

Physics and the communications industry

W. F. Brinkman and D. V. Lang

Bell Laboratories, Lucent Technologies, Murray Hill, New Jersey 07974

This review explores the relationship between physics and communications over the past 125 years. The authors find that four eras of major change in communications technology can be traced to the corresponding major discoveries of physics that directly influenced the communications industry. The four areas of physics that define these periods are electromagnetism, the electron, quantum mechanics, and quantum optics. [S0034-6861(99)02602-1]

I. INTRODUCTION

Today's communications industry is a leading force in the world's economy. Our lives would be vastly different without the telephone, fax, cell-phone, and the Internet. The commonplace and ubiquitous nature of this technology, which has been evolving over a period of 125 years,¹ tends to overshadow the dominant role that physics and physicists have played in its development.² The purpose of this review is to explore this coupling and to show that the communications industry has not only made use of the results of academic physics research but has also contributed significantly to our present understanding of fundamental physics. Due to limitations of space, we shall primarily use examples from the Bell System and Bell Laboratories.

The foundations of communications technology lay in the discoveries of the great physicists of the early 19th century: Oersted, Ampere, Faraday, and Henry. The telegraph was invented only seventeen years after the discovery of electromagnetism by H. C. Oersted in 1820. In spite of this connection, however, much of the early work on communications was done by inventors, such as Morse, Bell, and Edison, who had no formal scientific background. The telegraph was a fairly simple electromechanical system which did not require the development of new scientific principles to flourish commercially in the mid 19th century. Closer coupling between physics and communications occurred shortly after the invention of the telephone by Alexander Graham Bell in 1876. The telephone concept immediately captured the imagination of the scientific community, where the "hot physics" of the period was electromagnetism and wave propagation.

Over the next century, industrial physics research on communications improved the technology, as well as spawning fundamental results of interest to the broad physics community. The types of physics specifically devoted to communications have varied continuously with the evolution of the technology and discontinuously with major physics discoveries over the past 125 years. We

can identify four broad eras of physics that have impacted communications: (1) the era of electromagnetism (starting in 1820); (2) the era of the electron (starting in 1897); (3) the era of quantum mechanics (starting in the 1920s); and (4) the era of quantum optics (starting in 1958).

II. THE ERA OF ELECTROMAGNETISM

This era dates from 1820, when Oersted discovered that an electric current generates a magnetic field. The first electromagnet was built in 1825, and in 1831 Faraday and Henry independently discovered that electric currents can be induced in wires moving in a magnetic field. The concept of the electromagnet was exploited in two independently invented telegraph systems in 1837—an analog system by Cooke and Wheatstone in Britain and a digital system by Samuel Morse in the U.S., using a dot-dash code that he also invented for the purpose. Because of its simple and robust design, the latter system achieved widespread commercial use in less than fifteen years. In 1861 Western Union had completed the first transcontinental telegraph line across the U.S. The British physicist William Thomson (later Lord Kelvin) was largely responsible for the construction of the first successful transatlantic telegraph cable in 1866 and developed the mirror galvanometer needed to detect the extremely weak signals.

One of the many young inventors excited by telegraph technology was Alexander Graham Bell, who was also interested in teaching the deaf to speak. Bell was working on devices that would enable deaf people to visualize the sounds they could make but not hear. He constructed a mechanical strip chart recorder, known as a phonograph, which used human ear bones to couple the vibrations of a diaphragm to a stylus that traced the voice oscillations on a moving glass slide covered with lampblack. He was interested in making an electromagnetic analog for "electric speech" and as early as 1874 had the concept of vibrating a small magnet with sound waves and inducing a speech current in an electromagnet. In 1875, during his concurrent work on the multiplexed "harmonic telegraph," he accidentally discovered that useful audio-frequency currents could indeed be induced by a vibrating magnetic reed over an electromagnet and could be transmitted over such a system. By coupling the vibrating magnet to a diaphragm, as in his

¹Much of this review is based on the telecommunications histories compiled by Fagen (1975) and Millman (1983, 1984).

²The solid-state physics history is based on Hoddeson *et al.* (1992) and Riordan and Hoddeson (1997). The laser and optical communications history is based on Whinnery (1987), Agrawal and Dutta (1993), and Kaminow and Koch (1997).



FIG. 1. Alexander Graham Bell's first telephone. Voice sounds were transmitted for the first time on June 3, 1875, over this gallows-shaped instrument. From Fagen, 1975.

phonograph, he could transmit speech sounds. Figure 1 shows this first crude telephone, which Bell patented in 1876. For the next several years he empirically optimized the telephone while he and his financial backers started various corporations to exploit his invention.

Bell was not the only person exploring speech transmission. The idea of transmitting voice using “harmonic telegraph” technology also occurred to Elisha Gray of the Western Union Company. Working with Thomas Edison, who invented the carbon microphone in 1877, Gray developed and patented a telephone design that was technically superior to Bell's. The sound transmission efficiency of the transmitter-receiver pair in Bell's original telephone was roughly -60 dB. Edison's granular carbon microphone transmitter increased this efficiency by 30 dB. As a result, Western Union had 50 000 of Gray's telephones in service by 1881 when various patent lawsuits were settled giving the American Bell Telephone Company complete control of the technology. Note that the rapid commercialization of the telephone occurred within only five years of its invention—a time scale usually associated with contemporary computer technology.

