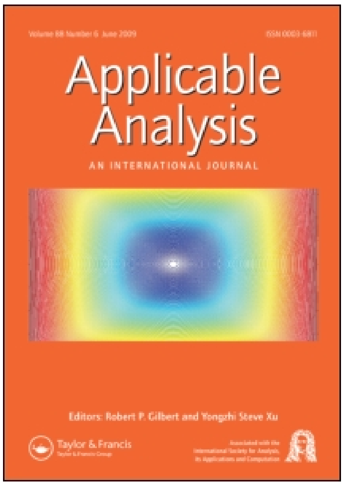

Applicable Analysis
Publication details, including instructions for authors and subscription information:
http://www.informaworld.com/smpp/title~content=t713454076

Multiplex signal transmission and the development of sampling techniques: the work of Herbert Raabe in contrast to that of Claude Shannon

P. L. Butzer[a]; M. M. Dodson[b]; P. J. S. G. Ferreira[c]; J. R. Higgins[d]; O. Lange[e]; P. Seidler[f]; R. L. Stens[a]
[a] Lehrstuhl A für Mathematik, RWTH Aachen University, 52056 Aachen, Germany [b] Department of Mathematics, University of York, York YO1O 5DD, UK [c] Signal Processing Laboratory, DETI/IEETA, Universidade de Aveiro, 3810-193 Aveiro, Portugal [d] I.H.P., 11250 Montclar, France [e] Institut für Technische Informatik, TU Hamburg-Harburg, 21071 Hamburg, Germany [f] Institut für Elektrische Nachrichtentechnik, RWTH Aachen University, 52056 Aachen, Germany

Online publication date: 15 March 2011

To cite this Article Butzer, P. L. , Dodson, M. M. , Ferreira, P. J. S. G. , Higgins, J. R. , Lange, O. , Seidler, P. and Stens, R. L.(2011) 'Multiplex signal transmission and the development of sampling techniques: the work of Herbert Raabe in contrast to that of Claude Shannon', Applicable Analysis, 90: 3, 643 — 688

To link to this Article: DOI: 10.1080/00036811.2010.549474

URL: http://dx.doi.org/10.1080/00036811.2010.549474

Applicable Analysis
Vol. 90, Nos. 3–4, March–April 2011, 643–688

Multiplex signal transmission and the development of sampling techniques: the work of Herbert Raabe in contrast to that of Claude Shannon†

P.L. Butzer[a], M.M. Dodson[b], P.J.S.G. Ferreira[c*], J.R. Higgins[d], O. Lange[e], P. Seidler[f] and R.L. Stens[a]

[a]*Lehrstuhl A für Mathematik, RWTH Aachen University, 52056 Aachen, Germany;* [b]*Department of Mathematics, University of York, York YO1O 5DD, UK;* [c]*Signal Processing Laboratory, DETI/IEETA, Universidade de Aveiro, 3810-193 Aveiro, Portugal;* [d]*I.H.P., 4 rue du Bary, 11250 Montclar, France;* [e]*Institut für Technische Informatik, TU Hamburg–Harburg, 21071 Hamburg, Germany;* [f]*Institut für Elektrische Nachrichtentechnik, RWTH Aachen University, 52056 Aachen, Germany*

Communicated by R.P. Gilbert

(*Received 15 November 2009; final version received 10 December 2010*)

This article discusses the interplay between multiplex signal transmission in telegraphy and telephony, and sampling methods. It emphasizes the works of Herbert Raabe (1909–2004) and Claude Shannon (1916–2001) and the context in which they occurred. Attention is given to the role that the exceptional research atmosphere in Berlin during the 1920s and early 1930s played in the development of some of the ideas underlying these works, first in Germany and then in the USA, as some of the protagonists moved there. Raabe's thesis, published in 1939, describes and analyses a time-division multiplex system for telephony. In order to build his working prototype, Raabe had to develop the theoretical tools he needed and achieved a thorough understanding of sampling, including sampling with pulses of finite duration and sampling of low-pass and band-pass signals. His condition for reconstruction was known as 'Raabe's condition' in the German literature of the time. On the other hand, Shannon's works of 1948 and 1949 contain the classical sampling theorem, but go much further and lay down the abstract theoretical framework that underlies much of the modern digital communications. It is interesting to compare Raabe's very practical approach with Shannon's abstract work: Raabe independently developed his methods to the degree he needed, but his main purpose was

*Corresponding author. Email: pjf@ua.pt

†This article presents a broader version of the invited lecture under the title 'Raabe's work in multiplexing and sampling: Implementing the sampling theorem in the 1930's' held by Paulo Ferreira at the Workshop 'Approximation Theory and Signal Analysis' conducted by Wolfgang zu Castell, Frank Filbir, Rupert Lasser and Jürgen Prestin at Lindau (Lake Constance), 21–24 March 2009. It presents joint work carried out by the seven authors, four mathematicians and three electrical engineers, during a 3-year period, partly in several joint get-togethers in the exceptional Library of the *Institut für Elektrische Nachrichtentechnik* (IENT), RWTH Aachen University (near the Guest House of the University).

ISSN 0003–6811 print/ISSN 1563–504X online

DOI: 10.1080/00036811.2010.549474
http://www.informaworld.com

to build a working prototype. Shannon, on the other hand, approached sampling independently of practical constraints, as part of *information theory* – which became tremendously influential.

Keywords: sampling theorem; multiplexing; Raabe's condition; Nyquist rate; Shannon theory; historical review

AMS Subject Classifications: 01-02; 01A60; 94-03; 94A20

1. Introduction

Multiplex signal transmission had its origin in telegraphy [1,2]. The laying of the telegraphic cables was a difficult and expensive process, and even the fastest telegraphists could not use the line to its full speed. Multiplex systems such as those invented by Baudot and Delany made a more rational use of the cables, by allowing a single line to be shared among a number of telegraphists.

As telephony became increasingly more widespread, a new opportunity for multiplex signal transmission arose. But the simplicity of telegraphy signals, which assume only a few distinct values, is absent from telephony signals. The problem of time-division multiplex in telephony is in fact closely related to a sampling problem – one in which the samples can be instantaneous or have nonzero duration. If one channel is observed through only a fraction of the time, how can the rest of the signal be reconstructed? What is the magnitude of the reconstruction error? How does it depend on the width of the observation window, or the frequency with which the channel is visited, or sampled? In the 1930s there were no good answers to these questions.

The purpose of this article is not only to discuss the role of multiplex signal transmission in the development of sampling methods, emphasizing the works of Herbert Raabe (1909–2004) and Claude Shannon (1916–2001), their similarities and differences, but also the context in which they occurred, some of the main forerunners and the role of the exceptional research atmosphere in Berlin in the 1920s and until the mid-1930s played in the development of some of the underlying ideas.

Early work on the multiplexing of telephony signals, such as that done by Willard Miner, was experimental [3]: the sampling frequency was determined by trial and error. In his patent of 1903, Miner writes [3], [4 p. 7]:

> ... a frequency or rapidity approximating the frequency or average frequency of the finer or more complex vibrations which are characteristic of the voice or of articulate speech, ... , as high as 4320 closures per second, at which rate I find that the voice with all its original timbre and individuality may be successfully reproduced in the receiving instrument...

Raabe's thesis[1] [5], published in 1939, goes further than mere experimentation. The importance of his contribution became known only after the important historical studies by Lüke [3,6–8]. Raabe describes and analyses a time-division multiplex system for telephony, demonstrating a thorough understanding of sampling, including sampling with pulses of finite duration and sampling of low-pass and band-pass signals. Raabe shows that a number of channels carrying telephony signals could be multiplexed and reconstructed with arbitrarily small error, provided that a certain condition be met. The condition, known as 'Raabe's condition' in the German literature of the time, states that the sampling frequency must exceed twice the maximum frequency of the multiplexed signal. The other outstanding characteristic of Raabe's work is that he *implemented the system* – in 1939.

Shannon's works of 1948 and 1949, on the other hand, contain the classical sampling theorem in a precise form but go much further: they lay down the theoretical framework that underlies much of modern digital communications.

It is interesting to compare Raabe's practical engineering approach with Shannon's theory: Raabe independently developed his methods to the degree he needed, his goal being *to build a working prototype*. He gives mathematical proofs of his arguments, using Fourier series as the main tool, but he seems interested in obtaining *a prototype rather than a theory*. Shannon, on the other hand, developed his theory in an abstract, technology-independent way. For several reasons, Raabe's work remained poorly known despite its real importance. On the other hand, Shannon's information theory, as it is well known, was tremendously influential.

2. The research milieu in Berlin 1920–1940

To understand Raabe's and Shannon's work in sampling in the context of information theory and its scientific growth correctly, it is appropriate to look at the exceptional research climate existing not only at the TH Berlin and at Berlin's Friedrich-Wilhelms university but also in Berlin itself and at its international industrial firms with their research centres during this period. As to Raabe, the situation is clear since his thesis of 1939 was written at the *Technische Hochschule Berlin* (TH Berlin), his advisors being Stäblein and Küpfmüller. As to Shannon, it seems that basic ideas of his work can also be traced back to the Berlin atmosphere. First, he cites Bennett[2] [9], who had already cited Raabe. Bennett had joined the communications research department at Bell Labs in 1925; thus, Raabe's work must have been known in the Bell Labs circle. Perhaps, Shannon also knew of the work of Küpfmüller himself as it has many ideas in common with that of Hartley and Nyquist which he cited. Second, Shannon probably became aware of the work of Szilard through von Neumann and possibly also Wiener, as we will now discuss.

In the 1920s and until the mid-1930s Berlin was the capital of modern physics and also a nucleus of electrical engineering. At the university there were giants such as Max Planck, Max von Laue, Walter Nernst, Fritz Haber and James Frank, as well as the younger future Nobel laureates Wolfgang Pauli and Werner Heisenberg, Planck's successor in Berlin in 1926. Berlin University's fame was enhanced by its school of thermodynamics, the world's leading school by that name, which was known for three fundamental principles: the 'energy conservation principle' by H. Helmholtz, the 'principle of entropy' by R. Clausius, and the 'zero entropy condition' at absolute zero temperature by Nernst [10]. Famous scientists who connected with the Berlin school include the American Josiah Willard Gibbs (1839–1939, who spent a year in Berlin in the time of Clausius and Helmholtz), A. Horstmann, C. Carathéodory, who studied there in 1900, and Erwin Schrödinger.

It was at the famous Berlin colloquia, organized by von Laue, where people reported on recent publications from the literature, that young graduate students could communicate with the leaders of physics. Einstein, who was based at the Prussian Academy of Sciences, but had been introduced to the circle of Berlin physicists by Planck and Nernst, was also a regular visitor to the colloquia [11].

It was at these colloquia where, for example, Szilard, Wigner, von Neumann and Gabor, who were actually students at the TH Berlin and not at the university, learned to know the giants of science.

The TH Berlin had as electrical engineers R. Franke (1870–1962), Ernst Max Orlich (1868–1935), K.W. Wagner[3] (1883–1953), who founded the *Heinrich Hertz Institut für Schwingungsforschung* (Oscillation and Vibration Institute) in 1927, H. Fassbender (1884–1970), Karl Küpfmüller (1897–1977), who first worked in 1921 at the renowned Siemens & Halske (becoming a director in 1937), W. Stäblein (1900–1945), W. Cauer (1900–1945) and the high-frequency engineer Helmut Schreyer (1912–1985) who influenced Konrad Zuse (1910–1995) on the design of his calculating machine, the first binary digital computer in the world, the Z3, completed in 1941.[4]

Its physicists included Ferdinand Kurlbaum (1857–1927) and his successor Gustav Hertz (1887–1975), whose aim was to raise the level of physics at the TH to that at the university, Richard Becker (1887–1955), Wilhelm Westphal (1882–1972) and its mathematicians included the geometer Georg Scheffers (1866–1945), Georg Hamel (1877–1954), the well-known student of Klein and Rudolf Rothe (1873–1942). The future Nobel laureates D. Gabor and E. Wigner left Germany in 1933 when Hitler came to power.

Two famous employees at the Siemens & Halske research laboratory must be mentioned, namely the electrical engineer Fritz Lüschen (1877–1945) and the physicist Felix Strecker (1892–1951). The former joined its laboratory in 1920, where also B. Pohlmann (1884–1958), Gabor and R. Feldtkeller (1901–1981) were employed, the latter in 1923, becoming its plenipotentiary in 1935. Strecker discovered the stability criterion usually associated with the name of Nyquist in 1930.

Leo Szilard,[5] who took many courses in engineering at the TH Berlin from 1920 onwards, receiving his *Diplom* there in 1922, wrote his doctoral thesis at the university under von Laue in 1923, with Einstein's encouragement. The thesis [12, pp. 34–102] was related to the Second Law of Thermodynamics. His Habilitation thesis of 1927 [12, pp. 103–29] dealt with Maxwell's Demon [10, 13–15], the famous 1867 *Gedankenexperiment* of Maxwell that would stimulate the investigation of the connections between entropy and information; in this sense, it can be regarded as one of the fundamental contributions to the development of information theory. Szilard's work on the subject, which we will consider in more detail later, is regarded as the earliest known paper in the field of information theory and exposes the connection between the concepts of information and statistical entropy, which Boltzmann had studied 70 years earlier.

Szilard was a personal friend of John von Neumann[6], who came to Berlin University in 1921. Between 1921 and 1923 von Neumann attended, among others, chemistry lectures by Fritz Haber and statistical mechanics by Einstein, who had been asked by Szilard to lecture on the topic. Einstein's audience included, in addition to von Neumann, Szilard, Wigner and Gabor [11]. In 1926, von Neumann became *Privatdozent* in Berlin after receiving his doctorate in mathematics at Budapest. In 1930, Szilard and von Neumann taught together with Schrödinger a theoretical physics seminar.

Szilard's influence may have reached Shannon through von Neumann, who convinced him to call information by the name 'entropy' [16, p. 45]. In this respect, Shannon had met von Neumann already in Princeton. Von Neumann 'pointed out

Boltzmann's observation (...) that entropy is related to *missing information*, inasmuch as it is related to the number of alternatives which remain possible for a physical system after all the macroscopically observable information concerning it has been recorded' [17, p. 156], [18, p. 3]. We will return to some of these points later.

Von Neumann had considered thought experiments similar to the Maxwell's Demon example [19] and was aware of Szilard's explanation of the paradox, which turns around the connection between entropy and information; he would refer to Szilard's work in his research on automata. In fact, Von Neumann himself had introduced, as early as 1927, the quantum mechanical analogue of entropy in [20], more than two decades before Shannon examined in detail its classical limit.

Whereas the capital of physics and engineering at the time was Berlin, that of mathematics was Göttingen. Von Neumann also frequented it, especially because of David Hilbert, the recognized leader in areas of mathematics in which he was interested. Norbert Wiener, who had taken the courses of Hilbert and Landau in 1914, was also a regular visitor to Göttingen. Wiener had come to conclusions similar to those of Szilard in the mid–1920s. In fact, from 1922 to 1927 Wiener travelled to Europe practically every summer where he received considerably more encouragement from mathematicians than he did at home. At his seminar talk at Göttingen in 1924 the 27-year-old Heisenberg was a listener, who was then grappling with the failure of the classical laws in atomic physics as was also Max Born at the time. Wiener again lectured in Göttingen in 1925 when he attended the IMC in Grenoble, as well as in 1926–27 when he gave there a series of lectures with a Guggenheim Fellowship, the second term being at Copenhagen. It seems to have been at Göttingen that Wiener was led to the nexus between communication engineering and statistical mechanics, to the idea – a dream – of a comprehensive quantum theory of entropy, embracing both matter and radiation, in which photons would carry information, and the Second Law of Thermodynamics would become 'rigidly true' (see [21, p. 155–158]).

Wiener's book of 1948 entitled 'Cybernetics', with the subtitle 'Control and communication in the animal and the machine' [22], included a theory for the amount of information in a signal and the transmission of this information through a channel. In his original introduction, written in 1947, Wiener states (p. 10) in this respect:

> To cover this aspect of communication engineering, we had to develop a statistical theory of the amount of information in which the unit amount of information was that transmitted as a single decision between equally probable alternatives. This idea occurred at about the same time to several writers, among them the statistician R.A. Fisher, Dr Shannon of the Bell Telephone Laboratories, and this author. Fisher's motive in studying this subject is to be found in classical statistical theory, and that of Shannon in the problem of coding information; and that of the author in the problem of noise and message in electrical filters. Let it be remarked parenthetically that some of my speculations in this direction attach themselves to the earlier work of Kolmogorov in Russian, although a considerable part of my work was done before my attention was called to the work of the Russian school.

As to the work of Sir Ronald Fisher (1890–1962), Wiener probably had that of 1934/1935 [23,24] in mind (see also [18, p. 95]), and regarding Andrei N. Kolmogorov (1903–1987) it is certainly his paper [25] of 1941 (see also Masani's description of the matter [21, pp. 153–159, 251–261]).

Although Shannon was influenced by Wiener's ideas – Wiener had been Shannon's teacher at MIT in the 1930s – they seem to have had little contact during the years when they worked in communication theory. Speculating, it seems possible that during Wiener's stay in Göttingen 1925 he had heard of Szilard's work or even visited nearby Berlin and that it thus became known in the USA around 1930–1935; the published versions of both theses had appeared in 1925 and 1929, respectively.

