I. HISTORY AND CONTEXT

Raymond Alphonsus Heising (Fig. 1), born in 1888, contributed to the development of radio by demonstrating (in 1923, at the age of 35) the first transatlantic radio communication using single-sideband modulation.

Heising’s contribution built on the work of a small number of prominent pioneers in the development of radio communication [1]. The German physicist Heinrich Hertz, in a series of spectacular experiments (1885–1889), confirmed the little appreciated or understood electromagnetic theory of James Clerk Maxwell (1864) by demonstrating electromagnetic waves and confirming their finite velocity. Hertz showed no interest in the practical applications of his discoveries. It was Guglielmo Marconi, an Italian working in England, who was determined to create a wireless telegraphy system and who received faint signals in Newfoundland transmitted from Poldhu, Cornwall (1901). Marconi did not realize that the success of his experiment depended on the reflection of the ionosphere, which was not discovered until 1924. Had he carefully analyzed the propagation, he probably would have concluded that his objective was not feasible!

Both Hertz and Marconi used a noise-like signal generated by a spark gap and filtered by the resonance of an antenna system and a receiving antenna tuned by a similar resonance with no amplification. Marconi had complete faith in the telegraph and did not believe it compelling to transmit speech or music by radio. It was Reginald Fessenden, a Canadian working in the United States, who pursued a vision of transmitting “words without wires” by continuous wave (CW) modulation. His transmitter, patented in 1901 but not demonstrated until 1906, consisted of a high-frequency alternator (rotating machine) connected to an antenna system with a series carbon microphone (technology borrowed from the telephone). On Christmas Eve, 1906, Fessenden made the first wireless radio broadcast from Brant Rock, MA, heard by radio operators using his receivers as far away as the West Indies. CW was not only a more polished method of transmitting speech or music but also had the benefit of cleanly separating distinct transmissions by carrier frequencies sufficiently separated, which we now call “frequency division multiplexing” (FDM).

Interestingly, Alexander Graham Bell’s invention of the telephone in 1876 arose out of his attempts (begun in 1873) to build a “harmonic telegraph.” The harmonic telegraph could transmit multiple telegraph signals on a single wire by separating them in frequency [3]. While Bell did not
actually perfect such a device, he might be credited with the notion of FDM.

A different thread of invention had a profound impact on the development of radio [4]. Building on a discovery by Thomas Edison in the United States arising out of his work on the light bulb (in 1883, four years before the discovery of the electron), J. A. Fleming, in England, invented the “Fleming valve” (1904), later known as a “vacuum tube diode.” The diode rectification property noted by Fleming later proved to be an excellent detector for the amplitude CW modulation of Fessenden. L. de Forest of the United States added a third element to this contraption (1906) called a “grid” by creating an “audion,” later called a “vacuum tube triode.” The triode later proved valuable as an oscillator—eliminating the need for the cumbersome high-frequency alternator—and also as a demodulator and amplifier.

Arguably the most important idea in the development of CW transmission was Fessenden’s invention (and naming) of the “heterodyne method” of detecting CW by combining two frequencies to yield the sum and difference frequencies (1902). Heterodyne was made practical by H. Armstrong in 1913, using de Forest’s audion in a circuit with positive feedback (called “regeneration”).

This brings us to Heising’s demonstration of a single sideband. Transatlantic radio telephony had first been demonstrated in 1915 [3]. Heising built a working prototype of a single-sideband modulation (earlier invented by J. R. Carson) transmitter and receiver and demonstrated its operation across the Atlantic Ocean. His work depends on the CW modulation of Fessenden but is remarkably more sophisticated, as we will discuss shortly. His implementation is completely dependent on Fleming’s and de Forest’s vacuum-tube technology, and he himself made important advances in the application of vacuum tubes to modulation and detection in radio as early as 1915. He generously gives credit to others for the basic ideas and reports in detail on his implementation [5].