While the telephone entrepreneurs were busy deploying the new technology, academic physicists were laying the groundwork that would be necessary to create a long-distance telephone network. James Clerk Maxwell developed the unified equations governing electromagnetism in 1864. Electromagnetic wave propagation, predicted by Maxwell, was observed in the 1880s by Heinrich Hertz. Also in the 1880s, the British physicist Oliver Heaviside applied Maxwell's theory to show that the propagation of speech currents over wires in telephone systems needed to be understood on the basis of wave propagation, not simple currents. In 1884, Lord Rayleigh showed that such speech currents would be exponentially attenuated in a telegraph cable, calculating that a 600-Hz signal would be reduced by a factor of 0.135 over 20 miles in the transatlantic telegraph cable (0.27 dB/km). His paper was very pessimistic about the pros-

pects for use of telephone technology for long-distance communication, compared to the well-established telegraph.

Results such as Rayleigh's, as well as the poor sound quality of the early telephone system, created the first real opportunity for physics in the infant telephone industry. In 1885, the managers of the engineering department of the American Bell Telephone Company, which Bell had left in 1881 and which later became AT&T, realized that it would not be easy to improve the transmission distance and quality on long-distance lines by trial and error. Therefore they formed a research department in Boston specifically focused on the physics of electromagnetic propagation on long-distance telephone lines. Hammond Hayes, one of Harvard's first physics Ph.D.s, organized the department and over the next twenty years hired other physicists from Harvard, MIT, Yale, Chicago, and Johns Hopkins to explore this new area of applied physics. This marked the beginning of industrial research in applied physics for communications.

The problem faced by early telephone researchers can be described rather simply from the vantage point of today's understanding. The original telephone transmission system was based on telegraph technology and used a single iron wire for each circuit with a return path through the ground. The attenuation A (in dB per unit length) of such a line at telephone frequencies ($\sim$ kHz) is given approximately by

$$A \sim R\sqrt{C/L} + G\sqrt{L/C} \quad (1)$$

where R is the series resistance, L the series inductance, C the shunt capacitance, and G the shunt conductance, all per unit length. At telegraph frequencies (~ 10 Hz) such a line is almost purely resistive. For a multiple-wire telephone cable, which has much higher capacitance and lower inductance, the attenuation is 10 to 25 times greater than Eq. (1) and can be approximated by

$$A \sim \sqrt{RC}f \quad (2)$$

where f is the frequency. The first solution to the telephone attenuation problem was to reduce R by an order of magnitude by replacing the iron wire with copper. A second problem was that the interference from outside sources picked up by the single, unshielded wire was an order of magnitude greater at telephone frequencies than for the telegraph. This was solved by adding a second wire to make a so-called “metallic circuit.” However, such solutions dramatically increased costs. First, the copper wire had to be of large enough gauge to be self-supporting, and second, twice as much copper was needed in a metallic circuit. The situation in long-distance cables was even worse, where the higher intrinsic attenuation of Eq. (2) could only be solved by using even heavier-gauge copper (the longest cables used one-tenth-inch-diameter wire). Thus cables were only cost effective within cities to solve the congestion and weather problems illustrated in Fig. 2.

By the turn of the century, an understanding of the physics of transmission lines had produced a dramatic

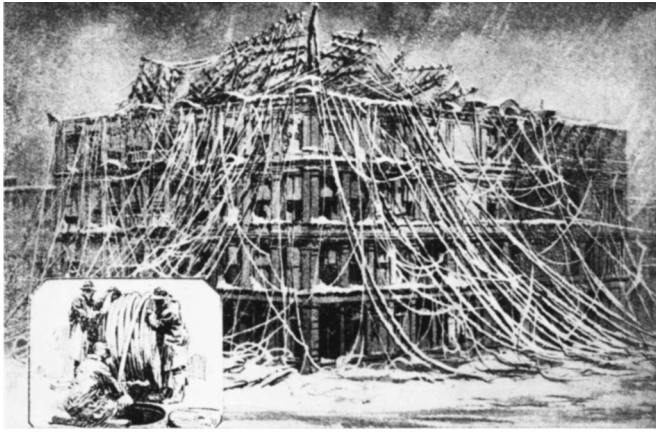


FIG. 2. Boston central telephone station at 40 Pearl Street after the blizzard of 1881. Inset shows the installation of underground cables which solved the weather and congestion problems of thousands of open wires. From Fagen, 1975.

solution to these problems. In 1887, Heaviside developed the transmission line theory which we now understand as Eqs. (1) and (2). He pointed out that the attenuation could be reduced by increasing the series inductance per unit length, which is obvious from Eq. (1) when the first term dominates, as is usually the case. In 1899, George Campbell, at AT&T's Boston laboratory, and Michael Pupin, at Columbia University, almost simultaneously concluded that discrete inductors could simulate the continuous inductance of Eq. (1) as long as the spacing was not larger than one-tenth of a wavelength. For telephone frequencies this corresponded to a spacing of eight miles in open lines and one mile in cables. The effect of these so-called "loading coils" was dramatic. The maximum transmission distance of open lines nearly doubled, thus allowing the long-distance network to extend from New York to Denver by 1911. The effect on telephone cables was even more dramatic. The attenuation was decreased by a factor of four and the frequency distortion of Eq. (2) was greatly reduced. Since cables were primarily used for relatively short distances, these gains were traded off against the series resistance of the conductor and allowed smaller-gauge wires to be used, thus saving an order of magnitude in the cost of copper for the same cable length.