Shannon refers to Wiener's influence in *Mathematical Theory of Communication* [26] several times. He points out that the theory is concerned, as Wiener has emphasized, not with operations on particular functions, but with operations on ensembles of functions. And in a footnote he adds that

> Communication theory is heavily indebted to Wiener for much of its basic philosophy and theory. His classic NDRC report, *The Interpolation, Extrapolation and Smoothing of Stationary Time Series* (Wiley, 1949), contains the first clear-cut formulation of communication theory as a statistical problem, the study of operations on time series. This work, although chiefly concerned with the linear prediction and filtering problem, is an important collateral reference in connection with the present paper. We may also refer here to Wiener's Cybernetics (Wiley, 1948), dealing with the general problems of communication and control.

In *Cybernetics*, Wiener acknowledges, among others, McCulloch, Pitts, Turing, von Neumann and Shannon. He emphasizes the importance of the work of Gibbs; later, in the 1950s, he would write

> ... I gradually came to realize the scope of statistics in my work and to apply them not merely to one communication engineering problem, but to all. I was forced to see that the basis of all measurement of information was statistical, and that the frame for it had in fact already been provided by the work of Willard Gibbs.

This sentence appears in a book published posthumously [27]. Wiener went on to add that once 'the public in general had been alerted to the statistical element in communication theory, confirmation began to flow in from all sides'. He referred to the work of Shannon as an example, and added

> I am inclined to believe that from the very start, a large part of [Shannon's] ideas in communication theory and its statistical basis were independent of mine, but whether they were or not, each of us appreciated the significance of the work of the other.

To conclude this section, Szilard, von Neumann, Wiener and Shannon appear to have worked on topics somehow related to entropy and information independently of each other. However, their ideas seem to have influenced each other in subtle ways. The extent of the influence is impossible to determine exactly, but there is ample evidence that the Berlin circle played an important role in the exchange of ideas, and that the exchange continued in the USA as some of the main protagonists, especially von Neumann, moved there.

As we have seen, the works of Szilard and von Neumann on entropy, information and quantum mechanics appeared in the 1920s and originated in the physics community. Interestingly, the first attempts to quantify information in communication systems also appeared in the 1920s, but originated in the electrical engineering community, with Nyquist and Hartley, in the USA, and Küpfmüller, in Germany. The ideas in these works are important to fully appreciate Shannon's work in the USA and Raabe's work in Berlin. The multiplexing problem, to which we now turn, played an important role in the process.

3. The time-division multiplexing problem in telephony

To the engineer, the word 'multiplexing' immediately suggests the simultaneous (or apparently simultaneous) transmission of multiple signals through the same wire or channel, and their recovery at the receiving end. The two simplest ways of achieving it are time division and frequency division multiplexing [28, pp. 279–285, 363–368], [29, Sections 2.9 and 3.5].

In the case of frequency division multiplexing, the frequency band available for transmission is divided into intervals, called sub-bands. Each signal to be transmitted is assigned one of the sub-bands, through a process called modulation. A discussion of some pioneering developments related to frequency-division multiplexing can be found in [30,31].

Time division multiplexing, as the name suggests, is based on time sharing. The period of time available for the transmission is divided in intervals, and each transmitter is assigned one such interval in turn. The duration of each interval is typically very small compared with the duration of the messages, so that to an observer the multiple transmissions appear to be happening simultaneously.

The mathematical model that describes the multiplexing of any finite number of signals is simple. Assuming as an example that there are three channels, the relevant signals would be related as shown in the diagram in Figure 1. The multiplexing system can be regarded as a multi-channel system, with inputs $x_1(t)$, $x_2(t)$ and $x_3(t)$ and one output $m(t)$ given by

$$m(t) = \sum_{i=1}^{3} s_i(t)\, x_i(t).$$

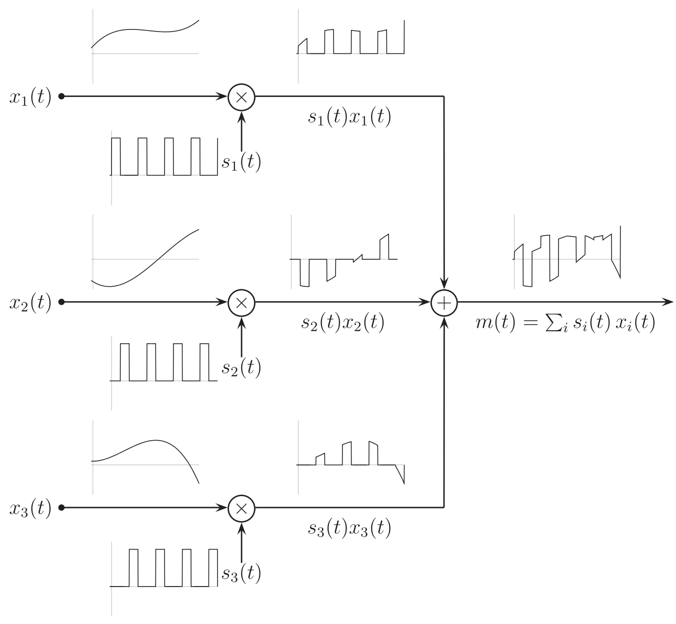

Figure 1. The signals in a three channel time division multiplexing system. The blocks labelled '×' and '+' perform, respectively, multiplication and addition of the indicated inputs.

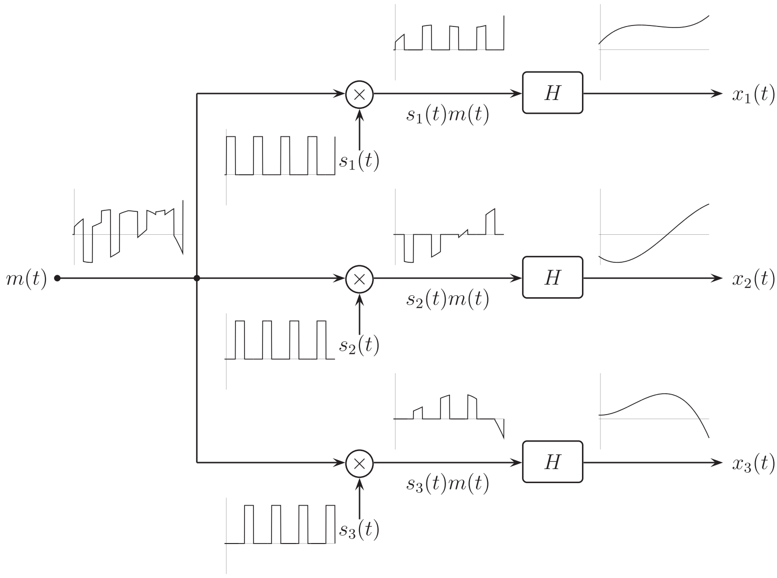

Figure 2. The receiver (or de-multiplexer) of the three channel time division multiplexing system.

The extension to any finite number of channels should be clear.

The multiplexed signal $m(t)$ would be transmitted to the receiver, where the separation of $x_1(t)$, $x_2(t)$ and $x_3(t)$ from the input $m(t)$ takes place. The receiver is illustrated in the diagram of Figure 2. It is again a multi-channel system that takes $m(t)$ as the input and ideally would yield $x_1(t)$, $x_2(t)$ and $x_3(t)$ as outputs. The existence of the operator H, which reconstructs the signals $x_i(t)$ from $s_i(t)m(t)$, cannot be taken for granted due to the apparent loss of information in the process.

The idea of multiplex signal transmission found wide application in telegraphy and was the subject of articles as early as 1883 [32]. The laying of the telegraphic cables was a difficult and expensive process, particularly in the case of transatlantic cables, and not even the fastest telegraphists were able to use them at maximum speed. Multiplex transmission systems, by allowing a single line to be shared among a number of telegraphists, resulted in a more rational use of the cables and attracted the attention of a number of researchers and inventors (Figures 3 and 4), as we shall see in the next section.

4. Raabe's work

Raabe mentions two time-division multiplex systems, associated with the names of Baudot[7] and Delany.[8] The Baudot system followed prior work by F.C. Bakewell (1848), A.V. Newton (1851), M.B. Farmer (1853) and especially B. Meyer[9] (1870). The Delany system built on previous work by P. la Cour[10] and had advantages over the Baudot system. It impressed W.H. Preece, FRS, [33], and was adopted by the British post-office.

Figure 3. The instruments on display at the museum of the *Lehrstuhl und Institut für Nachrichtentechnik* (Institute of Communications Engineering) at RWTH Aachen University give perspective on the evolution of telegraphy (including early telegraph multiplexers mentioned by Raabe, such as Baudot's, shown in Figure 4), the evolution of telephone engineering (including exchange techniques, long distance amplifiers, electro-acoustics, valves and radio receivers) and the evolution of image transmission (including picture telegraphy and early television engineering). The museum provides background information on telecommunications at the time Raabe and Shannon worked.

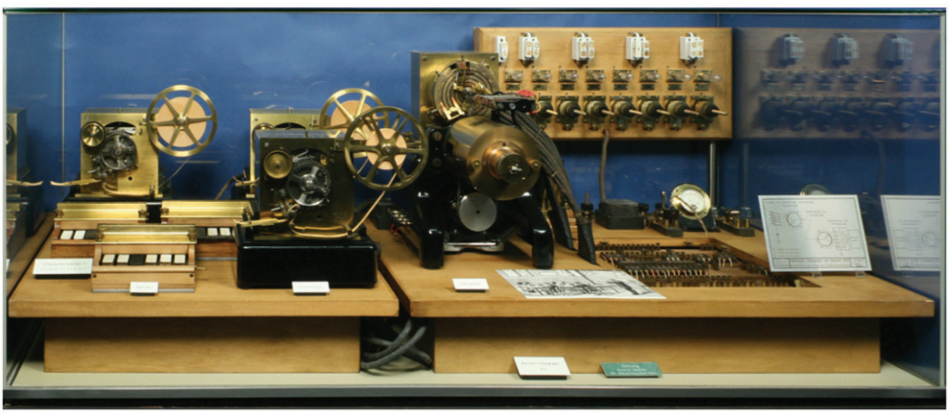

Figure 4. Baudot's time division multiplexer for telegraphy at the museum of the *Lehrstuhl und Institut für Nachrichtentechnik* at RWTH Aachen University (see also Figure 3).

Raabe's opening sentences and first diagram explain time division multiplexing: two synchronized rotating switches connect several transmitters to an equal number of receivers, so that each transmitter is connected to the corresponding receiver throughout a certain interval, in a cyclic way.

As a result of the switching, there is an apparent loss of information that Raabe describes in the following terms:

> (...) those sections of the function which fall within the instants of switching will not be transmitted. The received signal is therefore different from the transmitted one.

Raabe writes that the application of multiplexing systems to telephony requires a 'totally different technical setup'. He provides the necessary theory as well: he gives a condition that guarantees that there is no 'distortion', that is, each of the input (multiplexed) signals can be recovered (theoretically) without any error. The condition is expressed in terms of the spectrum of the signals that are to be multiplexed and in terms of the sampling frequency, that is, the rate at which each particular channel is connected to the multiplexing system – given by the frequency of any of the square waves $s_1(t)$, $s_2(t)$ or $s_3(t)$ in Figures 1 or 2.

The 'sections of the function' mentioned by Raabe could as well be called 'local averages' or 'samples'. They are determined by the product of the function by square waves such as $s_1(t)$, $s_2(t)$ and $s_3(t)$ in our example. In one of the Raabe's examples the pulses are relatively wide, but in others they are so narrow that they appear as 'line needles', as he refers to them.

The recovery of each multiplexed signal implies the recovery of each signal from its 'samples'. Since Raabe is interested in *building* the system, he cannot merely state a condition for transmission without distortion – he has to describe a practical recovery procedure as well. In other words, he determines a practical approximation to the reconstruction operator H in Figure 2.

4.1. *Raabe's condition for distortionless transmission*

To understand Raabe's point of view, it is necessary to obtain the Fourier expansion of the square waves $s(t)$ that will be multiplied by the signals to be multiplexed. Without loss of generality, they can be translated to become an even function. Their Fourier series is then given by

$$s(t) = \frac{k}{2\pi}\, y_m + \sum_{n=1}^{\infty} (-1)^n \frac{2}{n\pi} y_m \sin\frac{nk}{2} \cos n\omega_1 t.$$

Here, y_m is the height of each pulse, ω_1 is the angular sampling frequency and k/ω_1 is the width of the pulse. In electrical engineering, the ratio of the width of the pulse to the period of the wave is usually called the duty cycle. Thus we see that Raabe is considering a square wave of frequency ω_1 and duty cycle $k/2\pi$. His parameter k, the range of which is $(0, 2\pi)$, corresponds to what he calls 'the channel width', 'expressed in terms of phase'.

The received signal will be the product of the square wave $s(t)$ and the signal that needs to be transmitted, denoted by $f(t)$. For Raabe, the product

$$r(t) = s(t) f(t)$$

is 'the modulated signal', the square wave $s(t)$ is 'the unmodulated signal' and the signal $f(t)$ is 'the modulating signal', a terminology borrowed from modulation theory. Raabe is using a square wave instead of the usual sinusoidal carrier, but the amplitude of this square wave is modulated by $f(t)$, as in the standard amplitude modulation scheme.

To understand the effect of the modulating signal $f(t)$ on the square wave $s(t)$, Raabe also expands $f(t)$ in a Fourier series, and investigates the effect of the general term of this Fourier series on each frequency of the square wave. So the problem is reduced to that of understanding products of square waves and sinusoids. The result for a more general $f(t)$ follows by superposition. The square wave had been decomposed in a Fourier series as well. As Raabe writes

> (...) The modulation of a carrier representable as a Fourier series is equivalent to modulating each single frequency of the whole spectrum on its own.

Raabe's next step is to assume that the signal to transmit is simply $\cos(m\omega_1 t)$, where m is a real number. Multiplication of this signal by the Fourier series of the square wave $s(t)$ yields the received signal $r(t)$:

$$r(t) = \frac{k}{2\pi} y_m \cos m\omega_1 t + \sum_{n=1}^{\infty} (-1)^n \frac{1}{n\pi} y_m \sin \frac{nk}{2} [\cos(n-m)\omega_1 t + \cos(n+m)\omega_1 t]. \tag{A}$$

He notes that the received signal $r(t)$ includes the transmitted signal, which has replaced the 'direct voltage term', namely

$$\frac{k}{2\pi} y_m \cos m\omega_1 t.$$

We will now consider the remaining terms, following Raabe's reasoning.

Essentially, Raabe argues that to avoid ambiguity m has to satisfy $m < 1 - m$, that is, $m < 1/2$. In other words, the transmitted frequency must be below $\omega_1/2$, that is, one half of the sampling frequency. The condition guarantees that the spectral line m (the one that replaced the 'direct voltage term') stays to the left of the 'complementary frequency' $1 - m$, which in the context of amplitude modulation is called the 'lower sideband of the first harmonic'.

For signals consisting of a superposition of terms with different frequencies, the distortion can be avoided if *all the frequencies* are below one half of the sampling rate. This is how Raabe puts it:

> (...) the sampling frequency is determined by the range of signal frequencies. If these are kept below half of the sampling frequency, all of the noise frequencies above this limit can be kept away from the receiver by a low-pass filter. The transmission of a signal may thus be completely distortionless, if the sampling frequency is twice the highest signal frequency. The upper limitation of the signal frequencies is therefore a vital condition of distortionless transmission by time division multiplex transmission.

This paragraph, also quoted by Lüke [3], summarizes what has been called 'Raabe's condition'. The argument is based on the analysis of the Fourier spectrum of the product of the modulating square wave and the signal. The conclusion has a legitimate mathematical basis, valid for the class of signals Raabe considers.

4.2. *Raabe's condition in the band-pass case*

Raabe realizes that the reconstruction of the sampled signal can be accomplished by a 'low-pass filter', but he notes that sampling at 'twice the highest signal frequency' is not always necessary. For band-pass signals, Raabe states, 'special conditions apply when the frequency range is limited to an octave', in which case

> (...) one can lay down the sampling frequency onto the upper frequency limit, and the noise frequencies remain partly above the useful range, and partly, as in the case of the complementary frequency, below.

The value 'one octave' is of course correct, as discussed in [34]. For yet another perspective, note that the set $[-2a, -a] \cup [a, 2a]$, which is similar to the support of the Fourier transform of a real signal with a bandwidth of one octave, satisfies the disjoint translates condition for translates by $2a$. Thus, a spectrum with this type of support can be repeated every $2a$ to give a $2a$-periodic function. As Raabe writes, this means that the sampling rate $2a$ will work for this signal.

Raabe also knows that 'frequencies above the sampling frequency can also be transmitted free from distortion, but the relative range of frequencies shrinks more and more'. As is well known today, the minimum sampling density is determined by the bandwidth of the signal and not necessarily by its highest frequency.

Raabe also mentions the recovery procedure in the band-pass case:

> (...) one has to keep away from the receiver those noise frequencies lying outside of each section by a band pass filter.

From his perspective of circuit design, the only change necessary is the replacement of the low-pass filter by a band-pass filter.