II. CW MODULATION REVISITED

A modern treatment of CW modulation would write the transmitted signal, in a form sufficiently general to represent all variations, as [2]

\[ y(t) = c(t) \cdot e^{j\omega t} \]

\[ = a(t) \cdot \cos(\omega t) - b(t) \cdot \sin(\omega t) \]  

(1)

where \( c(t) \) is a complex-valued signal, called the complex-baseband signal, expressed in terms of real and imaginary parts as

\[ c(t) = a(t) + j \cdot b(t). \]  

(2)

A. Fessenden: Transmitted-Carrier Amplitude Modulation (AM)

If we let the complex-baseband signal be real-valued and of the form

\[ a(t) = A + m(t) \]  

(3)

where \( m(t) \) is a real-valued message signal, such as voice, and \( |m(t)| \leq A \), then we have the transmitted-carrier AM demonstrated by Fessenden and still in widespread use today on the AM broadcast radio band. While AM is easy to demodulate with just a rectifier and low-pass filter—the only feasible technology available to Fessenden—the primary reason Heising and his colleagues were interested in abandoning AM was the large power (at least two-thirds of the total) consumed by the carrier \( A \). For transatlantic transmission, the required message signal power was large, and there was concern about energy conservation.

B. Suppressed-Carrier AM

A major advantage of AM is the ability, because the carrier is included with the message signal, to demodulate without precise knowledge of the carrier frequency or phase. If we try to reduce the transmitted power by eliminating the carrier, we immediately run into this problem.

Taking the general CW modulated signal with a suppressed carrier, we must generate our own carrier reference in order to demodulate. Assume that this reference has frequency \((\omega + \Delta \omega)\) rather than \(\omega\) and phase \(\xi\) rather than zero. The effect is illustrated by homodyne demodulation (direct to baseband without an intermediate frequency stage as in heterodyne), multiplying \( y(t) \) by \( e^{-j(\omega+\Delta \omega)t+\xi} \)

\[ y(t) \cdot e^{-j(\omega+\Delta \omega)t+\xi} = c(t) \cdot e^{-j(\Delta \omega t+\xi)} \]  

(4)

where we have neglected double-frequency terms easily removed by a low-pass filter. The effect is to yield the complex-baseband signal modulated by the low frequency \(\Delta \omega\) and phase shift \(\xi\).

If, like Heising, we are determined to rid ourselves of the transmitted carrier, we might use double-sideband amplitude modulation (AM-DSB), which is like AM but with \( A = 0 \). The demodulated signal is then \( m(t) \cdot e^{-j(\Delta \omega t+\xi)} \). In the absence of a reference carrier frequency and phase offset, the message can be recovered as the real part. Unfortunately, frequency and phase offset cause a low-frequency modulation and would be subjectively very poor. In the absence of frequency offset, a worst-case phase will result in a zero signal. Today, we know that we could estimate the carrier frequency and phase, for example by building a tracking loop that forced the imaginary part of the demodulated signal to zero, but with the technology of Heising’s day that would have been very difficult, especially since negative feedback was not invented until 1930. As a practical matter, Heising was forced to use a locally generated carrier without tracking.

C. Heising: Single-Sideband AM

Fortuitously, single-sideband amplitude modulation (AM-SSB) helps immensely, and in 1923 it was the only practical way to eliminate the transmitted carrier. We form a complex-baseband signal as

\[ c(t) = \text{positive frequencies in } m(t) \]  

(5)
that is, the message signal put through a complex-valued filter (called a “phase splitter”) that passes only positive frequencies. In this case, \( c(t) \) is truly complex-valued, and \( m(t) \) is the real part of \( c(t) \). Now, if we take the real part of the demodulated signal and ignore double-frequency terms, we get

\[
\text{Re} \{ c(t) \cdot e^{-j(\Delta \omega t + \phi)} \}. \tag{6}
\]

In the absence of frequency and phase offset, this is \( m(t) \). It turns out that, at least for speech, the result is subjectively good even in the presence of small frequency offsets (say, a few hertz). This is best appreciated by a picture of the magnitude spectrum, as in Fig. 2. Shown there is a hypothetical message-signal magnitude spectrum \( M(j\omega) \) and the magnitude spectrum of the demodulated signal with a carrier frequency offset for both DSB and SSB. For AM-DSB, the demodulated signal consists of the superposition of the message signal displaced by \( \pm \Delta \omega \) (one of the replicas is conjugated, which is not evident from the magnitude plotted in the figure) and these displaced spectra overlap, creating major distortion. For AM-SSB, the demodulated spectrum is faithful to the original except that all frequencies are displaced by \( \Delta \omega \). There is no overlapping of the two displaced replicas because one sideband is missing in each. Surely this frequency displacement would be deleterious to musical overtones (which would no longer be strictly overtones), but for speech (Heising’s intended application) the subjective effect is imperceptible for small frequency offsets.