In addition to the transmission enhancements of loaded lines, there was a strong desire to develop some kind of "repeater" that would strengthen the weakened signals and retransmit them. Electromechanical repeaters were commonly used in the telegraph system for many years and were the reason that the telegraph was quickly extended coast to coast. The early attempts to invent telephone repeaters were electromechanical analogs of the telegraph systems. A number of inventors patented telephone repeaters that were essentially a telephone receiver placed next to a carbon microphone. The carbon microphone modulated a large current by means of speech-induced vibrations of the weakly touching particles, and the resulting gain formed the basis of an amplifier. An improved version was developed at

AT&T's Boston laboratory and used commercially in the New York to Chicago route in 1905. Such electro-mechanical amplifiers had considerable distortion and narrow dynamic range, but their limited success served to focus research energies on finding a more useful type of amplifier. This leads us to the next era in the relationship between physics and the communications industry, when such an amplifier was indeed invented.

III. THE ERA OF THE ELECTRON

At about the same time that the telephone was invented, various electrical experiments in gas-discharge tubes were laying the physics groundwork for the next phase of communications technology. In about 1878, Sir William Crookes developed a specially designed tube to study the mysterious phenomenon of cathode rays which caused gas-discharge tubes to glow. In 1897, the cathode-ray studies of J. J. Thomson, Cavendish Professor of Physics at Cambridge, led to the discovery of the electron. The first thermionic vacuum-tube diode was invented by Sir John Fleming in 1904, following Edison's 1883 observation of current flow between the filament of a light bulb and a nearby electrode. Fleming's device was an excellent detector of wireless telegraph signals, which had been invented by Marconi in 1896. In 1907, Lee de Forest invented the vacuum-tube triode, which for five years remained almost the exclusive province of wireless entrepreneurs. It was a much better rf detector than Fleming's diode but was never used as a power amplifier.

Meanwhile, pressure to develop a telephone amplifier intensified. In 1909, Theodore Vail, the new president of AT&T, who was rapidly acquiring the small telephone companies that had sprung up after Bell's patents expired, set forth the vision that would define the Bell System for many years to come—"One Policy, One System, Universal Service." He probably did not know that his grand vision of coast-to-coast service was not possible without an effective telephone amplifier. In about 1910, Frank B. Jewett, who was in charge of transmission engineering at AT&T's Western Electric subsidiary, guessed that an improved amplifier might be possible with the "inertialess moving parts" of the electron beams that his friend Robert Millikan at the University of Chicago had been studying. Jewett asked Millikan to recommend one of his top students trained in the "new physics" whom he could hire to do amplifier research. The new hire, H. D. Arnold, started in 1911 and within two years achieved the goal of the useful vacuum-tube amplifier that Vail's vision required. In 1913, Arnold's vacuum-tube amplifier, shown in Fig. 3, was being used in commercial service. As a result, the Panama-Pacific International Exposition in San Francisco opened to great fanfare in 1915 with the dedication of the first transcontinental telephone circuit and a conversation between President Woodrow Wilson in the White House, Alexander Graham Bell in New York, and Bell's original assistant Thomas Watson in San Francisco. This unqualified success of physics convinced AT&T officials

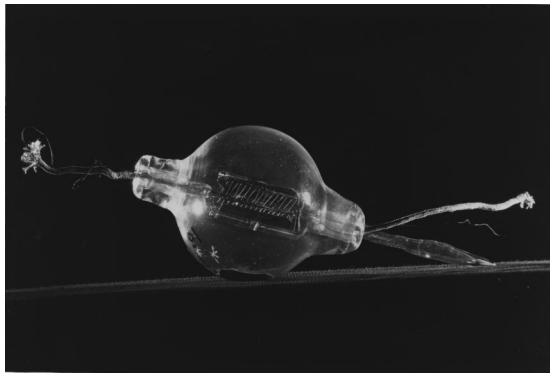


FIG. 3. Arnold's high-vacuum tube, first used as a telephone repeater at Philadelphia on a New York to Washington cable circuit in October, 1913. Other, later models, were used on the transcontinental circuit opened for service in 1915. From Fagen, 1975.

that paying topnotch Ph.D. physicists to do communications research was good business.

Arnold's vacuum-tube triumph was facilitated by de Forest's triode demonstration to Bell officials in 1912. Even though the demonstration failed under the output power levels required for telephone applications, Arnold immediately saw how to improve the tube. He made three improvements that were critical: a higher vacuum, an oxide-coated cathode, and a better grid position. Irving Langmuir of General Electric had independently pursued a high-vacuum design for improving the triode in 1913. These vacuum-tube amplifiers not only dramatically increased the transmission distance of telephone lines, but also made long-distance wireless telephony possible. In 1915, using high-power vacuum-tube amplifiers, Bell System scientists transmitted speech for the first time by wireless telephony from Arlington, Virginia, to both Paris and Honolulu. Wireless telephony was not widely used for two-way communications, however, until the development of microwaves during W.W. II and cellular wireless technology in the latter part of the 20th century.