4.3. *Admissible functions*

Raabe's main mathematical tool is the Fourier series. The square wave that the multiplexing system multiplies each signal by is expanded in a Fourier series, the fundamental frequency of which is the sampling frequency. The signal to be transmitted is also expanded in a Fourier series. The effect of the multiplication of the Fourier series is analysed, term by term, to obtain the main results. Raabe's conclusions apply to signals that can be expressed as a linear combination of sinusoidal terms (or complex exponentials), the frequencies of which do not exceed one half of the sampling frequency. The admissible functions include all band-limited periodic functions or trigonometric polynomials

$$f(t) = \sum_{k=-M}^{N} c_k e^{2\pi i k t/T}$$

but not band-limited square-integrable (finite-energy) signals. However, for the purposes of *using* Raabe's time division multiplexing system this is not a restriction. The system would be able to handle signals of arbitrarily large but finite duration.

Raabe's practical perspective is reflected in the way he deals with the reconstruction problem, which is essentially solved if the aliased terms or 'noise frequencies' are separated from the signal frequencies by a linear filter.

This practical perspective may help to understand why Raabe did not explicitly state a reconstruction formula based on the ideal low-pass filter, introduced by Küpfmüller and known as 'Küpfmüller-Tiefpass' in the German literature.

Instead of obtaining a theoretical reconstruction formula, for which he probably felt very little need, Raabe obtains a real reconstructed *signal*.

4.4. *Raabe and generalised sampling*

It is worthwhile to discuss the connection between the system that Raabe built and generalized sampling.[11] Let $f(t)$ be one of the input signals. The corresponding sampled signal is obtained by multiplying $f(t)$ by a square wave $s(t)$, which can be expressed as a sum of pulses $p(t)$. Hence, assuming for simplicity a unitary sampling period, we have

$$f(t)\,s(t) = f(t)\sum_n p(t-n).$$

Assume that the 'channel width' is small, so that the square waves look like 'line needles'. Then the sampled signal is *approximately* given by

$$g(t) = \sum_n f(n)\,p(t-n).$$

Let the response of the lowpass filter implemented by Raabe to a pulse $p(t)$ be denoted by $h(t)$. The reconstructed signal obtained using Raabe's system can then be written as

$$\sum_n f(n)h(t-n).$$

There are two main differences to the classical sampling theorem: the sinc kernel, the Fourier transform of the ideal lowpass filter,[12]

$$\mathrm{sinc}(t) := \frac{\sin \pi t}{\pi t},$$

has been replaced by the function $h(t)$, the pulse response of the filter used in the implementation; and there are only finitely many samples $f(n)$. Both differences would necessarily occur in any practical realization of sampling.

Although the sampling theorem is not to be found explicitly in Raabe's paper, it is a reasonable inference that in view of Raabe's approach and results he was aware of more general results, but did not state anything in further generality because his interests were in more practical directions.

4.5. *Raabe's condition for distorted transmission*

In order to implement his technical innovations in practice, Raabe also investigated the distortions which can arise during the transmission of square waves over a real transmission line.[13] To do this he first studied the possibility of a frequency limitation through real transmission lines on the basis of their inherent lowpass character. He shows that the cut-off of the carrier spectrum in general leads to a crosstalk across the neighbouring channels. If one neglects all frequency terms above a certain limit $n = p$ in the receiver function, one obtains a truncation of the square

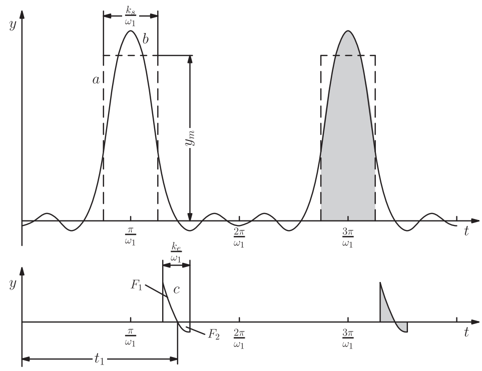

Figure 5. Because of the frequency limitation, the graph of the function a, shown in the figure, is changed to that of the function b. The section of the function b now represented by c is received by the receiver distributor of a neighbouring channel. The two shaded areas in the upper and lower diagrams represent I_0 on the one hand and I_d on the other.

wave function, namely

$$s_p(t) = \frac{k_s}{2\pi} y_m + \sum_{n=1}^{p} (-1)^n \frac{2y_m}{n\pi} \sin\left(\frac{nk_s}{2}\right) \cos(n\omega_1 t),$$

the approximation of which improves with increasing p (Figure 5).

Raabe's investigations therefore concentrate on the necessary frequency range $p\omega_1/2\pi$ in order to reach a transmission with the highest possible crosstalk attenuation. His procedure is as follows: first, he considers that part of $s_p(t)$ which lies on the outside of the channel width k_s/ω_1, and which extends across the neighbouring channels. With a Fourier series approach together with a subsequent formation of the associated average, he obtains the distorted part I_d in a width of k_e/ω_1, namely

$$I_d = \frac{k_s k_e}{4\pi^2} y_m + \sum_{n=1}^{p} (-1)^n \frac{2y_m}{n^2\pi^2} \sin\left(\frac{nk_s}{2}\right) \sin\left(\frac{nk_e}{2}\right) \cos(n\omega_1 t_1).$$

Likewise he determines the average value of $s_p(t)$ within the channel width k_s/ω_1, namely

$$I_0 = \frac{k_s k_e}{4\pi^2} y_m + \sum_{n=1}^{p} \frac{2y_m}{n^2\pi^2} \sin\left(\frac{nk_s}{2}\right) \sin\left(\frac{nk_e}{2}\right);$$

and finally the crosstalk attenuation $\vartheta = \ln|I_0/I_d|$. This attenuation should be large and I_d therefore small, and the values k_s, k_e, t_1 and p should be chosen correspondingly. If simultaneously the crosstalk attenuation of all neighbouring channels should be large, then the analysis in this general case is hopeless. Nevertheless, to obtain a clear picture, Raabe chooses a four-channel and a two-channel system as concrete examples. For these he determines the channel width, the phase position of the interfering channel and the neighbouring channels as well as the necessary frequency ranges under the régime of high crosstalk attenuation.

Finally, Raabe treats the influence of the transmission distortions on cross-talking. The ratio of a complex input voltage U_1 and the output voltage U_2 of a transmission line of length l terminated by the characteristic wave impedance is given by $U_1/U_2 = e^{\gamma l}$, with the transmission rate $\gamma = \beta + j\alpha$. The attenuation rate β and phase rate α are frequency dependent and therefore cause distortions. If now $s_1(t)$ is the input voltage, then the output voltage becomes

$$s_2(t) = \frac{k_s}{2\pi} y_m e^{-\beta_0 l} + \sum_{n=1}^{\infty} (-1)^n \frac{2y_m}{n\pi} e^{-\beta_n l} \sin\left(\frac{nk_s}{2}\right) \cos(n\omega_1 t - \alpha_n l).$$

Raabe now applies the above approach to evaluate the corresponding I_d and I_0, from which he determines the crosstalk attenuation $\vartheta = \ln|I_0/I_d|$. Since the evaluation of this result is extremely complicated, Raabe presents an alternative approach for determining the crosstalk attenuation. He represents $s_1(t)$ as a periodic sequence of superposed unit step functions. These functions degenerate at the end of the transmission line due to distortions. The output function can be interpreted as a periodic sequence of superposed distorted step functions. In the course of his analytical treatment, he obtains a series for which he just needs to determine a few terms. In this way he arrives at a distorted carrier and finally the crosstalk attenuation $\vartheta = \ln|I_0/I_d|$.

5. Shannon's sampling theorem in the frame of his information theory

It is difficult to determine when Shannon began to develop his theory of communication. In a letter of February 1939 to Vannevar Bush he still represents the 'general systems for the transmission of intelligence, including telephony, radio, television, telegraphy, etc.' as a transformation of time-functions. At that time he dealt with continuous frequency functions; discrete transmission of information arose as a result of technical trends and cryptography during war research. In June 1941 in a letter to Dean Eisenhart of Princeton University, he already reports that he is working on a 'general theory of transmission and transformation of information'; it is this theory for which he became famous.

Shannon may have began working on information theory as early as 1940 when he was a National Research Fellow at Princeton. In an interview recorded by F.W. Hagemeyer,[14] Shannon said his ideas concerning information theory were mostly developed around 1943–1945, but were not published until 1948 because 'information theory was not considered first priority work' during the war. He was grateful to Bell Labs for *tolerating* his work, which ironically seemed of no practical interest to AT&T [35]. It was cryptography that allowed him to work on information theory. Pressed as to what gave him the basic idea to his theory, he said it was Hartley's paper of 1928 on the transmission of information [36].

One should note, however, that in his 1982 interview of Shannon, 'Price tried hard to get into the mind of the grey haired man sitting next to him' [16] – but Shannon resisted. He seemed hesitant in answering 'these complex hypothetical questions', as he called them, or gave answers such as 'I have no idea'. More recently, the analysis of an unpublished manuscript of Shannon has raised some further doubts: Shannon could have been working on the details of his information theory [37] as late as 1948, shortly before its publication [38].

This landmark work is singular: it signals the birth of a discipline, yet it also contains some of its fundamental results. Initially, it was not unanimously well received: engineers found it too mathematical, and mathematicians criticized its lack of rigour. A review [39] asserted that 'the discussion is suggestive throughout, rather than mathematical, and it is not always clear that the author's mathematical intentions are honorable'. The view would rapidly change, however.

Only two years after the publication of his theory, Shannon was presenting a summary of its main results under the title 'Communication theory, exposition of fundamentals' at the Symposium on Information Theory held at the Royal Society, London, 26–29 September 1950 (see W. Jackson, Ed., Proceedings of the Symposium on Information Theory, Royal Society, London, Ministry of Supply, 1951). The proceedings were also published in 1953 in the first issue of IRE Transactions on Information Theory (Shannon contribution, also reproduced in [40, pp. 173–176], is [41]). This was the only time Shannon presented a summary of his theory in front of an international audience (over 130 participants from at least eight countries) including experts such as J. Loeb, D. Indjoudjian, P. Aigrain, J. Oswald, J. Ville, P. Chavasse, S. Colombo, Y. Delbrod, J. Icole, P. Marcou and E. Picault [42]. Information theory was spreading quickly.

In the Soviet Union, Andrei Kolmogorov (1903–1987) wrote [43]:

> The significance of Shannon's work for pure mathematics was not fully appreciated at the outset. I remember that even at the International Congress of Mathematicians held in Amsterdam in 1954 my American colleagues, specialists in probability, believed that my interest in Shannon' works was somewhat exaggerated since techniques prevailed over mathematics in them. Nowadays, such opinions hardly need a refutation. It is true that in situations of any degree of complexity Shannon left the strict mathematical 'validation' of his ideas to his followers. However, his mathematical intuition is amazingly correct...

Shannon's work was called 'incomparably deep' by Kolmogorov, who recognized its connections with ergodic theory (see also [44]) and started a seminar at Moscow University in the early 1950s to explore the mathematical foundations and implications of information theory [45]. Kolmogorov reported some of the results already in 1956 in the IRE Transactions on Information Theory [46]. He would retain his interest in the foundations of information theory, probability and computational complexity, and return to it as late as 1968 from a very different angle [47]. A survey of the Soviet research on information theory can be found in [48].

It is pointless to elaborate on the impact of Shannon's information theory, which, as Slepian wrote in the introduction to [49], altered man's understanding of communication as perhaps no other work in the twentieth century. Even its terminology became standard: Shannon called his measure of information 'entropy', maybe as we have seen at the suggestion of von Neumann; he used the term 'bit', which he attributed to J.W. Tukey (who had in fact proposed it in 1946 [50]); and he introduced the expression 'Nyquist rate' as a tribute to Nyquist. All these terms have stood the test of time.

Shannon introduces band-limited ensembles of functions and uses the sampling theorem, which he regards as a sum of orthogonal functions, to map a band-limited $f(t)$ to a vector of samples in an infinite-dimensional space. For a proof of the sampling theorem he refers to his paper of 1949 [26], which had been submitted in 1940. Shannon writes that a function is limited to a time T if all samples outside that interval of time are zero. Functions band-limited to W and time-limited to T can be represented by $2WT$ coordinates.

The sampling theorem is a crucial tool in Shannon's theory since it provides a way of defining an ensemble of band-limited and time-limited functions by means of a probability distribution $p(x_1, x_2, \ldots, x_n)$ in n-dimensional space. Note, however, that time-limitation and frequency-limitation are both essential. If the time-limitation condition is not satisfied, Shannon suggests the consideration of $2WT$ coordinates that correspond to an interval of duration T 'to represent substantially the part of the function in the interval T and the probability distribution $p(x_1, x_2, \ldots, x_n)$ to give the statistical structure of the ensemble for intervals of that duration'. From this point onwards Shannon makes no further direct references to the sampling theorem – but the bridge between the continuous and the discrete domains that it has established will be repeatedly used.

Shannon's work was rapidly followed up and it became very popular. In fact, it became *too popular* for Shannon's own taste: by 1956, he felt the need to write the editorial 'The Bandwagon' [51], in which he argues that maybe information theory had been 'ballooned to an importance beyond its actual accomplishments'. He insists that information theory, as a branch of mathematics, is a deductive system; that a thorough understanding of its mathematical foundations and applications in communication are a prerequisite to other applications; and that the 'house must be kept in first-class order' since the subject had been 'oversold'. Shannon suggests that attention should be directed to raising the publication standards: 'a few first rate research papers are preferable to a large number that are poorly conceived or half-finished' – good advice!

But let us first recall a few facts from Shannon's life, and then turn into the history.

5.1. *A brief biography of Shannon*

Claude Elwood Shannon was born on 30 April 1916 in Petoskey, Michigan. Through his father he was a distant cousin of Thomas Edison, both being descendants of the early colonial pilgrim John Ogden. His mother, Mabel Wolf, daughter of a German immigrant, was a language teacher and Principal of Gaylord High School for some years.

Shannon graduated from that school in 1932, and then entered the University of Michigan, receiving his Bachelor of Science degree both in Electrical Engineering and Mathematics in 1936. This dual interest continued throughout his career. As a boy he constructed model planes, a radio controlled boat and a telegraph system to a neighbour.

Although he was not outstanding in mathematics, he went to MIT, as a research assistant in Electrical Engineering, where he studied with both Norbert Wiener and Vannevar Bush. The latter had built an analogue computer called the Differential Analyzer. Shannon was fascinated with the 'Laws of Thought' of George Boole, which he had studied at Michigan; in his Master's thesis [52] 'A symbolic analysis of relay and switching circuits' (issued 1940 but worked out in 1937) he was able to show with the help of Boolean algebra how logical symbols could be treated as a series of on or off switches, and how binary arithmetic – manipulation of strings of 0s and 1s – could be carried out by electrical circuits. The work, which led to a paper published in 1938 in the AIEE Transactions [53], earned Shannon the 1939 Alfred Noble Prize of the combined engineering societies of the USA. The thesis has been called 'one of the most important master's theses ever written' [16] and essentially founded digital circuit design.

In September 1938 Bush suggested to Shannon that he should change from the Engineering to the Mathematics Department at MIT and that algebra might be as useful in organizing genetic knowledge as it was in switching. Thereupon Shannon wrote his doctoral thesis 'An algebra for theoretical genetics' [54]; the genetics he learned from Dr Barbara Burks, of Cold Spring Harbor, NY; the supervisor of the thesis at MIT was the algebraist Frank L. Hitchcock, who had already supervized Shannon's Master's work. See [55] for some comments about Shannon's work in genetics and the almost simultaneous and related work by Cotterman and Malécot. Shannon received MS degree in Electrical Engineering and his PhD in Mathematics at the same commencement, in the Spring of 1940.

He spent the (academic) year 1940–1941 at the Institute for Advanced Study, working under Hermann Weyl. It is here that he began to work seriously on his ideas relating to information theory and efficient communication systems. As we have seen, he had written already in February 1939 to Bush about these ideas. He spent 15 years with Bell Labs, first working on anti-aircraft directors – devices to observe enemy planes or missiles and calculate the aiming of counter missiles. Some of the foremost scientists and mathematicians then at Bell Labs were: T.C. Fry, head of its mathematics department, John Pierce, known for satellite communication, Harry Nyquist, well known for his work in telegraphy and signal theory, Hendrik Bode of feedback fame, Bardeen, Brattain and Shockley – the transistor inventors. Slepian and George Stibitz, with his relay computer of 1938, were also there.

At Bell Labs Shannon worked in information theory, as explained, which would lead to his seminal works on the subject: 'Communication in the Presence of Noise' and 'A Mathematical Theory of Communication' [26,37]. Shannon focuses on the problem of how to reliably reconstruct at a receiving point the information a sender has transmitted. In yet another notable paper of 1949, 'Communication Theory of Secrecy Systems' [56], he gave essential results in the mathematical theory of cryptography.

From 1957 to 1978 Shannon was Donner Professor of Science at MIT. He was known as 'the Irascible Genius'. He worked alone, although he had a chance to meet some of the brightest scientists and engineers of his time. He preferred to work out everything in his head, instead of on paper, and it has been claimed that he would write entire academic papers by dictating from memory alone, without correction [35,57]. Colleagues accused him of not being sufficiently rigorous.

He held honorary degrees from Yale, Michigan, Princeton, Edinburgh, Pittsburgh, Northwestern, Oxford, East Anglia, Carnegie-Mellon, Tufts and University of Pennsylvania. In addition to the already mentioned Alfred Noble Prize, Shannon received at least a dozen other awards, including the Morris Liebmann award of the Institute of Radio Engineers (1949), the Stuart Ballantine Medal of the Franklin Institute (1955), the Medal of Honor of the IEEE and the National Medal of Science (1966), the Kyoto Prize (1985) and the Eduard Rhein Award[15] (1991) (Figures 6 and 7). He was afflicted with Alzheimer's disease and died in 2001.