It is interesting to compare the analysis just given with Heising’s paper, which uses trigonometric identities and pictures and no complex signals or equations. Is our modern approach unnecessarily analytical, perhaps simply meant to impress, or is there some practical utility to introducing complex-baseband signals? There definitely is utility, for in not using complex-baseband signals, Heising had to reverse the heterodyne process in the transmitter. He first generates an AM-DSB signal at an intermediate frequency, uses a highpass filter to remove the upper sideband, and modulates again to the final radio-frequency carrier frequency. Had he realized that he could do the phase-splitter filtering at baseband, albeit at the expense of dealing with a complex-baseband signal (in practice two signals rather than one), he would have had an alternative implementation path avoiding two stages of modulation. There are additional insights flowing from the complex analysis above, as we describe shortly.

III. DEVELOPMENTS FOLLOWING AM-SSB

AM-SSB has been used extensively since Heising’s demonstration, predominantly in analog telephony coaxial carrier systems and, since the 1950’s, by radio amateurs. It is interesting that it has not been used in radio broadcast. We still use AM for broadcast, as well as H. Armstrong’s frequency modulation (FM). AM persisted all these years initially because of the inexpensive receiver and more recently because of inertia. FM offered higher signal bandwidth (although AM could have accomplished this just as well), greater noise immunity (because it is a nonlinear modulation method), and eventually stereo broadcast (by adding an AM-DSB subcarrier). AM-SSB may not have been used in broadcast because it cannot faithfully reproduce frequencies near dc, as no practical filter can perfectly separate the lower and upper sidebands. This makes AM-SSB suitable for audio (as long as there is an accurate carrier frequency reference) but totally unsuitable for video (which has a prominent low-frequency content).

The reason AM-SSB was used extensively in coaxial carrier systems is a feature mentioned by Heising, the lower bandwidth of the transmitted signal (equal to the message-signal bandwidth rather than double that bandwidth, as in AM-DSB). This doubles the number of voiceband signals carried by coaxial cable. This bandwidth savings is also important for television broadcast, so a variation on AM-SSB, vestigial sideband (AM-VSB), was invented. VSB faithfully renders the low frequencies by transmitting just a small portion of the lower sideband.

Arguably the greatest long-term significance of the work of Heising and his colleagues is that AM-SSB was the first modulation system to use two quadrature carriers (although Heising was probably not aware of this). As seen in (1), because the AM-SSB complex-baseband signal is complex, both an in-phase and a 90°-offset quadrature carrier are superimposed. This is characteristic of all subsequent advanced modulation systems, including AM-VSB and FM, but most importantly leads directly to quadrature amplitude modulation (QAM), in which two real-valued (double-sideband) signals \( a(t) \) and \( b(t) \) are simultaneously imposed on in-phase and quadrature carriers.

QAM is arguably the most important modulation system for the future because it is widely used in digital communications, and all communications are rapidly becoming digital. The complex-baseband signal for QAM is of the...
where $A_k$ is a sequence of complex-valued data symbols chosen from a two-dimensional signal constellation representing transmitted data bits. A problem for QAM is that mentioned above for AM-DSB, namely, the need for an accurate phase and frequency carrier reference. However, data transmission has a big advantage in this respect in that the constellation is discrete and it is relatively simple to track carrier phase and frequency by observing the rotation and alignment of the discrete data points in the received constellation. A further advance in communications in the presence of noise is the observation by Claude Shannon (1948) that it is advantageous to group $2N$ complex-valued transmit symbols together, considering them to be a $2N$-dimensional signal vector. Choosing these vector symbols in a higher dimensional space under a power constraint allows us to keep them further apart, improving noise immunity [2].

QAM heavily exploits the simultaneous modulation of in-phase and quadrature carriers. Heising and his colleagues were the first to demonstrate this concept, although they did not formulate or recognize their contribution in those terms.

IV. Reflections on the Process

Of the pioneers of radio mentioned in Section I, it is interesting to note that Hertz, Fleming, Armstrong, and Fessenden were university faculty, while Edison, Marconi, and de Forest were independent inventors and entrepreneurs. Heising is the only pioneer mentioned who worked for a large corporation, Western Electric. Heising, however, is but one in a long line of important contributors to radio, telecommunications, and electronics technology from AT&T research (after 1925 called Bell Laboratories) and other industrial laboratories such as RCA and GE. Industrial laboratories soon came to dominate innovations in radio and electronics. De Forest for one lamented the declining importance of the independent inventor during the later years of his life [4].