Vacuum tubes also made possible dramatic increases in the capacity of the long-distance telephone system, which lowered the cost per call. This was done by multiplexing a number of calls in parallel over the same pair of wires using independently modulated carriers of different frequencies, a technique sometimes called "wired wireless." The idea of multiplexing was originally developed for the telegraph, using electromechanical resonators to generate and separate the various carrier tones. Indeed, Bell had been experimenting with just such a "harmonic telegraph" when he invented the telephone. However, to apply the same principle to voice-frequency signals required carriers of much higher frequency, ~ 5 –25 kHz. This was not possible without vacuum tubes for oscillators, modulators, and amplifiers. In addition, technology from the era of electromagnetism was needed for the bandpass filters to demultiplex the signals at the receiver. In about 1910, the Bell System theoretician George Campbell, who also introduced loading

coils, invented the so-called "wave filter." The value of this invention was not immediately appreciated, however, and the concept was not patented until 1917. It is of interest to note that today's technology for increasing transmission capacity on a glass fiber is wavelength-division multiplexing (WDM), which, although digital, is nevertheless almost an exact optical analog of the electronic multiplexing developed in the 1920s.

The fundamental physics underlying telephone transmission was fairly well established by 1920. Thus communications-related physics research turned to an exploration of the underlying materials. The primary areas were thermionic emission, noise in vacuum tubes, and the magnetic materials used in transformers and loading coils. During the 1920s physicists at the Western Electric laboratory, which became Bell Telephone Laboratories in 1925, made a number of important physics contributions in these areas. In 1925, J. B. Johnson made the first observation of the thermal noise predicted in 1918 by W. Schottky. Thermionic emission was first described theoretically by O. W. Richardson in Britain in 1901, and the benefit of alkaline-earth oxides in enhancing thermionic emission was observed by A. Wehnelt in Germany in 1904. About 10 years later, C. D. Child at Colgate and Langmuir at GE independently described the physics of space-charge-limited current, which is essential to the operation of a vacuum tube. Nevertheless, by 1920 there was still a long-standing controversy as to whether the enhanced performance of oxide-coated cathodes was due to chemical or physical effects. This was settled by Arnold, who showed in 1920 that the enhanced emission was indeed due to thermionic emission. In 1917, Clinton J. Davisson, a student of Richardson's at Princeton, came to Western Electric in order to understand the fundamental physics of why alkaline-earth oxides had a lower work function than tungsten. As a result of the invention of the ionization vacuum gauge by Oliver E. Buckley in 1916, it was possible to characterize the vacuum conditions necessary for this early surface-physics research, thus allowing Davisson and Lester Germer to do some of the key physics on thermionic emission from oxides and tungsten in the early 1920s. Physics research on magnetic materials was also important at Bell Labs during the 1920s. Permalloy had been developed at Western Electric for use in the loading coils discussed earlier. The first scientific paper on permalloy was given by Arnold and G. W. Elmen at the 1923 Spring Meeting of the American Physical Society. Another highlight of this period was Richard Bozorth's internationally acclaimed work on the Barkhausen effect.

Perhaps the most famous physics experiment at Bell Labs during the 1920s was the observation of the wave nature of the electron, for which Davisson received the 1937 Nobel Prize in Physics. This is indicative of the systematic business practice of recruiting the very best physicists to do applied research in the communications industry. In order to attract and retain top physics talent, it was necessary to allow such researchers the freedom to pursue the fundamental questions uncovered in their

applied research. The history of Davisson's Nobel Prize is instructive. At the same time that Davisson was working on thermionic emission, he was also pursuing experiments to study the atomic structure of magnetic metals using inelastic electron scattering. This was relevant to both magnetic materials and the problem of secondary electron emission in vacuum tubes. During these experiments, Davisson noticed that some of the electrons were elastically scattered. While pursuing this unexpected result, he noticed some angular structure in the scattering pattern. In 1925, W. Elsasser suggested that Davisson's data could be evidence of "de Broglie waves," which had been proposed in 1923. During a 1926 trip to Europe, Davisson obtained a copy of Schrödinger's paper on wave mechanics, which had been published earlier that year. Upon returning to Bell Labs, he and Germer repeated the experiments with a single crystal of nickel, looking at the specific angles where electrons would be diffracted according to Schrödinger's equation. The results led to their famous 1927 paper, which also introduced low-energy electron diffraction (LEED) as an important tool for surface physics. Similar stories hold for Karl Jansky, who discovered radio astronomy in 1933 while studying noise in transatlantic radio telephony, and Arno Penzias and Robert Wilson, who discovered the 3K microwave background radiation of the big bang in 1964 while studying noise in telecommunications satellites.

IV. THE ERA OF QUANTUM MECHANICS

The fundamental-physics roots of this era began with the explosive growth of quantum mechanics in Europe in the 1920s. The first application of this new theory to solids was Felix Bloch's 1928 quantum theory of metals. The foundations of semiconductor physics quickly followed with Rudolf Peierls's 1929 theory of the positive Hall effect due to holes, Brillouin's 1930 concept that band gaps are related to the Bragg scattering conditions, and Alan Wilson's 1931 band theory of semiconductors, including the effects of doping. A major development in the physics of real materials was E. Wigner and F. Seitz's 1933 approximate method for calculating band structure. This marked the beginning of a shift from the fundamental studies of the 1920s to the practical solid-state physics which would dominate the second half of the 20th century.

The experimental roots of semiconductor physics date from the 19th century. In the 1870s, at almost exactly the same time that Bell was inventing the telephone, physicists working on selenium, copper oxide, and various metallic sulfides (all materials we know today to be semiconductors) were discovering diode rectification behavior, the Hall effect, photoconductivity, and the photovoltaic effect. In fact, even the idea of inventing a solid-state analog of the vacuum tube had occurred to a number of people during the 1920s and 1930s—J. E. Lilienfeld patented the field-effect concept in 1926, and

Walter Brattain and Joseph Becker at Bell Labs contemplated putting a grid into copper oxide rectifiers during the 1930s.