Figure 6. Claude E. Shannon (1916–2001) addressing the members of the Eduard Rhein Foundation and invited guests in October 1991 at the German Federal Guest House Petersberg (near Bonn), where he received the Eduard Rhein Award.

Figure 7. Shannon between his wife Betty and Hans-Dieter Lüke (1935–2005), signing the document of the Award, with the signature shown. Both photos were kindly supplied by Mrs. Bernhardine Lüke.

5.2. *Thoughts on the early history of information theory*

The rapid replacement of optical telegraphy [2] and its expensive towers, devices, telescopes and observers with electromagnetic telegraphy, which was capable of operating almost independently of the weather conditions and at much higher transmission rates, set the stage that would ultimately lead to modern information theory.

The first telegraphy experiments by Gauss and Weber in 1833 [2,58] were the work of scientists, not entrepreneurs. Both men lacked the means and the motivation to develop the idea further; they do not seem to have had any intention of exploring it for commercial purposes. For a while steps were under way to use the new invention on the railroad, but the project was soon abandoned for financial reasons. Thus, as Dunnington wrote, 'Germany lost the honor of being the first to produce a practical telegraph' [58, p. 150].

Meanwhile, in the United States, the Morse-Vail method of telegraphy and its efficient code of 1837 were quickly gaining acceptance. By 1844, a line was already connecting Baltimore and Washington, DC – and the work of Gauss and Weber in Europe was being for all practical purposes forgotten. New problems and challenges were appearing and demanded attention, from transatlantic cables to multiplex systems. Although they would lead to a lot of practical work and many patents, they would also lead to theoretical work of great importance.

In a sense, information theory grew out of this effort to master specific problems in telegraphy and then increasingly more general and abstract problems of communication. It is impossible to understand the work of Raabe and Shannon without putting it into perspective – and to do that one must at least briefly sketch their connection with the work of Hartley and Nyquist in the USA and the work of Karl Küpfmüller, Fritz Lüschen and Felix Strecker in Germany. These names are also firmly associated with the birth of information theory.

We begin with Harry Nyquist,[16] whose fundamental work on telegraphy was published in 1924 and 1928 [59,60]. In these papers, Nyquist shows a thorough understanding of the connection between signalling speed, number of bits per symbol and bandwidth. He also fully understood intersymbol interference and how to avoid it. These are landmark achievements, for which he is justly remembered.

He had started to work on the subject much earlier, as he himself writes in a memorandum of 1934: 'At the end of 1917, I was transferred to current work on telegraph developments and later signalling work'. In his 1924 paper 'Certain factors affecting telegraph speed' [59], Nyquist argued that the transmission rate of a telegraph system is proportional to the logarithm of the 'number of current values', that is, the number of signal levels. He gives the formula $W = K \log m$ where K is a constant and m is the number of distinct values of the current. The proportionality to the bandwidth is one of the results of his second, more general paper on the subject, 'Certain topics in telegraph transmission theory', published in 1928 [60]. In that paper Nyquist was especially interested in the 'maximum speed of transmission of intelligence'.

In 1928, Ralph V.L. Hartley,[17] another engineer, published his 'Transmission of Information' [36], also in the *Bell System Technical Journal* (not citing Nyquist or anyone else). Hartley, whose first paper on electrical communication had appeared in 1918 [61, p. 211], had performed research on voice and carrier transmission and was also interested in picture transmission and television. For him, information was 'a very elastic term', and his first task was 'to set up for it a more specific meaning as applied to the present discussion'. He reached the conclusion that Nyquist had reached in his 1924 paper: information was proportional to the logarithm of the number of possible messages. Unlike Nyquist, who spoke of 'intelligence', Hartley used the term 'information'. Hartley's law of 1928 states 'that the total amount of information that can be transmitted is proportional to the frequency range transmitted and the time of the transmission'. In other words, the information content is proportional to a product of bandwidth and time.[18]

Hartley not only considered telegraphy-like signals but also continuous-time signals, similar to those that appear in telephony. To handle the latter, he approximated them using step functions (see Figure 3 of [36, p. 543] and the discussion in [61, p. 235]). The step-width was determined by the frequency, the step-height by the magnitude of the intersymbol interference.

Meanwhile, in Germany, Karl Küpfmüller,[19] in his paper 'Über Einschwingvorgänge in Wellenfiltern' [62] of 1924, which appeared slightly earlier than Nyquist's paper of 1924, studied for the first time the maximum telegraph signalling speed sustainable by bandlimited linear systems. In fact, Küpfmüller discussed a relation between bandwidth and time similar to that of Nyquist. The result of the two independent studies was the so-called Nyquist-Küpfmüller law, stating that in order to transmit telegraph signals at a given rate a certain definite frequency bandwidth was required. This provoked vehement protest from Küpfmüller's colleagues and also those of Nyquist, even after the latter had produced the theoretically precise results in his paper of 1928 [60].

One should note that Küpfmüller, in his paper 'Ausgleichsvorgänge in der Telegraphen-und Telefontechnik' of 1931, which appeared only in a Swedish version, was aware of the papers of Hartley and Nyquist (but did not cite them). He established a version of Hartley's theorem of 1928, his 'Zeitgesetz der Telefonie und Telegrafie'.

Küpfmüller's paper [63] contains one of the first discussions of stability in the context of closed-loop systems [64] (see also the footnote on Felix Strecker). Küpfmüller appears to have been the first researcher to use abstract, idealized linear systems (characterized by input–output relationships described in the time or frequency domains).

The landscape of scientific discovery at this period is complex, but the work of Hartley, Nyquist and Küpfmüller is without doubt important and would have lasting influence – in information theory and in sampling.

Fritz Lüschen[20] in his paper 'Moderne Nachrichtenysteme' of 1932 [65] first referred to both Hartley's law and Küpfmüller's 'Zeitgesetz' as the basic papers in communication theory of the time [61, p. 251]. Lüschen presented this paper to the IEE at their meeting in London in 8 April 1932, and probably also made Hartley's results first known in Britain. Hartley and Küpfmüller met at the International Congress of Telegraphy and Telephony held at Lake Como (Italy) in honor of Alessandro Volta (1745–1827) in September 1927 (where the physicists also held their congress, with Bohr, Bragg, Frank, Gerlach, Rutherford, Millikan, Zeeman). With the publication of Shannon's papers of 1948/1949, references to the 'law of Hartley' stopped, as did references to Küpfmüller.

In 1935, Felix Strecker[21] was interested in the application of Hartley's results to telephony and concluded 'that in a telephone system the effective bandwidth is at most equal to that bandwidth which is equivalent to the actual one'. In other words, any manipulation in the time and frequency distribution of telephone conversations cannot reduce the bandwidth absolutely necessary for comprehensibility. In connection with Strecker's conclusion, one should note that in the USA, Carson had written in 1922 that [66]

> ... a great deal of inventive thought has been devoted to the problem of narrowing the band of transmission frequencies. Some of the schemes which are directed to this end are very ingenious; all, however, are believed to involve a fundamental fallacy.

Strecker's conclusion confirms Carson's belief.

Let us also mention the 'Theory of Communication' presented by the physicist Dennis Gabor[22] in 1946 [67]. Gabor is one of the several remarkable Hungarian-born scientists who studied in Berlin for some time and then left Germany. Some other names are Theodore von Kármán (1881–1963), Eugene Paul Wigner (1902–1995), Edward Teller (1908–2003) and of course the already mentioned Leo Szilard and John von Neumann [11].

Gabor defined a 'quantum of information', which he called a 'logon', in terms of the product of uncertainties of time and frequency of an electrical signal, a concept he used to analyse waveforms in communication systems. Gabor's theory is non-statistical in nature. He refers to the Lüschen paper [65] of 1932; both had worked at Siemens in 1932.

Now turning to the statistical aspects of Shannon's theory, we recall that the connection between entropy as it was known in statistical physics and Shannon's entropy in communication had been realized, among others, by Szilard and von Neumann, who had pointed it to Shannon [16, p. 45], [18, p. 3], [68]. According to Lanouette [15], Szilard 'saw the key elements of information theory some three decades before it became popular'. Müller [69] argues that Szilard was one of the first, if not the first, to recognize that information could be converted into 'negative entropy'. In fact, by 1927–1929 [12, pp. 103–129], [70] Szilard had explained the relation of physical entropy to information and 'successfully exorcized' Maxwell's Demon. For a detailed discussion, see [71, Chapter 13] and [72, Section 8.4], or [73,74].

Still concerning the concept of entropy, it is worth noting that its quantum mechanical version (the von Neumann entropy) was introduced more than two decades before the classical limit was discussed at length by Shannon in the context of information theory: Von Neumann first considered it in an article [20] of 1927, in which he associates an entropy operator with a statistical operator (see also [75]).

The fact that both von Neumann and Szilard considered the concept of entropy, although in very different contexts and for different purposes, is not totally unexpected given that both were Hungarian-born, frequented the Berlin circle and had some common interests. In Berlin both attended a statistical physics course by Einstein, who had been persuaded by Szilard to lecture on the topic. Later, in 1930, they taught together with Schrödinger a theoretical physics seminar. It is a fact that von Neumann knew Szilard's work very well; he mentioned it in lectures (an account can be found in [76]) and in a review of Wiener's *Cybernetics*. In the late 1940s, when he became interested in automata, he investigated the extent to which reliable systems could be built out of unreliable components by exploring the role of redundancy and error-correcting codes; he mentioned not only Szilard and Shannon, but also Nyquist, Hartley and others.

As Shannon was completing his *Mathematical Theory of Communication*, William Tuller (born 1918) was working on his MIT PhD thesis [77] (1948) on the limits of information transmission in the presence of noise. William Tuller regards his thesis as a follow-up of Hartley's work, and places his work on the side of Carson, Nyquist, Küpfmüller and Gabor, whose works he cites. He can be regarded as one of the founders, with Shannon, of the mathematical theory of communication.

Tuller's work comes closer to Shannon than any of his predecessors in that he considers the effects of noise. Parts of Tuller's thesis were published in 1949 as an article [78]. Tuller points out that the previous workers in the field 'failed to include noise in their reasoning' and mentions Wiener's work on prediction and filtering of stationary time series as well as its similarity with the problem of information transmission. Tuller shows that the quantity of information H that may be transmitted over a given circuit satisfies

$$H \leq 2BT \log(1 + C/N),$$

where B is the transmission link bandwidth, T the time of transmission and C/N the carrier-to-noise ratio. In a footnote, he writes that he became aware of Shannon's work only in the spring of 1946, after completing the basic work underlying his paper. He adds that he remained unaware of the details of Shannon's theories until the summer of 1948, eight months after completing his work.

When Shannon considers the capacity with an average power limitation P and obtains the famous formula [37]

$$C = W \log_2\left(1 + \frac{P}{N}\right),$$

which he had already discussed in [26], he mentions that Wiener, Tuller and Sullivan had also obtained similar results 'although with somewhat different interpretations'. He gives no specific reference to H. Sullivan but cites Wiener's book on cybernetics [22] and Tuller's thesis [77].

Now turning to mathematics and the sampling theorem itself or its variants, it is well known that – unknown to the mentioned authors – Whittaker [79] had already found in 1915 how to interpolate the sampled values of bandlimited functions (see also [80] for a discussion of contributions by Hardy and other authors and several other historical notes). In fact, the sampling theorem with the sinc function for functions of finite duration was first established by the mathematician de la Vallée Poussin [81] in 1908; for a discussion of his contribution and its influence, see [82]. The sampling expansion for trigonometric polynomials is in fact much older, being due to Euler, who discussed it more than once, as had Cauchy. See [83] for details about the method used by Euler to derive expansions such as

$$\cos vt = \frac{\sin \pi t}{\pi}\left\{\frac{1}{t} + 2t\sum_{n=1}^{\infty} \cos vn \frac{(-1)^n}{t^2 - n^2}\right\}, \quad (|v| \leq \pi)$$

and other related historical information.

In Russia, the engineer Vladimir Kotel'nikov[23] had formulated the sampling theorem already in 1933 [84] in an interesting direct and converse form:

> Any function $F(t)$ which consists of frequencies between 0 and f_1 is representable by the series
>
> $$F(t) = \sum_{k=-\infty}^{\infty} D_k \frac{\sin \omega_1 \left(t - \frac{k}{2f_1} \right)}{t - \frac{k}{2f_1}}$$
>
> where k is an integer number, $\omega_1 = 2\pi f_1$ and D_k are constants depending on $F(t)$. Conversely, any function $F(t)$ represented by this series consists only of frequencies between 0 and f_1.

Kotel'nikov's famous paper, prepared for a conference never held, did not appear in an internationally accessible form until the publication in 2001 of the English translation by V.E. Katsnelson [85, pp. 27–45]. It became known in the West only much after its publication, possibly through two works of Kolmogorov and Tichomirow[24] published in 1956 and 1960 [46,86].

In Japan, the publication of Isao Someya's book 'Waveform Transmission', in 1949 [87] provides another example of a work that includes the sampling theorem and that also remained unknown in the West for a long time.

5.3. *Sampling techniques in the course of Shannon's work in information theory*

Nyquist, Hartley, Küpfmüller and Gabor had not accounted for noise. Nor had they considered probabilistic models of information sources. Much of the credit for importing random processes into communication engineering is due to Wiener [88] and Rice [89,90]. Nyquist and Hartley had also made no explicit distinction between source, channel and destination. For Shannon, this distinction is essential. In his ground-breaking abstraction of the communication process, his definition of the amount of information is presented in semi-axiomatic form, the capacity of a channel is defined for channels with or without noise and the source of information is modelled as a random process. Let us consider it in some more detail.

5.3.1. *Communication in the presence of noise*

In 'Communication in the Presence of Noise' [26], Shannon begins by describing the blocks that compose a communication system: an information source, the transmitter, the channel, the receiver and the information destination. He then introduces the unit of information, the bit, 'following Nyquist and Hartley' and citing [36,59]. He notes that when it is possible to reliably distinguish M different signal functions of duration T on a channel, the channel transmits $\log_2 M$ bits in the interval T and the rate of transmission is $(\log_2 M)/T$ bits/second. He then defines the channel capacity as

$$C = \lim_{T \to \infty} \frac{\log_2 M}{T},$$

adding that the requirement of 'reliable resolution' will be clarified in the sequel. This coincides with the definition of capacity that Shannon uses in the first part of

'A Mathematical Theory of Communication' [37], in reference to the discrete noiseless channel.

Shannon assumes that signals are band-limited to W cycles/second starting at zero frequency (low-pass signals) and that the channel is available throughout T seconds. He then considers the sampling theorem, stating it as follows: if a function $f(t)$ contains no frequencies higher than W cycles/second, it is completely determined by giving its ordinates at a series of points spaced $1/2W$ seconds apart. He adds that this is 'common knowledge in the communication art' and gives a proof based on the Fourier series expansion of the Fourier transform of $f(t)$. He takes the interval $[-W, W]$ as the fundamental period and recognizes that the coefficients of the Fourier series are the samples of $f(t)$ at multiples of $1/2W$. His equation (5) expresses this clearly:

$$f\left(\frac{n}{2W}\right) = \frac{1}{2\pi}\int_{-2\pi W}^{2\pi W} F(\omega)e^{i\omega\frac{n}{2W}}\,\mathrm{d}\omega.$$

Since these samples determine the Fourier series, they determine the Fourier transform of $f(t)$ in the range $[-W, W]$. Hence, they determine $f(t)$ as well.

Here Shannon is asserting that the integers, suitably scaled, form a set of uniqueness for the Paley–Wiener space. Shannon's proof is virtually the same as that given by Borel [91], and this kind of uniqueness appears in the literature many times after that (e.g. Whittaker [92] in a more general form).

Shannon then gives the usual series

$$f(t) = \sum_{n=-\infty}^{\infty} x_n \frac{\sin \pi(2Wt - n)}{\pi(2Wt - n)},$$

where 'x_n is the nth sample', that is,

$$x_n = f\left(\frac{n}{2W}\right).$$

The question of quantization, raised by the real-valued character of the samples and the need to represent them with finite precision, is explicitly addressed in connection with sampling in another famous article of 1948, the paper on PCM[25] by Oliver et al. [93].

Shannon then argues that the series represents $f(t)$ because it 'satisfies the conditions on the spectrum and passes through the sampled values'. It is not clear whether Shannon was aware that another kind of uniqueness holds, namely, that only the sinc functions can give rise to a sampling series of the above form (see [94] for further details).

Shannon adds that the theorem has been given in other forms by mathematicians and mentions Whittaker's book [95], but that 'in spite of its evident importance seems not to have appeared explicitly in the literature of communication theory'. He cites the work of Bennett [9], which cites Raabe's paper [5], adding that it contains 'a result similar to [the sampling theorem] but on a steady-state basis' (see [34] for a detailed discussion). Finally, he adds that because Nyquist had pointed out the 'fundamental importance of the time interval $1/2W$ in connection with telegraphy' he will call the interval 'the Nyquist interval corresponding to the band W'.