Bell Laboratories defined and demonstrated the concept of industrial research, a model that was successfully emulated by other corporations. These laboratories proved successful by having sufficient freedom to encourage risky and high payoff explorations while making it relatively easy on the inventor by providing built-in funding, facilities, and a collegial environment.

Heising’s research was typical of both industrial and academic engineering research during his age in that it was directed at a very specific and practical goal. Following World War II, both industrial and academic engineering research took a partial turn from being “goal-oriented” toward a “curiosity-driven” model long practiced by scientists like Hertz, where a foreseeable practical outcome was not necessary. While today we lament the declining support for such research in both academe and industry, in fact this is really a return to an earlier prewar age of goal-oriented engineering research. Interestingly, we are at the same time witnessing a renaissance of the independent inventor and entrepreneur, particularly in the software industry, where the barriers to entry are relatively low.

Today, it may well be true that goal-driven research with longer time horizons is in danger of being slashed along with curiosity-driven research. A significant issue here is the increasing tendency to apply traditional financial metrics to the support of research, especially research financed by industry. A financial metric discounts revenues earned $k$ years in the future resulting from research supported today by $p^k$, where $p = 1/(1 + r)$ and $r$ is a prevailing annual interest rate or “cost of money.” The payoff of Heising’s work in modern QAM would be assigned a very low value in these financial terms, given the long gestation period, although it would be greatly valued in technical or societal terms. Similarly, the research 25 years ago in defining the Internet could not have been highly valued financially, even though today we would value it highly in both financial and societal terms. Much of the success of Bell Laboratories and its predecessors up until the 1970’s arose from the distinctly different societal-benefit model under which it operated. The research was valued from the perspective of achieving “universal telephone service” through a progressive reduction in the cost of telephony and was largely paid for not from future revenues but from an assigned fixed percentage of every telephone bill. This provided a stable source of support (another important characteristic of successful long-time-horizon research) divorced from direct coupling to discounted future revenues. Furthermore, after the “consent decree” of 1954, Bell Laboratories was obligated to license its technology freely to the entire industry. In other words, this successful industrial model of research was actually supported by a broad-based taxation mechanism, and the benefits were widely disseminated to society at large. The fundamental problem with a financial metric is that it maximizes our own wealth while ignoring the wealth or well-being of our children and grandchildren.

Today, we are increasingly decreeing that industrial research must be justified in conventional financial terms based on the resultant revenues accruing to the organization financing that research. Research outcomes often diffuse rather quickly, becoming “public goods,” in the terminology of economists. (Of course, patents are designed to counter this, with some effectiveness.) Research output as public goods results in economic inefficiencies in that each company has little incentive to contribute to these public goods but every incentive to exploit them, resulting in inadequate (from a societal view) investment in the research breeding these public goods. The question is, if such a financially driven system of research were in place during the period 1864–1923, would it have been possible to develop radio technology as effectively as we observed then? I think not. It is critically important that government (and industry insofar as it is possible) continue and expand support of long-time-horizon
and curiosity-driven research. While the time horizons may be long, the return on investment is eventually very high.

REFERENCES


David G. Messerschmitt (Fellow, IEEE) received the B.S. degree from the University of Colorado, Boulder, and the M.S. and Ph.D. degrees from the University of Michigan, Ann Arbor.

Prior to 1977, he was with AT&T Bell Laboratories in Holmdel, NJ. He currently is the Roger A. Strauch Professor of Electrical Engineering and Computer Sciences at the University of California at Berkeley. From 1993 to 1996, he was Department Chair. His current research interests include issues overlapping signal processing (especially video and graphics coding) and transport in broadband networks with wireless access, network services, and protocols for multimedia, wireless multimedia computing, network management, and the economics of networks. He has served as a consultant to a number of companies and is a cofounder and member of the Board of Directors of TCSI Corporation and a Director of Coastcom.

Dr. Messerschmitt is a member of the National Academy of Engineering and serves on the Computer Sciences and Telecommunications Board of the National Research Council.