By the late 1930s, solid-state physics was well established and had the potential for major applications. In a move reminiscent of earlier eras, Mervin Kelly, Bell Labs' Director of Research, sought out the best of the new breed of solid-state physicists to explore the potential of semiconductors for communications; in 1936 he hired William Shockley from John Slater's group at MIT. However, the effort to make devices of possible use in communications, e.g., solid-state switches or amplifiers, did not start seriously until 1946 when nonmilitary research resumed at Bell Labs after W.W. II. Shockley was put in charge of a new solid-state research group specifically chartered to obtain a fundamental understanding of the device potential of silicon and germanium, which had been developed into excellent microwave detectors during the war. One of his first moves was to hire John Bardeen. The subsequent path to success was as rapid as Arnold's development of the vacuum-tube amplifier in 1912. The point-contact transistor was demonstrated within two years, by the end of 1947. The birth of the transistor is covered in a number of 50th Anniversary reviews (Brinkman *et al.*, 1997; Riordan and Hoddeson, 1997; Ross, 1998), including one in this volume (Riordan *et al.*, 1999, which includes a photograph of the first transistor). Therefore our focus will be to review the relationship of the transistor to the technology changes that have revolutionized communications over the past 50 years.³

The application of the transistor to communications occurred in two phases. The first, during the 1950s, was simply the replacement of vacuum tubes in various circuits. The first commercial use of the transistor in the Bell System was in 1954; the first fully "transistorized" product (the E6 repeater) was in 1959. There were some benefits of size and power reduction, but the functionality and design of the telephone system was not changed. In the second phase, the transistor made possible digital transmission and switching—an entirely new communications technology that revolutionized the industry. The concept of digital voice communications, known as pulse code modulation (PCM), was first demonstrated in 1947 at Bell Labs. This early demonstration was based on voice coding ideas developed in the 1930s and telephone encryption devices used by the military during the war. Commercial use of PCM, however, was not possible without transistors to make the complex circuits practi-

³We should note in passing, however, that Bell Labs' broader focus on physics research in areas relevant to communications—magnetism, semiconductors, and surfaces—grew substantially during the postwar period and is still a major effort today. Even though much of this work is directly relevant to communications, it is well beyond the scope of this review to cover it extensively. One of the highlights is P. W. Anderson's theoretical work on magnetism and disordered solids during the 1950s, for which he received the 1977 Nobel Prize in Physics.

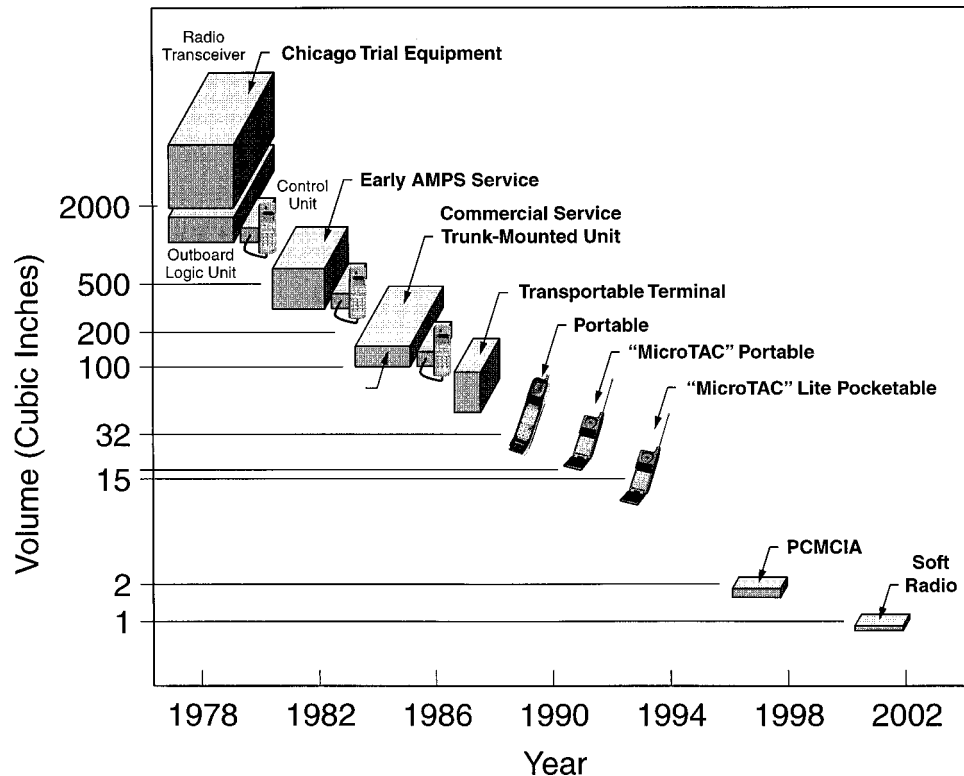


FIG. 4. Size reduction of cellular telephones as a result of progress in increasing the number of transistors in an integrated circuit chip, as expressed by Moore's Law.

cal. The first digital transmission system, the so-called *T1* carrier, was introduced in 1962 and carried 24 digital voice channels with an overall bit rate of 1.5 Mbit/sec. Even though a combined digital switching and transmission system was demonstrated in 1959 at Bell Labs, the first commercial use of fully digital switching and transmission was not until the introduction of the 4ESS switch for long-distance traffic in 1976.