Shannon then introduces time-limitation along with band-limitation as in [37]: a signal is time-limited to an interval if its samples outside that interval vanish. As a result, a signal time-limited to T and band-limited to W is determined by giving $2WT$ numbers. He adds that Nyquist [59,60] and Gabor [67] have pointed out this need for $2WT$ numbers, using the Fourier series expansion of the function on the interval T. He closes the section on sampling with a remark on nonuniform sampling: the $2WT$ numbers need not be the equally spaced samples. However, if nonuniform sampling is used, the reconstruction will be more involved and there is the possibility of noise sensitivity if there is 'considerable bunching'. He mentions that the value of the function and its second and third derivatives at every third sample point would also work and that, in general, any set of $2WT$ independent numbers would perfectly determine the function. These remarks would inspire a number of later works.

In the next section, Shannon stresses the parallel between the sampling theorem and an expansion in an orthonormal basis. More precisely, if $f(t)$ is determined by $2WT$ numbers, these numbers can be thought of as its coordinates. He shows that the 'energy' of $f(t)$ can be written as

$$E = \int_{-\infty}^{\infty} f(t)^2 \,\mathrm{d}t = \frac{1}{2W}\sum_n x_n^2,$$

using the orthogonality of the shifted sinc functions. Thus, the distance from the origin to the point represented by the x_n is $d^2 = 2WE$ or $d = \sqrt{2WTP}$, where P is the power of the signal.

He uses this to give several geometric interpretations of messages: the input and output messages are regarded as points in the input and output vector spaces; the transmitter establishes the correspondence between them; the receiver establishes a correspondence in the inverse direction. Shannon summarizes the entities in the communication system and their geometrical counterparts in tabular form. For example, 'noise in the channel' corresponds to 'a region of uncertainty about each point'.

The essence of the geometrical viewpoint used by Shannon had already become clear by 1907, when F. Riesz and E. Fisher independently proved that the classes of square-summable functions (L^2) and square-summable sequences (ℓ^2) are isometric (see [96] for details on how geometrical ideas in Hilbert space evolved). Any complete orthonormal system of functions would establish a correspondence between L^2 and ℓ^2; Shannon, who was dealing with a subspace of L^2, uses the sampling theorem and the shifted sinc functions. The extent to which his geometrical ideas were re-invented is unclear, but the idea of combining them with the sampling theorem to obtain a probabilistic description of the signals was certainly original.

Shannon then discusses the dimension of the input and output spaces and its effect. As an example, he gives a mapping from a line to a square (Figure 8) and argues that when this is done the effect of noise becomes small relative to the length of the line; but this is true only if the noise level does not exceed a critical value. From this strikingly simple example he draws an emphatic conclusion about the 'threshold effect' due to noise: there is a critical noise amplitude that causes the message to be very badly distorted.

Shannon also considers the opposite direction, when the dimensionality is reduced to compress bandwidth or time or both. He points out that the effective

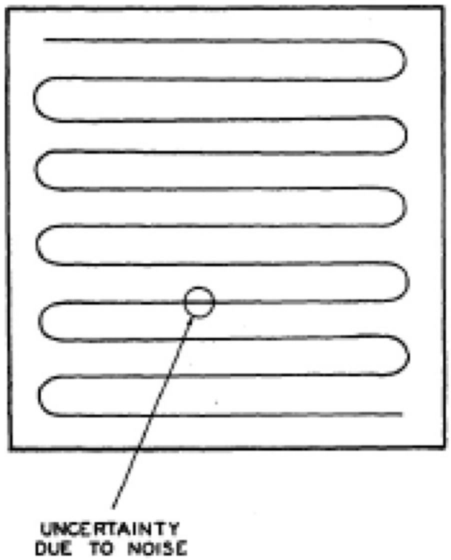

Figure 8. Shannon's example of a mapping from a line into a square, showing that the uncertainty due to noise relative to the length of the line can be controlled only if the noise level is below a certain critical value.

dimension of the message space may be well below $2WT$ and that this is what makes the transmission of signals such as speech at reduced bandwidths possible. He adds that one-to-one mappings between the square and the line, of the type Cantor introduced, can be used to reduce dimensionality even further but that it is impossible to do so continuously. The unavoidable discontinuity will produce threshold effects (which can also be understood geometrically, again from Figure 8). Only in the absence of noise it is possible to arbitrarily reduce the time-frequency product, with exact recovery of the original message.

Shannon remarks that Hartley's law, which sets an upper limit to the amount of information which may be transmitted in terms of the available bandwidth-time product, is true in a sense and false in another sense. He argues again in terms of the discontinuity of the mappings between input and output spaces of possibly different dimensions and the effect of noise. To better understand its effect, he proceeds to discuss the inherent limitations of a channel under additive white Gaussian noise of bandwidth W. If the signal has power P and the noisy signal has power $P+N$, the number of amplitudes that can be distinguished will be, roughly,

$$L = K\sqrt{\frac{P+N}{N}}.$$

Here, the constant of proportionality K determines the desired separation (with smaller values corresponding to better separation). Since there are $2WT$ independent amplitudes, the number of distinct signals will be $M = L^{2WT}$, and the number of bits that can be transmitted during the T seconds will be

$$\frac{\log_2 M}{T} = \frac{WT\log_2 L^2}{T} = W\log_2 K^2\frac{P+N}{N}.$$

Shannon makes this precise and gives the exact result: first, there exists an encoding method that leads to transmission at the rate

$$C = W \log_2 \frac{P+N}{N},$$

with arbitrarily small error probability; second, it is impossible to send at a higher rate and still have arbitrarily small error probability.

Shannon argues that since the perturbations due to noise are independent and follow the Gaussian distribution, the probability of a perturbation having coordinates $(x_1, x_2, \ldots, x_n)$ is the product of the individual Gaussian probabilities. Due to the characteristics of the Gaussian function, this product depends only on $\sum x_i^2$, meaning that the region of uncertainty is spherical.

But, using the geometrical interpretation discussed before, $\sum x_i^2$ is equal to $2WT$ times the average noise power during time T. As T increases, $\sum x_i^2$ will approximate $2WTN$. Thus, for sufficiently large T, $\sum x_i^2$ will lie within a sphere of radius $\sqrt{2WT(N+\epsilon)}$. As Shannon puts it, when $2WT$ is large the noise regions can be thought of 'as sharply defined billiard balls'.

As for the received signals, they have an average power of $P+N$, and so they must lie on the surface of a sphere of radius $\sqrt{2WT(P+N)}$. The number of distinct signals cannot exceed the ratio of the volumes of the signal spheres and noise spheres. This immediately leads to the bound

$$M \leq \left(\sqrt{\frac{P+N}{N}}\right)^{2WT}$$

and therefore to

$$C = \frac{\log_2 M}{T} \leq W \log_2 \frac{P+N}{N}.$$

It remains to show that there exists an encoding that achieves the upper bound with error probability ϵ. The M signal functions must be chosen in such a way that, when a perturbed signal is received, the nearest signal point is, with probability greater than $1-\epsilon$, the original signal. Surprisingly, as Shannon writes, one signal point selected at random from inside each sphere of radius $\sqrt{2WTP}$ will do. Each collection of M points obtained in this way corresponds to an encoding of the input. Shannon then shows that the frequency of errors averaged over all such selections is less than ϵ. This establishes the existence of an encoding with the required property (obviously, if the average of M numbers is below ϵ, at least one of the numbers will be smaller than ϵ).

The last section of this article briefly deals with continuous sources. It points out that it is impossible to send continuous information exactly over a channel of finite capacity, but that in practice a certain amount of discrepancy has to be tolerated. The rate of generating information must therefore be understood in connection with a given criterion of fidelity. Shannon then gives bounds for the rate of generating information in bits per second, in terms of the maximum tolerable mean square error and the message source and entropy powers.

5.3.2. *A Mathematical Theory of Communication*

Shannon's masterpiece 'A Mathematical Theory of Communication' [37], later reprinted in [18], considers some of the issues addressed in 'Communication in the Presence of Noise' and goes further in several directions. It begins by pointing out the interest of a general theory of communication in connection with modulation methods such as PCM and PPM.[25] Shannon mentions Nyquist and Hartley [36,59,60] on the subject and adds that he will address the effect of noise and explore the statistical nature of the messages.

The first part of his paper addresses the discrete noiseless channel. Pointing out teletype and telegraphy as examples of discrete channels, he defines its capacity as

$$C = \lim_{T\to\infty} \frac{\log_2 N(T)}{T},$$

where $N(T)$ is the number of allowed signals of duration T.

Shannon suggests statistical descriptions of a discrete source based on Markov models, from which he singles out the class of ergodic models as particularly suited to communication theory. He then asks for a quantity that measures the rate at which information is produced by such a source. After examining the properties that such a measure should have, he reaches the expression

$$H = -\sum_{i=1}^{n} p_i \log_2 p_i,$$

which is the only continuous function of the p_i with the desired properties (up to a multiplicative factor that fixes the units). He points out its similarity with 'the entropy as defined in certain formulations of statistical mechanics' and suggests the name 'entropy'.

Shannon then proves the fundamental theorem for a noiseless channel: let a source have entropy H and a channel have capacity C, then it is possible to encode the output of the source to transmit at the average rate of $C/H - \epsilon$ symbols/second, where ϵ is positive and arbitrarily small. Furthermore, the result is sharp: it is not possible to transmit at an average rate greater than C/H.

Shannon then turns to the discrete channel with noise. He observes that although at first sight errorless transmission cannot ever be achieved in the presence of noise, it is possible to reduce the probability of errors by introducing redundancy – for example, by repeating the message sufficiently many times. This suggests that to force the error probability to arbitrarily small values, the redundancy has to be increased to arbitrarily large values, which seems to imply that the rate of information transmission would approach zero.

Surprisingly this is not the case. Shannon's definition of capacity of a noisy channel, which reduces to the one already given in the noiseless case, is

$$C = \max\big(H(x) - H_y(x)\big)$$

where the maximum is taken with respect to all possible information sources used as input to the channel. The fundamental theorem asserts that if the entropy of the source, H, satisfies $H \le C$, there exists a coding system such that the output of the source can be transmitted over the channel with an arbitrarily small error probability. This is not the case if $H > C$.

The second part is devoted to the mathematically much more subtle continuous-time case. After a few preliminaries, Shannon introduces band-limited ensembles of functions and uses the sampling theorem, again regarded as an orthogonal decomposition, to map a band-limited $f(t)$ to a vector of samples in an infinite-dimensional space. For a proof of the sampling theorem he refers to 'Communication in the Presence of Noise' [26] (which had been submitted in 1940) and in which he extensively discusses the geometrical interpretation of band-limited signals as vectors of samples by means of the sampling theorem, as we have seen.

Shannon [26] writes that a function is limited to a time T if all samples outside that interval of time are zero. Functions band-limited to W and time-limited to T can be represented by $2WT$ coordinates.

The sampling theorem is a crucial tool in Shannon's theory since it provides a way of defining an ensemble of band-limited and time-limited functions by means of a probability distribution $p(x_1, x_2, \ldots, x_n)$ in the n-th dimensional space. Note, however, that time-limitation and frequency-limitation are both essential. If the time-limitation condition is not satisfied, Shannon suggests the consideration of $2WT$ coordinates that correspond to an interval of duration T 'to represent substantially the part of the function in the interval T and the probability distribution $p(x_1, x_2, \ldots, x_n)$ to give the statistical structure of the ensemble for intervals of that duration'. From this point onwards no further references to the sampling theorem are needed – but the bridge that it establishes between the continuous and the discrete has been established and will remain in the background.

6. Raabe's condition, Shannon and the Nyquist rate, treatment of transmission systems

It is difficult to determine the extent to which Raabe's work was recognised (or was ignored or unknown) through the scientific community at the time, mainly in Germany and in the United States. We overview some of the main issues in this section. For details and an English translation of the relevant parts of Raabe's thesis, see [34].

6.1. *Background on Raabe's condition*

Prof. Hans Dieter Lüke (Aachen), who first drew attention to the importance of Raabe's work in connection with sampling, exchanged an extensive correspondence with Raabe from 1978 onwards. On 10 October 1989 he congratulated Raabe on the occasion of his 80th birthday and the publication of his thesis 50 years before; Raabe's vita and a photograph were published in [7]. In a letter dated 5 October 1978, Raabe writes that he had read with great pleasure Lüke's article [6], where Raabe's dissertation is discussed with admiration. Lüke had stressed its practical aspects and the fact that it was independent of work in the same area by Kotel'nikov [84].

Raabe also mentions Koch's work [8], who recalled the term *Raabsche Bedingung* (Raabe's condition). The term continued to be used in the German electrical engineering terminology, but Koch expressed regret that it was not used in the USA and that elsewhere its influence was dying out.

Raabe recalls that he first heard the expression 'Raabe's condition' at a conference during World War II chaired by Prof. Hans Rukop (1883–1958) [97, pp. 148–150], Research Director of Telefunken at Berlin. Raabe also mentions that in the USA there never was any reference to his article. The sampling theorem had been credited to Shannon and one spoke of the 'Nyquist rate'. People were astonished whenever Raabe mentioned his 1939 paper.

In his reply of 30 October 1978 Lüke mentioned, without giving specific examples, that Raabe's paper was cited quite regularly in the German post-war literature – in contrast to the situation in the USA, in which he knew only the reference made by Bennett [9].

Bennett's paper was published in 1941, only two years after the publication of Raabe's article, and is cited by Gabor and Shannon. According to Shannon, in it 'a result similar to [the sampling theorem] is established, but on a steady-state basis' [26, p. 12]. The comparison of Bennett and Raabe's work is interesting. Unlike Raabe, Bennett considers a general 'switching function'. Raabe specifically considers a square wave and its Fourier series expansion, but goes on to consider band-pass inputs as well. Both use Fourier series and periodic inputs and both compare switching with amplitude modulation. Both papers can be considered to contain sampling results on a 'steady-state basis', as Shannon wrote.

Raabe's paper is mentioned by Bennett at the end of his paper in a group of three 'Further References', which are actually the essential ones for his own work. Bennett must have been aware of the content of Raabe's paper although it had been written in German.

6.2. *Shannon and the Nyquist rate*

As we have seen, Nyquist's classical papers of 1924 and 1928 on telegraphy [59,60] show a thorough understanding of the connection between signalling speed, number of bits per symbol and bandwidth. However, an explicit formulation of the sampling theorem cannot be found there and it would take some more time to appear in the engineering literature. Nyquist did establish that 'the speed with which intelligence can be transmitted' increases linearly with the line speed or the number of symbols that can be transmitted per unit time, and linearly with the number of bits per symbol. His view of telegraphy is strikingly modern and contains key ideas underlying digital communication systems such as PCM, which was discussed two decades later.

PCM was the subject of the important paper by Oliver et al. [93]. The paper connected key concepts such as sampling, quantization, coding and channel capacity. Its Appendix I contains a proof of the sampling theorem (along the lines given in [26]), but more importantly it described the essence of the problem in simple terms, which would appeal strongly to the engineering audience:

> To reconstruct the signal it is merely necessary to generate from each sample a proportional impulse, and to pass this regularly spaced series of impulses through an

> ideal low-pass filter of cutoff frequency W_0. The output of this filter will then be (except for an overall time delay and possibly a constant of proportionality) identical to the input signal. Since the response of an ideal lowpass filter to an impulse is a sin x/x pulse, and since the total output is the linear sum of the responses to all inputs, this method of reconstruction is simply the physical embodiment of [the sampling theorem].

This expository account is a synthesis of work that had its origin with Nyquist and the earlier precursors of telegraphy and culminated in Shannon's theory of communication. As the advantages of PCM started to be recognized, the importance of Nyquist's legacy became clear. Oliver et al. wrote:

> If it surprises the reader to find that $2W_0T$ pieces of data will describe a continuous function completely over the interval T, it should be remembered that the $2W_0T$ coefficients of the sine and cosine terms of a Fourier series do just this, if, as we have assumed, the function contains no frequencies higher than W_0.

The approach of Nyquist (as that of Raabe) was of course firmly based on the analysis of the Fourier series of periodic signals, as these lines bring to mind. What Shannon wrote when he discussed the sampling theorem in [26] is, therefore, not surprising:

> Nyquist pointed out the fundamental importance of the time interval $1/2W$ seconds in connection with telegraphy, and we will call this the Nyquist interval corresponding to the band.

Shannon's 'crisp statement and proof of the sampling theorem', as Verdú puts it [98, p. 2063], was instrumental in popularizing sampling in the engineering mainstream. But although Shannon mentioned Whittaker, Hartley and Nyquist, one should recognize the differences between their works. E.T. Whittaker, as pointed by Khudiakov [99], considers the repetition rate of interpolation nodes; Nyquist considers the repetition rate of elementary pulses and Kotel'nikov considers the sampling frequency. Raabe, as we have seen, considers the multiplexing frequency, which is nothing but the sampling frequency for each multiplexed channel. In one way or the other, these authors are all interested in the relation between these rates and the properties of the signal itself, and particularly in their frequency content. Nyquist is concerned with the rate of transmission and the bandwidth required and Whittaker is interested in the interpolation of regularly spaced samples by analytic functions. To build his multiplexer, Raabe determines the required sampling frequency for a given signal bandwidth, in the lowpass and bandpass cases; he implements sampling without giving the sampling theorem explicitly. Kotel'nikov, on the other hand, gives the reconstruction formula without implementing it. Shannon recognizes the importance of sampling as a part of information theory, separates the concepts of source, channel and receiver and considers stochastic inputs; he is then able to go much further by introducing concepts such as entropy, mutual information, code and capacity, around which the central results of information theory would turn.

6.3. *Raabe's versus Shannon's approach in treating transmission systems*

As is well-known, electrical circuit theory depends heavily on simple linear circuit elements which can, in general, be called impedances. Mathematically, the relation between the voltage applied to an impedance and the current that flows through it can be expressed by a linear differential operator. This explains the usefulness and success of Fourier techniques in circuit analysis.

The most natural Fourier analysis tool to use in circuit analysis when the signals of interest are periodic is Fourier series. This is especially true in steady-state analysis, in which the engineer is interested in quantities such as gain or magnitude as functions of the frequency.

Since Fourier series were well known in Raabe's time, and since the multiplexing problem involves a periodic multiplexing signal and relatively simple circuit elements, it is not surprising that both Raabe and Bennett have relied on them.

Generally speaking, harmonic analysis is not as successful in connection with the investigation of highly non-linear or very complex systems. As communication systems and communication problems became increasingly complex, the importance of harmonic analysis gradually decreased while that of statistical methods increased.

This is true, for example, in the case of Shannon's fundamental theorem for the discrete channel with noise, a result that opened an entirely new research field. It would take 50 years of intense efforts to find capacity-achieving codes [100,101] – and probabilistic methods were essential to reach the goal.

6.4. *Bennett's work in treating transmission systems*

Since Shannon cited Bennett's [9] multiplex systems paper, which cited Raabe [5], it is appropriate to give an overview of Bennett's work on time division multiplex systems. In fact, Shannon writes [26, p. 12] with respect to his sampling theorem that 'a result similar to Theorem 1 is established [by Bennett], but on a steady-state basis', exactly the same basis as with Raabe.

Bennett, just like Raabe, assumes an N-channel system with a sinusoidal signal $E_j(t) = E_j \exp(i\omega_j t)$ impressed on the j-th channel (see Figure 1 in [9, p. 201]). Then he lets the switching between the j-th channel and the transmission line at the sending end to be represented by $I_{sj}(t) = F_j(t)E_j(t)$, where the function $F_j(t)$ is periodic in time with fundamental frequency $q = 2\pi/T$, where T is the time occupied by one cycle of the switching operation. Here he assumes for $F_j(t)$ a somewhat general function of time, choosing a general Fourier series[27] approach:

$$F_j(t) = \sum_m A_{mj} e^{i(mqt-\theta_{mj})}.$$

It follows that

$$I_{sj}(t) = F_j(t)E_j(t) = E_j \sum_m A_{mj} e^{i(mqt+\omega_j t-\theta_{mj})}. \tag{1}$$

The next step is the transmission of the wave over a line, the properties of which are in general specified by a complex transfer impedance $E_r/I_s = Z(i\omega)$. The result of applying the wave (1) to the line is then the open circuit voltage

$$E_{rj} = E_j \sum_m A_{mj}\, Z[i(mq + \omega_j)]\, e^{i(mqt+\omega_j t-\theta_{mj})}.$$

At the receiving end, another switching process takes place synchronously with that at the transmitting end. Thus the switching process between the k-th channel and the line can be represented by $I_{rk}(t) = G_k(t)E_{rj}(t)$, where $G_k(t)$ is a periodic function of time with fundamental frequency q, and will be expressed in a manner analogous to the corresponding $F_j(t)$, namely as a general Fourier series:

$$G_k(t) = \sum_n B_{nk} e^{i(nqt - \phi_{nk})}.$$

In combining all this Bennett finds

$$I_{rk}(t) = E_j \sum_m \sum_n A_{mj}\, B_{nk}\, Z[i(mq + \omega_j)]\, e^{i[(m+n)qt + \omega_j t - \theta_{mj} - \phi_{nk}]}.$$

In essence, this is his received wave equation (10) [9, p. 203], which thus consists of a doubly infinite set of side frequencies involving harmonics of q. It is however possible to set up conditions under which the original signal may be selected and the frequencies involving the switching rate may be suppressed by filtering. Such a separation is possible provided $\omega_j < q/2$, because it then follows that a lowpass filter with cutoff frequency at $q/2$ will not pass any of the components with frequencies dependent on q. The condition $\omega_j < q/2$ is exactly the Raabe condition. Then assuming that it is fulfilled, he calculates channel output $I_{ck}(t) = Y_{jk} E_j \exp(i\omega_j t)$. Raabe, in Bennett's terminology, arrives at $I_{cj}(t) = Y_{jj} E_j \exp(i\omega_j t)$.

In his further treatment Bennett investigates crosstalk between adjacent channels, as Raabe did, but he can carry this out in a greater generality due to Y_{jk}. Bennett treats on-and-off switching where he now selects suitable switching functions $F_j(t)$ and $G_k(t)$ for crosstalk suppression and minimum bandwidth. Further on, he investigates transmission requirements in the case that the transfer impedance Z is acted by frequency-dependent phase shifts and gain.

All in all, Bennett's approach is more general but in the same spirit as that of Raabe. Both use Fourier series and periodic inputs and both compare switching with amplitude modulation. Bennett considers a general 'switching operation' in which the signal is multiplied by a periodic function, represented by a Fourier series. Raabe specifically considers a square wave and its Fourier series expansion, but goes on to consider band-pass inputs as well. As already mentioned in Section 6.1, both papers can be considered to contain sampling results on a 'steady-state basis', as Shannon wrote.

Bennett [102, pp. 140–152] still uses his results of 1941 without working in the more modern statistical setup. In this sense this book – the basic results of which are understandable in terms of operations on sine waves – represents a 'back to basics' approach.

7. A brief biography of Raabe

Herbert P. Raabe was born on 15 August 1909 in Halle, Germany. He attended the *Humboldt Oberrealschule* in Zeitz, receiving the graduation certificate in 1929. Subsequently, he studied Electrical Engineering and Telecommunications at the TH Berlin, receiving the *Diplom* in 1936. In 1937 he became a Research Assistant and in December 1939 an *Oberingenieur* at the Chair in Telecommunications of Professors Wilhelm Stäblein (1900–1945) and Karl Küpfmüller, who between 1935 and 1936 had been responsible for the vacant chair in Telecommunications.

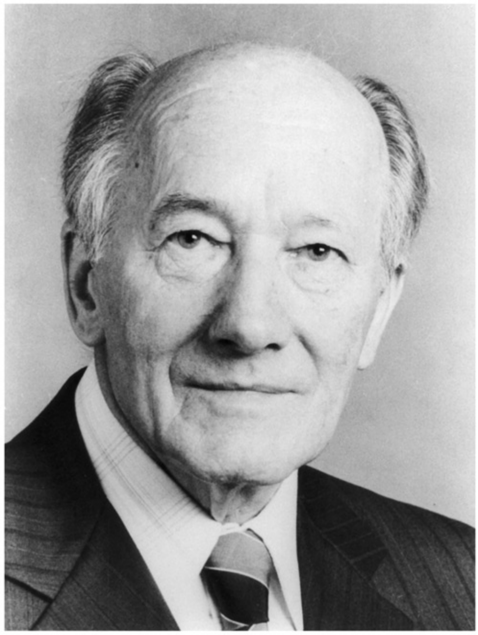

Figure 9. Herbert Raabe around 1989, when he was about 80 years old.

In 1939 Raabe earned his doctorate (*summa cum laude*) with the dissertation *Untersuchungen an der wechselzeitigen Mehrfachübertragung* which appeared in 1939 in the journal *Elektrische Nachrichtentechnik* [5] and in which he describes his time division multiplexing system.

Raabe's thesis was supervized by Stäblein and Heinrich Fassbender (1884–1970), who had been professor of high-frequency engineering at TH Berlin since 1918.

Raabe sent his *Erinnerungen* (personal recollections) [103] to Prof. P. Noll (Berlin), in a communication dated 20 December 1994. These highlight his time in Berlin up to 1947 and show his practical experiences in signal processing and radar in a new perspective. He speaks of his influences on his profession, and also singles out some of his teachers for special mention. These include three electrical engineers whom we have already met; Fassbender (whose lectures especially appealed to him), Küpfmüller and Stäblein, among others.

In 1989 Raabe (Figure 9) was honoured at a celebration at Kleinheubach, Germany, on the 50th anniversary of the publication of his doctoral thesis. Raabe, who was the author of 15 articles and 19 patents, died on 25 August 2004, in Potomac, Maryland, where he had resided since 1968.

8. Conclusion

There is convincing evidence for giving Raabe credit for discovering the minimum sampling rate for errorless transmission, independently of Kotel'nikov. He also gave, for the first time, the condition for the band-pass case. It turns out that his conclusions hold in more general contexts, but he formulated them in this setting, one decade before Shannon named the condition after Nyquist, and deserves credit for that.

Why was the impact of Raabe's work not more widely felt? We have suggested several reasons. His thesis was written in German, published in a German journal and appeared just before World War II. After the war Raabe moved to the USA where he began working in other areas. By that time Shannon's work had appeared in the USA and was having an enormous impact. The sampling theorem became associated with the name of Shannon and the minimum sampling rate with that of Nyquist, although for Shannon sampling was 'common knowledge in the communication art' [37]. As has been pointed out, he cited Bennett [9], who had already cited Raabe. However, as we discussed, this did not attract attention to Raabe's work.

By the time Shannon stated and proved the sampling theorem, a number of mathematicians and engineers had already contributed to sampling. It is fair to emphasize, however, one characteristic that distinguishes Raabe's work: at a time when the principles underlying his subject were still unclear, Raabe managed not only to understand the problem, but also to build and test a system that carried the idea to practice. This practical side of his work, which appeared before Shannon's communication theory had popularized sampling, is indeed outstanding.

Acknowledgements

The authors would like to thank Prof. Dr Ing. J.R. Ohm, Prof. Lüke's successor in the office of IENT, for giving them access to the exceptional library of his institute, and for kindly preparing the two Shannon photos (Figures 6 and 7) for publication, with the assistance of Mrs. Myrjam Schiermeyer. This work was partially supported by an FCT grant (PTDC/EEA-TEL/108568/2008, COMPETE reference FCOMP-01-0124-FEDER-009920).

Notes

1. The article by Butzer et al. [34], an in-depth study of the life and work of Raabe, contains an English translation of the relevant parts of Raabe's doctoral thesis [5]; Raabe's contribution is analysed in detail, and the concepts used at the time are put into the present day terminology. This article is, in contrast, centred on the life and work of Shannon, which is described in some detail and set against the earlier work of Raabe.
2. William Robert Bennett was born on 5 June 1904, in Des Moines, Iowa. He received his BS degree in electrical engineering from Oregon State College, in 1925, and A.M. and PhD degrees from Columbia University, in 1928 and 1949, respectively. He joined the research department of Bell Labs in 1925 and worked in multichannel communication, including multiplex telephony. He also investigated the effects of nonlinear distortion and the spectra of quantized signals [104]. His books on data transmission and noise include [102,105,106]. He had a gift for exposition and could make a complex problem understandable to 'even the most inexperienced beginner' [107]. He became a fellow of the IEEE in 1956 and retired from Bell Labs in 1965, as Head of the Data Theory Department. He accepted a Professorship at Columbia University and was appointed Charles Batchelor Professor of Electrical Engineering there in 1968; Emeritus, 1972. He received the Mervin J. Kelly Award in Telecommunications from IEEE in 1968 and died on 21 August 1983.
3. Karl Willi Wagner was removed from office in 1936 since he refused to dismiss his Jewish employees. He escaped being sent to a concentration camp in view of a thrombosis.

His research assistant and later co-author of his well-known book on operational calculus, Dr Alfred Thoma, who denounced certain accusations made against his teacher, was fired immediately. This information was kindly supplied by his son Ulrich O.E. Thoma and is to be found in his father's autobiography; see http:/www.ulrichthoma.de/alfredthoma/.

4. See [108] and also the web link http:/www.nue.tu-berlin.de/history/, which includes biographies and other useful historical information pertaining to the TH/TU Berlin.
5. Leo Szilard was born on 11 February 1898, in Budapest. He moved to Germany in December 1919 and became a student of the TH Berlin. Divided among engineering and physics, he started to attend the physics colloquia at Berlin University in January 1920; joining this eminent circle of physicists changed Szilard's life. He obtained his PhD degree from Berlin University in 1922, officially under von Laue, but with the support of Einstein, with whom he had developed a close relationship. Of the several patents and inventions that he filed from 1926 onwards, several were joint work with Einstein. In 1933 Szilard left from Germany to England. Later on, in 1938, he moved to the USA. During that period he developed his ideas about nuclear chain reaction and critical mass, for which he would file a patent. With Wigner, he organized the letter of Einstein to Roosevelt that would initiate the atomic bomb project. However, in 1945 he started two versions of a petition to prevent the deployment of atomic bombs against Japan. He would defend nuclear weapon control until the end of his life. He was appointed Professor at the University of Chicago in 1946, a Fellow of the American Academy of Arts and Sciences in 1954 and a member of the National Academy of Sciences of the USA in 1961. He received the USA Atoms for Peace Award in 1960 (see also [11,109]). Szilard died on 30 May 1964 in La Jolla, CA.
6. John von Neumann was born on 28 December 1903, in Budapest. In 1921 he enrolled in mathematics at the University of Budapest, but spent most of his time at the TU Berlin and the Zurich Polytechnic, where he studied mathematics and chemical engineering. He obtained his PhD degree in mathematics from the University of Budapest and the chemical engineering degree from Zurich at about the same time, in 1925. During the period 1927–1929 he taught as a *Privatdozent* at the University of Berlin. In 1929 he moved to the University of Hamburg, and in 1930 to the USA. In 1930 he became a visiting professor at Princeton University. He obtained a permanent position there in 1931 and in 1933 he was invited to join the Institute for Advanced Study, becoming the youngest of its six initial members. Von Neumann made fundamental contributions in a number of topics: the axiomatization of set theory, the mathematical foundations of quantum physics (included in this group is his famous book [75] on the mathematics of quantum mechanics), game theory and mathematical economics (among which stands out the influential book with Oskar Morgenstern on game theory and economic behaviour), spectral theory and operator algebras, ergodic theory and numerical mathematics and computer science. His contributions to modern computing are discussed in [17,19,110,111], among others. See also the paper by Ulam [112], which is part of a memorial issue of the *Bulletin of the AMS* that also includes an account written by Shannon [113] of von Neumann's work on automata theory. John von Neumann died on 8 February 1957, in Washington, DC.
7. Jean-Maurice-Émile Baudot (1845–1903) was born on 11 September 1845 in Magneux, France. A telegraph engineer and one of the pioneers of telecommunications, he worked in the development of fast telegraphy, and invented the Baudot system for simultaneous transmission of several signals over the same wire (Figure 4). Baudot's invention of 1875–1977 was based upon a distributor of B. Meyer of 1870/1871 and the code used by Gauss and Weber in their telegraph experiments of 1833 [2,58]. Baudot's name became attached to the Baudot code, a predecessor of the ASCII code. The term 'baud', a measure of the number of symbols transmitted per second, is named after him.
8. Patrick Bernard Delany (1845–1924) was born in King's County, Ireland, on 28 January 1845, and moved to the United States in 1854. By the age of 16 he was already an accomplished telegraphist and became Chief Operator of the Franklin Telegraph Company, Assistant General Superintendent of the Southern and Atlantic Telegraph Company and superintendent of the Automatic Telegraph Company. He left

telegraphy and became a newspaper correspondent at Washington and subsequently Editor of *Old Commonwealth*, Harrisonburg, Virginia. From 1880 onwards he turned his attention to inventing. He obtained over 100 patents, mostly related to telegraphy. For his work in high speed and synchronous telegraphy, the Franklin Institut awarded Delany the John Scott Medal in 1885 and the Elliot Cresson Medal in 1886 and 1896, respectively.

9. Bernard Meyer (1830–1884) was a French telegraph operator. Meyer's distributor of 1870/1871, which influenced Baudot, led to a multiplexer that was first used on the Paris–Lyon line already in 1872.
10. Poul la Cour (1846–1908), a Danish meteorologist, invented and patented in 1874 (the year Edison invented quadruplex telegraphy) a telegraphic device based on tuning forks. Today the main idea behind la Cour's device would be described as frequency-domain multiplexing (see [30,31] for an account of other pioneering efforts at frequency-domain multiplexing). In the United States, la Cour's invention was credited to Elisha Gray, who had worked along similar lines. la Cour protested but found himself unable to support the legal battle and withdrew his claim. Nevertheless, in 1886, the Franklin Institute awarded la Cour the John Scott Medal for his phonic wheel of 1877.
11. Generalized sampling theory was developed at Aachen from 1977 onwards. It was first studied in the doctoral dissertation of Splettstößer [114], continued by Stens [115], Ries–Stens [116] and Butzer et al. [117]; overview papers are [118–120]. Let us point out that it was Otto Lange who in 1975 introduced PLB and his research group, especially Wolfgang Splettstößer, to signal analysis, in particular to investigate Shannon's sampling theorem from a critical, mathematical point of view. It finally lead to some 150 papers in the broad area by this research group up to 1994 [121]. In 1970, while studying and doing graduate work in EE and information science at Aachen, Dr Lange obtained, in addition, his Dipl. Math. degree, one examiner being PLB.
12. The sinc function (which had been defined by Raabe's teacher Küpfmüller) is never mentioned by Raabe. However, it should be kept in mind that the ideal lowpass filter, and hence the sinc function, cannot be implemented. Therefore, it could not ever become part of the multiplexing system that Raabe was trying to design and build.
13. This subsection on distortion treats Sections 5 and 6 of Raabe's doctoral thesis [5] which were not discussed nor translated in [34]. The authors are planning to present an online translation of the whole Raabe thesis.
14. Recorded at Winchester, MA, 28 February 1977. Other sources of information by Shannon himself include an interview with Robert Price (28 July 1982) and a Kyoto Prize speech draft that Shannon wrote in 1985. See also the comprehensive thesis of Hagemeyer [61] and that of Guizzo [16].
15. Shannon received this award (of DM 200,000) for his 'fundamental research on information theory'. Prof. H.D. Lüke, chairman of the Board of Curators of the Eduard Rhein Foundation, had suggested Shannon for the Award. The foundation of Eduard Rudolph Rhein (1900–1993) is now managed by his nephew Dr Rolf Gartz.
16. Harry Nyquist was born on 7 February 1889 in Nilsby, Sweden. He emigrated to America, attended the University of North Dakota, Grand Forks, from 1912 to 1915, and received his BS and MS degrees in Electrical Engineering in 1914 and 1915, respectively. He attended Yale University, New Haven, Connecticut, from 1915 to 1917, and was awarded a PhD in 1917.