Digital technology has profoundly affected the communications industry. It could be argued that the 1984 breakup of the Bell System was due to the ease with which various competitors could develop digital telephone systems, as opposed to the complex electromechanical switching systems that required the substantial resources of the Bell System to develop and maintain. An additional factor was the close relationship between the digital technology for communications and for computers. Thus the 1958 invention of the integrated circuit by J. S. Kilby at *TI*, along with major improvements by R. N. Noyce at Fairchild and the 1960 invention of the MOSFET (metal-oxide-semiconductor field-effect transistor) by D. Kahng and M. M. Attala at Bell Labs, affected both industries in fundamental ways. As a result of the exponential growth in the number of transistors per chip (Moore's Law), the cost and size of electronic devices have changed by orders of magnitude since the 1960s. In communications this made possible the wireless revolution that we are seeing today. Figure 4 shows the reductions in size of cellular telephone equipment since the invention of the concept by Bell Labs in 1960; dramatic reductions have also occurred in the cost. In a

similar way, digital electronics and integrated circuits have made the "old" technology of facsimile transmission, introduced by AT&T in 1925, into the practical communication medium of today's ubiquitous Fax machine.

At this moment we stand on the threshold of yet another revolution brought on by the transistor and integrated circuit—the Internet. The widespread use of personal computers and the Internet have made data networking one of the hottest growth industries. Because of digital technology, telecommunications is being redefined to include data and video, as well as voice. It is ironic that after the telephone made the telegraph obsolete, the latest technology is essentially reverting back to an ultrafast version of the telegraph, in which coded digital messages, not analog voice, dominate the system. However, this revolution is not based solely on the transistor and integrated circuit. We require the ultrahigh transmission bandwidth of fiber optics to complete our story. This brings us to the fourth era in the relationship between physics and communications—learning how to communicate with light.

V. THE ERA OF QUANTUM OPTICS

Alexander Graham Bell invented the photophone in 1880, just a few years after the telephone. This device used a mirrored diaphragm to modulate a beam of sunlight with speech vibrations, analogous to the modulation of an electric current in Bell's telephone. The re-

ceiver was a selenium photocell, which had been discovered in 1876. The fiber-optic communications systems of today differ in only three major respects: (1) a glass waveguide, invented by the British physicist John Tyndall in 1870, replaces free-space propagation; (2) high-speed, multichannel digital modulation replaces the single analog voice channel; and (3) a coherent light source replaces the sun. The coherence of the source is critical for two reasons. First, coherent light has much less noise than incoherent, and second, coherent light can be focused into a high-power beam without much loss in intensity. In fact, the Bell Labs communications engineer W. A. Tyrrell pointed out in 1951 the advantages of using optical frequencies instead of microwaves for communications, but noted the lack of the required coherent source. The major impact of physics has been to provide the coherent light source—the laser.

The date commonly cited for the invention of the laser is the theory of A. Schawlow and C. H. Townes in 1958 (see the articles in this issue by Lamb *et al.* and by Slusher). However, the fundamental physics of stimulated emission was first recognized in 1917 by Einstein. Stimulated emission was first observed at microwave frequencies (24 GHz) in the ammonia beam maser by Townes in 1955. N. Bloembergen demonstrated in 1956 the importance of a three-level system in obtaining the population inversion for maser amplification. In 1957, Bell Labs developed the first solid-state maser, which was used as a low-noise microwave amplifier in the Telstar communications satellite in 1962. With such a background, Schawlow and Townes's 1958 theory for extending stimulated emission to optical frequencies was not terribly surprising. However, it was two years before the first laser (the pulsed ruby laser) was demonstrated by Theodore Maiman at Hughes Research Labs in 1960. Subsequent lasers of various types came at a rapid pace in the early 1960s.

The laser most directly relevant to communications is the semiconductor laser, demonstrated independently in 1962 by GE, IBM, and Lincoln Laboratory. This type of laser can be fabricated from a variety of direct-band-gap III-V semiconductors, which were introduced by H. Welker in 1952. However, it was not until 1970 that laser diodes could be made to operate continuously at room temperature, a prerequisite for optical communications. In order to explain this and other improvements in laser diodes for communications, we must return to the previous era of semiconductor physics and the concept of the heterojunction (see Riordan *et al.*, 1999 in this issue). A heterojunction is formed when two semiconductors of different band gap are joined together. William Shockley pointed out in 1951 that the performance of a bipolar transistor would be enhanced if the emitter had a wider band gap than the base, i.e., the emitter-base junction was also a heterojunction. In 1963, it was independently suggested by Herbert Kroemer and by Zh. I. Alferov and R. F. Kazarinov that a laser having such heterojunctions would be superior to the homojunction diodes first demonstrated. Materials of different band gaps, however, generally also have different lattice con-

stants, making good crystal growth difficult, if not impossible. Therefore heterojunctions remained a theoretical curiosity until 1967 when J. M. Woodall, H. Rupperecht, and G. D. Pettit at IBM produced good-quality GaAs/Al_xGa_{1-x}As heterojunctions by liquid-phase epitaxy. In 1970, I. Hayashi and M. B. Panish at Bell Labs and Alferov in Russia obtained continuous operation at room temperature using double-heterojunction lasers consisting of a thin layer of GaAs sandwiched between two layers of Al_xGa_{1-x}As. This design achieved better performance by confining both the injected carriers (by the band-gap discontinuity) and emitted photons (by the refractive-index discontinuity). The double-heterojunction concept has been modified and improved over the years, but the central idea of confining both the carriers and the photons by heterojunctions is the fundamental approach used in all semiconductor lasers.