    From 1917 to 1934 Nyquist was employed by the American Telephone and Telegraph Company in the Department of Development and Research Transmission, where he was concerned with studies on telegraph picture and voice transmission. From 1934 to 1954 he was with the Bell Telephone Laboratories, Inc., where he continued in the work of communications engineering, especially in transmission engineering and systems engineering. When he retired he had obtained 138 US patents and published 12 technical articles.

    He received many honours for his outstanding work. He was the fourth person to receive the National Academy of Engineer's Founder's Medal, 'in recognition of his many fundamental contributions to engineering'. In 1960, he received the IRE Medal of Honor 'for fundamental contributions to a quantitative understanding of thermal noise, data transmission and negative feedback'. Nyquist was also awarded the Stuart Ballantine

Medal of the Franklin Institute in 1960, and the Mervin J. Kelly award in 1961. He passed away on 4 April 1976.

17. Ralph Vinton Lyon Hartley was born on 30 November 1888 in Spruce, Nevada. He attended the University of Utah and received as a Rhodes Scholar the BA degree in 1912 and the BSc degree in 1913 from Oxford University. Upon returning to the USA, he worked for the Western Electric Company and was in charge of the Bell System's transatlantic radiotelephone tests. He invented the oscillating circuit that beared his name during that period. After World War I he worked at Bell Laboratories, doing research on repeaters and voice transmission. After a period of illness he returned to Bell Laboratories as a consultant. Hartley proposed the linear transformation that beared his name in 1942 [122]. He received the IRE Medal of Honor in 1946 'for his early work on oscillating circuits employing triode tubes and likewise for his early recognition and clear exposition of the fundamental relationship between the total amount of information which may be transmitted over a transmission system of limited band-width and the time required'. He retired in 1950 and died on 1 May 1970.
18. It is worth pointing out that McCulloch [123], a promoter of cybernetics, reports having received from Hartley in 1929 'a reference to the definition of information by C.S. Peirce', which McCulloch calls 'the bud of the American definition of information as a quantity'. C.S. Peirce (1839–1914), a philosopher and logician, was a co-founder of semiotics.
19. Karl Küpfmüller, born on 6 October 1897 in Nürnberg, was an Electrical Engineer and a pioneer in communication theory and control engineering [124, pp. 96–102], [125]. After attending school, he received a practical engineering education in *Siemens-Schuckert Werke* in Nürnberg in 1914–1915 and attended until 1919 the Ohm-Polytechnikum in Nürnberg, the period 1916–1918 being spent with the armed forces. During 1919–1921 he was fortunate to be an assistant of Karl W. Wagner at the *Telegraphentechnisches Reichsamt* in Berlin and in 1921 he entered the *Zentral-Laboratorium* of *Siemens* & *Halske AG* as an engineer. During his years at Siemens he attended for three semesters a variety of university courses, and in 1928, as a non-academic, he received a full professorship for general communication engineering at the TH Danzig. In 1935 he was with the TH Berlin and in 1937 back again at *Siemens* & *Halske*, where he became Director of Research and Development for Communication Theory. He remained at the TH Berlin as honorary professor. After the war, he was with the firm *Rohde* & *Schwarz* during 1946–1948 in Munich (but interned until 1947 by the US occupational forces). In 1948–1952 he was with *Standard Electric Lorenz AG*, and finally from 1952 to 1963 as Full Professor and Director of the Institute for Telecommunication Engineering at the TH Darmstadt, perhaps the best in the field at that time. There he was a co-founder of the ITG (*Informationstechnische Gesellschaft*).

    Küpfmüller's classic textbook *Einführung in die Theoretische Elektrotechnik* (1st edition, Berlin 1932, 285pp.) is still in print after 18 editions [126]. His *Die Systemtheorie der elektrischen Nachrichtenübertragung* (Stuttgart, 386pp.), which appeared in 1949, was based on lectures given from 1937 to 1943.

    He was highly honoured in his lifetime, being awarded, e.g. the Gauss-Weber Medal in 1932, honorary doctorates from the TH Danzig in 1944 and Erlangen in 1976, in addition to medals from Sweden and Austria. He died on December 26, 1977, in Darmstadt.
20. Fritz (Friedrich Heinrich) Lüschen, born on 19 March 1877 in Oldenburg, was one of the most important German communication engineers between WW I and WW II. He worked for 25 years for the *Deutsche Reichspost*, entering its administration as an 18 year old in 1895. After 10 years there, he spent five months at the *École Supérieure des Postes et Télégraphes* in Paris and then studied during 1905–1911 mathematics and physics at the University of Berlin, besides his work at the *Reichspost*. He also took in 1907 a nine-month engineering course at the *Reichspostversuchsamt*. Then in 1911 he became a telegraph engineering inspector. During WW I, he first served as a stage telegraph inspector and after 1917 as a second lieutenant at the headquarters of the German army in Turkey and Palestine. In 1920 Lüschen joined *Siemens & Halske AG*, first as Chief-Engineer of the laboratory for low-voltage cables, then in 1921 as Director of their Central Laboratory for Telecommunications, which he founded, and then in 1930 as director of their total cable community. Finally, in 1944 he was appointed by Albert Speer

as Chairman of Development and Production of the entire German electrical engineering industry. On 18 June 1945 he committed suicide in Berlin. He was awarded honorary doctorates from Cologne University in 1925 and TH Danzig in 1929. He received the prestigious Gauss-Weber Medal from Göttingen in 1933 (see also [127]).

21. Felix (Alexander Philipp) Strecker, born on 27 February 1892 in Dembno (Jarotschin District, now Poland), was an authority in communication theory, a subject to which he contributed substantially. He studied botany and zoology during 1910–1914 at the University of Halle and the TH Munich. He joined the armed forces during WW I, then he studied in 1921–1923 at Halle, receiving his Dr. rer. nat. in physics under Karl E.F. Schmidt in 1923. In that same year he joined *Siemens & Halske AG*, the main laboratory of which was just coming into being and would gather names like Lüschen, Bruno Pohlmann (1884–1958), Küpfmüller and Dennis Gabor. Strecker worked first as a Patent Engineer, then in 1929 he became in charge of the *Zentrallabor* headed by Fritz Berger, in 1934 he became Chief-Engineer and in 1935 head of the long distance telephone laboratory, among other duties. From 1937 to 1944 he was the authorized person (*Bevollmächtigter*) of the *Siemens* & *Halske AG*, Berlin-Siemensstadt. During this period alone he published more than 20 papers. After the war his health deteriorated, but he wrote two books and returned to the *Zentrallabor*. In 1950 he settled in Munich, where he passed away in 1951.

    Let us just mention two prominent results (see [128–132] for more information). In 1929 Strecker introduced together with Richard Feldtkeller (1901–1981) matrix calculus into the theory of linear networks and amplifiers, giving it the present form; he had been introduced to the calculus by his teacher Prof. H. Jung at Halle.

    In 1930 he discovered the stability criterion usually associated with the name of Nyquist. He lectured on it in 1931 and wrote a manuscript entitled 'Die Bedingungen der Selbsterregung in linearen Gebilden' which the journal *Elektrische Nachrichtentechnik* did not publish (the journal still exists). As a result, the first publication on the subject was Nyquist's paper of 1932 [133], who found the criterion independently. Strecker was able to publish his result of 1931 only in 1947 in his book 'Die elektrische Selbsterregung' [134] (see also [135]).

22. Dennis Gabor was born on 5 June 1900, in Budapest, Hungary. Already in high school he was demonstrating a deep understanding of physics. He attended the Budapest Technical University, and obtained a degree in Mechanical Engineering. Because he opposed the monarchy that had come to power in 1920, he fled to Germany and studied at the Technical University of Berlin, from which he received the Diploma in Engineering in 1924 and the doctorate in 1927. During this period, he invented a fast-response cathode-ray oscilloscope. After graduation he worked in the physics laboratory of Siemens and Halske but was forced to return to Hungary in 1933: the Nazi regime had started and his contract was terminated because he was not German. He emigrated to the UK and got a job at the British Thomson-Houston Company. During the war, his Hungarian citizenship proved to be an obstacle and his scientific work did not go well. The breakthrough occurred in 1947, when he invented holography. Although the potential of his invention was not fully appreciated until the invention of the laser, its importance would be recognized. He was awarded the Nobel Prize for Physics in 1971 as a result of his pioneering work in holography. He received other honours, including the Albert Michelson Medal of the Franklin Institute, in 1968, and the Medal of Honour of the IEEE, in 1970. He died on 9 February 1979, in London.

23. Vladimir Alexandrovich Kotel'nikov (1908–2005) was born on 6 September 1908 in Kazan, the capital of what was then the Republic of Tatarstan. He graduated from the Moscow Power Engineering Institute (MEI) in 1930. He became a postgraduate at MEI, and was promoted to Senior Laboratory Assistant and then to Assistant Professor. His doctor of science thesis of 1946, translated in 1959 [136], is another of his pioneering works that drew little attention at the time it appeared despite the importance of the results it contained (see, in this respect, [137]). Kotel'nikov also worked in cryptography and planetary radar. He was elected a full member of the USSR Academy of Sciences in 1953, received the State Prize (twice, in 1943 and 1946), the Lenin prize (1964), the Eduard Rhein Foundation award (in 1999, at age 91, following a proposal

made by Hans Dieter Lüke), the IEEE Alexander Graham Bell Medal (2000) and several other honours.

24. As late as 1995, Vladimir Tichomirow, the great Russian expert in the broad areas of signal analysis and approximation, and a student of Kolmogorov, told one of the authors (P.L.B.) at Aachen that this paper was not available in Moscow; he himself had never seen it.
25. In PCM, or pulse code modulation, an analogue signal is represented by uniformly sampling it and then quantizing the samples to a fixed, finite precision. In the compact disc, for example, the sampling rate is 44.1 kHz and the samples are quantized to 16 bits.
26. In PPM, or pulse position modulation, a message is conveyed by varying the position of a pulse. If the pulse can occupy any of 2^n equally likely positions, each message will consist of n bits.
27. We have slightly simplified Bennett's equations by taking his $\nu = 0$ and using complex exponentials.

References

[1] K.G. Beauchamp, *History of Telegraphy: Its Technology and Application*, IET, London, 2001.

[2] A.A. Huurdeman, *The Worldwide History of Telecommunications*, John Wiley & Sons, Hoboken, New Jersey, 2003.

[3] H.D. Lüke, *The Origins of the sampling theorem*, IEEE Commun. Mag. 37(4) (1999), pp. 106–108.

[4] W. Kester (ed.), *The Data Conversion Handbook*, Elsevier, Amsterdam, 2005.

[5] H. Raabe, *Untersuchungen an der wechselzeitigen Mehrfachübertragung (Multiplexübertragung)*, Elektrische Nachri. 16 (1939), pp. 213–228 (Text identical to Dissertation D83, 'Untersuchungen an der wechselzeitigen Mehrfachübertragung', TH Berlin, submitted April 24, 1939).

[6] H.D. Lüke, *Zur entstehung des abtasttheorems*, NTZ 31(4) (1978), pp. 271–274.

[7] H.D. Lüke, *Herbert P. Raabe der 'Vater' des abtasttheorems*, NTZ 43 (1989), p. 46.

[8] G.F. Koch, *Bemerkungen zu dem aufsatz: 'Zur entstehung des abtasttheorems' von H. D. Lüke*, NTZ 43 (1989), p. 498.

[9] W.R. Bennett, *Time division multiplex systems*, Bell Syst. Tech. J. 20 (1941), pp. 199–221.

[10] W. Ebeling and I.M. Sokolov, *Statistical Thermodynamics and Stochastic Theory of Nonequilibrium Systems*, World Scientific, Singapore, 2005.

[11] I. Hargittai, *The Martians of Science: Five Physicists who Changed the Twentieth Century*, Oxford University Press, Oxford, 2006.

[12] B.T. Feld and G.W. Szilard (eds.), *The Collected Works of Leo Szilard: Scientific Papers*, MIT Press, Cambridge, MA, 1972.

[13] J.R. Pierce, *Signals and Noise: The Nature and Process of Communication*, Harper & Row, New York, 1961.

[14] H.S. Leff and A.F. Rex (eds.), *Maxwell's Demon 2: Entropy, Classical and Quantum Information, Computing*, Institute of Physics Publishing, Bristol, 2003.

[15] W. Lanouette and B. Silard, *Genius in the Shadows: A Biography of Leo Szilard, The Man Behind the Bomb*, Charles Scribner, New York, 1992.

[16] E.M. Guizzo, *The essential message: Claude Shannon and the making of information theory*, Master's thesis, Massachusetts Institute of Technology, 2003.

[17] G. Israel and A.M. Gasca, *The World as a Mathematical Game: John von Neumann and Twentieth Century Science*, Birkhäuser, Basel, 2009.

[18] C.E. Shannon and W. Weaver, *The Mathematical Theory of Communication*, The University of Illinois Press, Urbana, 1964 (first edition by Urbana University Press, 1949).

[19] W. Aspray, *John von Neumann and the Origins of Modern Computing*, The MIT Press, Cambridge, MA, 1990.

[20] J. von Neumann, *Thermodynamik quantenmechanischer Gesamtheiten*, Gött. Nach. 1 (1927), pp. 273–291.
[21] P.R. Masani, *Norbert Wiener 1894–1964*, Birkhäuser, Basel, 1990.
[22] N. Wiener, *Cybernetics or Control and Communication in the Animal and the Machine*, MIT Press, Cambridge, MA, 1948.
[23] R.A. Fisher, *Indeterminism and natural selection*, Philos. Sci. 1 (1934), pp. 99–117.
[24] R.A. Fisher, *The Design of Experiments*, Hafner Press, New York, 1935 (8th ed., 1966).
[25] A.N. Kolmogorov, *Interpolation und extrapolation von stationären zufälligen Folgen* [*Interpolation and extrapolation of stationary random sequences*], Bull. Acad. Sci. U.S.S.R. ser. Math. 5 (1941), pp. 3–14 (Izvestiya Akad. Nauk).
[26] C.E. Shannon, *Communication in the presence of noise*, Proc. IRE 37 (1949), pp. 10–21.
[27] N. Wiener, *Invention: The Care and Feeding of Ideas*, The MIT Press, Cambridge, MA, 1993 (posthumous, originally written in the 1950s).
[28] A.B. Carlson, *Communication Systems: An Introduction to Signals and Noise in Electrical Communication*, 3rd ed., McGraw-Hill, Inc., New York, 1986.
[29] S. Haykin, *Communication Systems*, 4th ed., John Wiley & Sons, New York, 2001.
[30] J. Blanchard, *The history of electrical resonance*, Bell Syst. Tech. J. 20 (1941), pp. 415–453.
[31] J. Blanchard, *A pioneering attempt at multiplex telephony*, Proc. IEEE 51(12) (1963), pp. 1706–1709.
[32] Anonymous, *On the synchronous-multiplex telegraph*, Proc. Am. Philos. Soc. 21(114) (1883), pp. 326–328.
[33] W.H. Preece, *A visit to Canada and the United States in the year 1884*, Scientific American Supplement XIX(481), a lecture delivered before the Society of Telegraph Engineers and Electricians, London, 11 December 1884.
[34] P.L. Butzer, M.M. Dodson, P.J.S.G. Ferreira, J.R. Higgins, O. Lange, and P. Seidler, *Herbert Raabe's work in multiplex signal transmission and his development of sampling methods*, Signal Process. 90(5) (2010), pp. 1436–1455.
[35] S.W. Golomb, E. Berlekamp, T.M. Cover, R.G. Gallager, J.L. Massey, and A.J. Viterbi, *Claude Elwood Shannon (1916–2001)*, Notices Am. Math. Soc. 49(1) (2002), pp. 8–16.
[36] R.V.L. Hartley, *Transmission of information*, Bell Syst. Tech. J. 7 (1928), pp. 535–564.
[37] C.E. Shannon, *A mathematical theory of communication*, Bell Syst. Tech. J. 27 (1948), pp. 379–423, 623–656.
[38] S.W. Thomsen, *Some evidence concerning the genesis of Shannon's information theory*, Studies History Philos. Sci. 40 (2009), pp. 81–91.
[39] J.L. Doob, *Review of C. E. Shannon's 'The mathematical theory of communication'*, Math. Rev. 10 (1949), pp. 133–134.
[40] N.J.A. Sloane and A.D. Wyner (eds.), *Claude Elwood Shannon: Collected Papers*, John Wiley & Sons/IEEE, New York, 1993.
[41] C.E. Shannon, *Communication theory – Exposition of fundamentals*, IRE Trans. Inform. Theory 1(1) (1953), pp. 44–47.
[42] J. Segal, *Le zéro et le un: Histoire de la Notion Scientifique d'Information au 20e Siècle*, Éditions Syllepse, Paris, 2003.
[43] L.A. Bassalygo, R.L. Dobrushin, and M.S. Pinsker, *Kolmogorov remembered*, IEEE Trans. Inform. Theory 34(2) (1988), pp. 174–175.
[44] P.C. Shields, *The interactions between ergodic theory and information theory*, IEEE Trans. Inform. Theory 44(6) (1998), pp. 2079–2093.
[45] T. Berger, *Andrei Nikolaevich Kolmogorov*, IEEE Trans. Inform. Theory 34(2) (1988), pp. 173–174.
[46] A.N. Kolmogorov, *On the Shannon theory of information transmission in the case of continuous signals*, IRE Trans. Inform. Theory 2 (1956), pp. 102–108.
[47] A.N. Kolmogorov, *Logical basis for information theory and probability theory*, IEEE Trans. Inform. Theory 14(5) (1968), pp. 662–664.

[48] R.L. Dobrushin, *Survey of Soviet research in information theory*, IEEE Trans. Inform. Theory 18(6) (1972), pp. 703–724.

[49] D. Slepian (ed.), *Key Papers in the Development of Information Theory*, IEEE Press, Piscataway, New Jersey, 1974.

[50] H.S. Tropp, *Anecdotes, origin of the term bit*, Ann. History Comput. 6(2) (1984), pp. 152–155.

[51] C.E. Shannon, *The bandwagon (editorial)*, IRE Trans. Inform. Theory 2(1) (1956), p. 3.

[52] C.E. Shannon, *A symbolic analysis of relay and switching circuits*, Master's thesis, Massachusetts Institute of Technology, 1940.

[53] C.E. Shannon, *A symbolic analysis of relay and switching circuits*, Trans. AIEE 57(12) (1938), pp. 713–723.

[54] C.E. Shannon, *An algebra for theoretical genetics*, Ph.D. thesis, Massachusetts Institute of Technology, 1940.

[55] J.F. Crow, *Shannon's brief foray into genetics*, Genetics 159 (2001), pp. 915–917.

[56] C.E. Shannon, *Communication theory of secrecy systems*, Bell Syst. Tech. J. 28 (1949), pp. 656–715.

[57] R.G. Gallager, *Claude E. Shannon: A retrospective of his life, work, and impact*, IEEE Trans. Inform. Theory 47(7) (2001), pp. 2681–2695.

[58] G.W. Dunnington, *Carl Friedrich Gauss: Titan of Science, A Study of His Life and Work*, Exposition Press, New York, 1955.

[59] H. Nyquist, *Certain factors affecting telegraph speed*, Bell Syst. Tech. J. 3 (1924), pp. 324–346.

[60] H. Nyquist, *Certain topics in telegraph transmission theory*, Trans. AIEE 47 (1928), pp. 617–644.

[61] F.W. Hagemeyer, *Die entstehung von informationskonzepten in der nachrichtentechnik: Eine fallstudie zur theoriebildung in der technik in industrie- und kriegsforschung*, Doctoral dissertation, Free University of Berlin, Berlin, Germany, 1979, 569p.

[62] K. Küpfmüller, *Über einschwingvorgänge in Wellenfiltern*, Elektrische Nachr. 1 (1924), pp. 141–52.

[63] K. Küpfmüller, *Über beziehungen zwischen frequenzcharakteristiken und ausgleichsvorgängen in linearen systemen*, Elektrische Nachr. 5(1) (1928), pp. 18–32.

[64] C. Bissell, *Karl Küpfmüller, 1928. An early time-domain, closed-loop, stability criterion*, IEEE Control Syst. Mag. 26(3) (2006), pp. 115–116.

[65] F. Lüschen, *Modern communication systems*, J. IEE 71 (1932), pp. 276–98 (Lecture of 8 April 1932 at an IEE meeting in London; German version: Moderne Nachrichtensysteme, Europäischer Fernsprech Dienst 39 (1932), pp. 171–182).

[66] J.R. Carson, *Notes on the theory of modulation*, Proc. IRE 10(1) (1922), pp. 57–64.

[67] D. Gabor, *Theory of communication*, J. IEE 93(3) (1946), pp. 429–441.

[68] M. Tribus and E.C. McIrvine, *Energy and information (thermodynamics and information theory)*, Sci. Am. 225(3) (1971), pp. 179–188.

[69] S.J. Muller, *Asymmetry: The Foundation of Information*, Springer-Verlag, Berlin, 2007.

[70] L. Szilard, *Über die Entropieverminderung in einem thermodynamischen System bei Eingriffen intelligenter Wesen*, Zeitschr. Phys. 53 (1929), pp. 840–856 (English translation 'On the decrease of entropy in a thermodynamic system by the intervention of intelligent beings' published in Behavioral Science 9 (1964), pp. 301–310; also in [22, pp. 110–119] and [139, pp. 539–548]).

[71] L. Brillouin, *Science and Information Theory*, 2nd ed., Academic Press, New York, 1962.

[72] M. Li and P. Vitányi, *An Introduction to Kolmogorov Complexity and its Applications*, 2nd ed., Springer, New York, 1997.

[73] D. Petz, *Entropy, von Neumann and the von Neumann entropy*, in *John von Neumann and the Foundations of Quantum Physics*, M. Rédei and M. Stöltzner, eds., Kluwer Academic Publishers, Dordrecht, 2001, pp. 83–96.

[74] D. Petz, *Quantum Information Theory and Quantum Statistics*, Springer-Verlag, Berlin, 2008.

[75] J. von Neumann, *Mathematische Grundlagen der Quantenmechanik*, Springer, Berlin, 1932 (English translation: Mathematical foundations of Quantum Mechanics, Dover, New York, 1954; Princeton University Press, Princeton, 1955).

[76] J. von Neumann, *Theory of Self-Reproducing Automata*, University of Illinois Press, Urbana, 1966 (edited and completed by Arthur W. Burks).

[77] W.G. Tuller, *Theoretical limitations on the rate of transmission of information*, Ph.D. thesis, Massachusetts Institute of Technology, Cambridge, MA, 1948.

[78] W.G. Tuller, *Theoretical Limitations on the rate of transmission of information*, Proc. IRE 37(5) (1949), pp. 468–478.

[79] E.T. Whittaker, *On the functions which are represented by the expansions of the interpolation-theory*, Proc. Roy. Soc. Edinburgh 35 (1915), pp. 181–194.

[80] J.R. Higgins, *Five short stories about the cardinal series*, Bull. Am. Math. Soc., New Ser. 12(1) (1985), pp. 45–89.

[81] C.J. de la Vallée Poussin, *Sur la convergence des formules d'interpolation entre ordonnées équidistantes*, Bull. Acad. Roy. de Belgique, pp. 319–410.

[82] P.L. Butzer and R.L. Stens, *De La Vallée Poussin's article of 1908 on interpolation and sampling theory, and its inuence*, in *Ch. J de La Vallée Poussin: Collected Works. Volume III, Approximation Theory, Fourier Analysis, Quasi-analytic Functions*, P.L. Butzer, J. Mawhin, and P. Vetro, eds., Circolo Matematico di Palermo, Palermo, 2004, pp. 421–453, xxv+611pp.

[83] J.R. Higgins, *Two basic formulae of Euler and their equivalence to Tschakalov's sampling theorem*, Sampl. Theory Signal Image Process. 2(3) (2003), pp. 259–270.

[84] V.A. Kotel'nikov, *On the Carrying Capacity of the 'Ether' and Wire in Telecommunications. Material for the First All-Union Conference on Questions of Communications*, Izd. Red. Upr. Svyazi RKKA, Moscow, 1933 (Russian).

[85] J.J. Benedetto and P.J.S.G. Ferreira (eds.), *Modern Sampling Theory: Mathematics and Applications*, Birkhäuser, Boston, 2001.

[86] A.N. Kolmogorov and W.M. Tichomirow, *Arbeiten zur Informationstheorie III: ϵ-Entropie und ϵ-Kapazität von Mengen in Funktionalräumen*, Mathematische Forschungsberichte X, VEB Deutscher Verlag der Wissenschaften, Berlin, 1960.

[87] H. Ogawa, *Sampling theory and Isao Someya: A historical note*, Sampl. Theory Signal Image Process. 5(3) (2006), pp. 247–256.

[88] N. Wiener, *Extrapolation, Interpolation, and Smoothing of Stationary Time Series with Engineering Applications*, John Wiley & Sons, New York, 1949.

[89] S.O. Rice, *Mathematical analysis of random noise*, Bell Syst. Tech. J. 23 (1944), pp. 282–332.

[90] S.O. Rice, *Mathematical analysis of random noise*, Bell Syst. Tech. J. 24 (1945), pp. 46–156.

[91] E. Borel, *Sur l'interpolation*, C. R. Acad. Paris 124 (1897), pp. 673–676.

[92] J.M. Whittaker, *The 'Fourier' theory of the cardinal function*, Proc. Edinburgh Math. Soc. 1 (1929), pp. 169–176.

[93] B.M. Oliver, J.R. Pierce, and C.E. Shannon, *The philosophy of PCM*, Proc. IRE 36(11) (1948), pp. 1324–1331.

[94] J.F. Price, *Minimum conditions for the sampling theorem*, preprint, c. 1985.

[95] J.M. Whittaker, *Interpolatory Function Theory*, Cambridge University Press, Cambridge, 1935.

[96] F. Smithies, *The shaping of functional analysis*, Bull. London Math. Soc. 29(2) (1997), pp. 129–138.

[97] S. von Weiher, *Männer der Funktechnik, 70 Lebenswerken Deutscher Pioniere von Funk, Rundfunk und Fernsehen*, VDE Verlag, Berlin, 1983.

[98] S. Verdú, *Fifty years of Shannon theory*, IEEE Trans. Inform. Theory 44(6) (1998), pp. 2057–2078.
[99] G.I. Khudiakov, *Sampling theorem in signal theory and its originators*, J. Commun. Technol. Electr. 53(9) (2008), pp. 1096–1106.
[100] A.R. Calderbank, *The art of signaling: Fifty years of coding theory*, IEEE Trans. Inform. Theory 44(6) (1998), pp. 2561–2595.
[101] D.J. Costello Jr and G.D. Forney Jr, *Channel coding: The road to channel capacity*, Proc. IEEE 95(6) (2007), pp. 1150–1177.
[102] W.R. Bennett, *Introduction to Signal Transmission*, McGraw-Hill, Inc., New York, 1970.
[103] H. Raabe, *Erinnerungen an meine Studien – Forschungs – und Lehrtaetigkeit an der Technischen Hochschule Berlin*, 8pp., typewritten manuscript sent 20 December 1994 to Prof. Peter Noll.
[104] W.R. Bennett, *Spectra of quantized signals*, Bell Syst. Tech. J. 27 (1948), pp. 446–472.
[105] W.R. Bennett and J.R. Davey, *Data Transmission*, McGraw-Hill, Inc., New York, 1965.
[106] W.R. Bennett, *Electrical Noise*, McGraw-Hill, New Work, 1960.
[107] S. Rice, *In memoriam: W. R. Bennett – Some early memories*, IEEE Trans. Commun. 32(2) (1984), p. 109.
[108] W. Mathis and P. Noll (eds.) *2.ITG-Diskussionsitzung: Neue Anwendungen Theoretischer Konzepte in der Elektrotechnik mit Gedenksitzung zum 50. Todestag von Wilhelm Cauer* (Meeting in Berlin, April 1995); VDE-Verlag, Berlin, 1996 (Also treats Electrical Engineering at the TH Berlin).
[109] S.R. Weart and G. Weiss-Szilard, *Leo Szilard: His Version of the Facts, Selected Recollections and Correspondence*, The MIT Press, Cambridge, MA, 1978.
[110] S. McCartney, *ENIAC: The Triumphs and Tragedies of the World's First Computer*, Walker and Company, New York, 1999.
[111] P.E. Ceruzzi, *A History of Modern Computing*, 2nd ed., The MIT Press, Cambridge, MA, 2003.
[112] S. Ulam, *John von Neumann 1903–1957*, Bull. Am. Math. Soc. 64(3, Part 2) (1958), pp. 1–49.
[113] C.E. Shannon, *Von Neumann's contributions to automata theory*, Bull. Am. Math. Soc. 64(3, Part 2) (1958), pp. 123–129.
[114] W. Splettstößer, *Über die approximation stetiger Funktionen durch die klassischen und durch die neuen Abtastsummen mit Fehlerabschätzungen*, Doctoral dissertation, RWTH Aachen, Aachen, 1977.
[115] R.L. Stens, *Error estimates for sampling sums based on convolution integrals*, Inform. Control 45(1) (1980), pp. 37–47.
[116] S. Ries and R.L. Stens, *Approximation by generalized sampling series*, in *Constructive Theory of Functions*, B. Sendov, P. Petrushev, R. Maleev, and S. Tashev, eds., Publishing House of the Bulgarian Academy of Sciences, Sofia, 1984, pp. 746–756.
[117] P.L. Butzer, W. Engels, S. Ries, and R.L. Stens, *The Shannon sampling series and the reconstruction of signals in terms of linear, quadratic and cubic splines*, SIAM J. Appl. Math. 46(2) (1986), pp. 299–323.
[118] P.L. Butzer, W. Splettstößer, and R.L. Stens, *The sampling theorem and linear prediction in signal analysis*, Jahresber. Deutsch. Math.-Verein. 90(1) (1988), pp. 1–70.
[119] R.L. Stens, *Sampling with generalized kernels*, in *Sampling Theory in Fourier and Signal Analysis. Vol. 2: Advanced Topics*, J.R. Higgins and R.L. Stens, eds., Oxford University Press, Oxford, 1999, pp. 130–157.
[120] P.L. Butzer, G. Schmeisser, and R.L. Stens, *An Introduction to Sampling Analysis*, in *Nonuniform Sampling: Theory and Practice*, F. Marvasti, ed., Kluwer Academic Publishers, New York, 2001, pp. 17–121.
[121] P.L. Buzer, W. Splettstößer, and R.L. Stens, *Index of papers on signal analysis 1972–1994*, Lehrstuhl A für Mathematik, RWTH Aachen University, Aachen, 1994.

[122] R.V.L. Hartley, *A more symmetrical Fourier analysis applied to the transmission problems*, Proc. IRE 30(3) (1942), pp. 144–150.
[123] W.S. McCulloch, *Recollections of the many sources of cybernetics*, in *ASC Forum Volume VI*, Number 2, p. 2ff, 1974.
[124] E. Feldtkeller and H. Goetzeler (eds.), *Pioniere der Wissenschaft bei Siemens*, Publicis MCD Verlag, Berlin/München, 1994.
[125] W. Endres. *NTZ-Interview: Prof. Dr.-Ing. E.h. Karl Küpfmüller 80 Jahre alt.*, NTZ 30 (1977), pp. 774–776.
[126] K. Küpfmüller, W. Mathis, and A. Reibiger, *Theoretische Elektrotechnik: Eine Einführung 18. Auage*, 18th ed., Springer, Berlin, 2008.
[127] O. Blumtritt, *Lüschen, Friedrich (Fritz) Heinrich, nachrichtentechniker*, in *Neue Deutsche Biographie*, Vol. 15, Bayerische Akademie der Wissebshaften, eds., Duncker & Humblot, Berlin, 1987, pp. 475–476.
[128] F. Strecker, *Nachrichten und Frequenzbreite*, Veröffentlichungen aus d. Geb. d. Nachr. Tech. Siemens & Halske 5(4) (1935), p. 227.
[129] R. Feldtkeller, *Felix Strecker*, Arch. Elektr. Übertragung (AEÜ) 6 (1952), p. 46.
[130] K. Küpfmüller, *F. Strecker*, Elektrotechn. Zeitschrift (ETZ) 73 (1952), p. 251.
[131] F. Strecker, *Über die Ortskurven der Scheinwiderstaende elektrischer netzwerke in abhaengigkeit von der Frequenz*, Wissensch. Veroeffentl. d. Siemens–Konzerns 6 (1928), pp. 68–105.
[132] F. Strecker, *Praktische Stabilitaetspruefung Mittels Ortskurven und Numerische Verfahren*, Springer, Berlin, 1950.
[133] H. Nyquist, *Regeneration theory*, Bell Syst. Tech. J. 11 (1932), pp. 126–147.
[134] F. Strecker, *Die Elektrische Selbsterregung*, Hirzel, Stuttgart, 1947.
[135] C.C. Bissell, *Karl Küpfmüller, a German contributor to the early development of linear systems theory*, Int. J. Control 44(4) (1986), pp. 977–989.
[136] V.A. Kotel'nikov, *The Theory of Optimum Noise Immunity*, McGraw-Hill, New York, 1959.
[137] T. Kailath and H.V. Poor, *Detection of stochastic processes*, IEEE Trans. Inform. Theory 44(6) (1998), pp. 2230–2259.