The second essential ingredient for optical communications is low-loss silica fiber. Fiber-optic illuminators were developed in the mid 1950s, but this type of glass has an attenuation of about 100 dB/km and would only work for communications systems a few hundred meters in length. A major breakthrough occurred in 1970, when F. P. Kapron and co-workers at Corning produced the first fiber with a loss less than 20 dB/km at 850 nm, the GaAs/Al_xGa_{1-x}As laser wavelength. This marked the beginning of optical fiber communications technology. By 1976, fiber loss was reduced to 1.0 dB/km at 1300 nm, increasing the distance between repeaters to tens of miles. The first commercial trial of optical communications was carried out in Chicago in 1978; by 1982, fifty systems were installed in the Bell System, corresponding to 25 000 total miles of fiber. In 1985, the loss in silica fiber reached a low of 0.15 dB/km at a wavelength of 1550 nm. By this time, the systems had migrated to the lower-loss long-wavelength region of 1300 to 1550 nm, and the laser diodes were being fabricated in the InP/InGaAsP materials system. The first transatlantic optical fiber system (TAT-8), operating at 1300 nm, was installed in 1988.

The capacity of commercial optical fiber communications systems has increased about as fast as Moore's Law for integrated circuits, doubling every two years, as shown in Fig. 5. The recent rate of increase for experimental systems has been even faster. In the latest of a sequence of major advances, Bell Labs reported in 1998 the transmission of 1.0 Terabit/sec over a distance of 400 km on a single fiber using 100 different wavelengths, each modulated at 10 Gb/sec—the so-called dense wavelength-division-multiplexing (DWDM) technology. Three additional elements of physics are necessary, however, to complete our story and bring us to the technology of the 1990s: quantum wells, optical amplifiers, and nonlinear optics.

A quantum well is a double-heterojunction sandwich of semiconductors of different band gap, discussed above, with the central, lower-band-gap layer so thin (typically less than 100 Å) that the quantum states of the confined carriers dominate the properties of the mate-

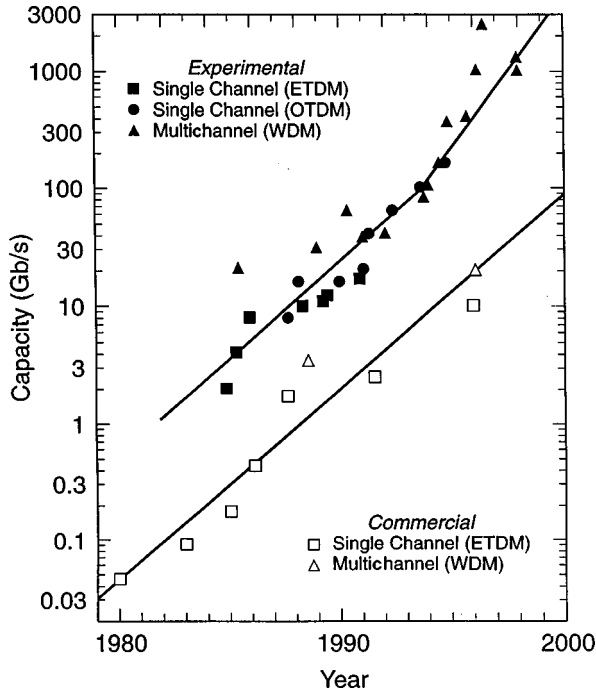


FIG. 5. Progress in optical fiber transmission capacity. The growth in commercial capacity is due to the increasing bit rate of electronic-time-division multiplexing (ETDM) and the introduction of multichannel wavelength-division multiplexing (WDM). Experimental systems used ETDM, WDM, and optical-time-division multiplexing (OTDM) to achieve record-setting results.

rial. This was first demonstrated by R. Dingle, W. Wiegmann, and C. H. Henry at Bell Labs in 1974. The first quantum-well laser was made at Bell Labs in 1975, and a patent on the quantum-well laser was granted to Dingle and Henry in 1976. Such structures were made possible by the crystal-growth technique of molecular-beam epitaxy, developed into a high-quality epitaxy method by A. Y. Cho in the early 1970s. By using multiple quantum wells in the active layer of a laser diode, the properties needed for communications systems are dramatically improved. In fact, most multiple-quantum-well lasers today also use intentionally strained quantum-well layers, a method first proposed by G. C. Osbourn at Sandia in 1982 to modify the band structure and further enhance device performance. The most recent device, which is essential for wavelength-division multiplexing systems, is a laser and electroabsorption modulator integrated on the same chip. In this case both the laser and modulator are based on quantum wells, with the modulator using the quantum-confined Stark effect discovered at Bell Labs by D. A. B. Miller and co-workers in 1985.

The physics of quantum wells is another example in which communications research has impacted the broad physics community. The 1978 invention of modulation doping by Stormer and co-workers at Bell Labs made possible ultrahigh-mobility two-dimensional electron systems which have dominated the research on mesoscopic and quantum-confined systems over the past twenty years. In 1982, D. C. Tsui, H. L. Stormer, and A.

C. Gossard at Bell Labs discovered the fractional quantum Hall effect using high-mobility quantum-well structures. Research on the fractional quantum Hall effect is still a hot topic in fundamental physics today (see Stormer and Tsui, 1999, in this issue). Another recent example of fundamental quantum-well research is the quantum cascade laser, invented by J. Faist, F. Capasso, and Cho at Bell Labs in 1994. This is the first semiconductor laser to use the quantum-well states, rather than the band gap, to define the wavelength of laser emission. Such lasers currently operate at wavelengths longer than those needed for optical communications; however, the insights gained will most likely impact communications systems in the future.

The optical amplifier is a major mid 1980s advance in communication systems. Such amplifiers are, of course, based on the same physics of stimulated emission as the laser and maser, but it was not at all clear whether such an amplifier would have the required low noise and low cross-talk properties essential for communications systems. Several amplifier designs were explored, including laser diodes with antireflective coatings and Raman amplifiers, but the best amplifier for communications systems proved to be optically pumped erbium-doped silica fiber. The considerable body of physics knowledge developed in connection with rare-earth ion lasers (such as Nd:YAG) invented in the mid 1960s greatly accelerated the development of the erbium-doped fiber amplifier (EDFA). Basic research at Bell Labs and elsewhere on the spectroscopy of rare-earth ions in various matrices, including silica glass, was a necessary precursor for the rapid development of EDFAs. By the late 1980s, EDFAs were widely used in experimental systems. Optical amplifiers make WDM systems possible, since the parallel wavelength channels can be simultaneously boosted without each having to be separated and routed through an expensive optoelectronic repeater typical of older fiber-optic systems.

Nonlinear optics was discovered in 1961 with the observation of two-photon absorption, almost immediately after the first laser was constructed in 1960. Even though the optical nonlinearities of silica are very small, such effects become important in communications systems because of the long distances involved. In 1975, Roger Stolen observed four-photon mixing in silica fibers. This third-order nonlinear effect, analogous to intermodulation distortion in electrical systems, emerged as a serious problem for WDM systems in the 1990s. The problem arose from the typical practice of designing optically amplified transmission systems with the chromatic dispersion of the fiber carefully adjusted to zero, to prevent pulse spreading as a result of the finite spectral width of the modulated source. Such a system creates good phase matching between the WDM channels and, hence, generates four-wave mixing products that seriously degrade the performance of very long systems. The solution was to introduce a new fiber design, the so-called True Wave fiber invented by Bell Labs in 1993, which introduces a small, controlled amount of dispersion to break up the

phase matching and prevent four-wave mixing. This makes WDM technology practical.

The extremely high transmission capacity possible with today's optical fiber systems (a Terabit/sec corresponds to over 15 million simultaneous phone calls) along with the extremely large number of transistors on a single chip (over one billion expected in 2001) will undoubtedly lead us into future eras of communications technology which will be poised to benefit from future discoveries of physics.

VI. CONCLUSIONS

We have shown the critical impact of the four major eras of physics on the communications industry over the past 125 years. The industry rapidly applied the major physics discoveries during this period and thus made dramatic improvements in communications technology, with demonstrable benefits to society. We note that in all four major eras of physics—electromagnetism, the electron, quantum mechanics, and quantum optics—the fundamental discoveries were applied by the communications industry within 15–20 years. Further, it is evident that the communications industry's practice of employing the best physicists to do both basic and applied research resulted in the successes noted in this review. Indeed, the development of marketable communications technology would most certainly not have occurred so rapidly had this not been the case. This reality was well understood by such visionary communications research leaders as Hayes, Jewett, Buckley, and Kelly. They recognized the potential of applying the latest physics dis-

coveries to enhance communications technology. The current leaders of the communications industry continue in this tradition.

REFERENCES

- Agrawal, G. P., and N. K. Dutta, 1993, *Semiconductor Lasers* (Van Nostrand Reinhold, New York).
- Brinkman, W. F., D. E. Haggan, and W. W. Troutman, 1997, *IEEE J. Solid-State Circuits* **32**, 1858.
- Fagen, M. D., 1975, *A History of Engineering and Science in the Bell System: The Early Years (1875–1925)* (Bell Telephone Laboratories, New York).
- Hoddeson, L., E. Braun, J. Teichmann, and S. Weart, 1992, *Out of the Crystal Maze* (Oxford, New York).
- Kaminow, I. P., and T. L. Koch, 1997, *Optical Fiber Telecommunications IIIA* (Academic, San Diego).
- Millman, S., 1983, *A History of Engineering and Science in the Bell System: Physical Sciences (1925–1980)* (Bell Telephone Laboratories, Murray Hill, NJ).
- Millman, S., 1984, *A History of Engineering and Science in the Bell System: Communications Sciences (1925–1980)* (AT&T Bell Laboratories, Murray Hill, NJ).
- Riordan, M., and L. Hoddeson, 1997, *Crystal Fire: the Birth of the Information Age* (Norton, New York).
- Riordan, M., L. Hoddeson, and C. L. Herring, 1999, *Rev. Mod. Phys.* **71** (this issue).
- Ross, I. M., 1998, *Proc. IEEE* **86**, 7.
- Stormer, H. L., and D. C. Tsui, 1999, *Rev. Mod. Phys.* **71** (this issue).
- Whinnery, J. R., 1987, *Lasers: Invention to Application* (National Academy, Washington, D.C.).