Arch. Hist. Exact Sci. (2011) 65:437–470
DOI 10.1007/s00407-011-0082-5

Tubes, randomness, and Brownian motions: or, how engineers learned to start worrying about electronic noise

Chen-Pang Yeang

Received: 14 January 2011 / Published online: 24 April 2011

Abstract In this paper, we examine the pioneering research on electronic noise—the current fluctuations in electronic circuit devices due to their intrinsic physical characteristics rather than their defects—in Germany and the U.S. during the 1910s–1920s. Such research was not just another demonstration of the general randomness of the physical world Einstein's work on Brownian motion had revealed. In contrast, we stress the importance of a particular engineering context to electronic noise studies: the motivation to design and improve high-gain thermionic-tube amplifiers for radio and wired communications. Engineering scientists' endeavors to understand electronic noise started in 1918, when Walter Schottky at Siemens formulated a theory of "shot noise," current fluctuations owing to the random emissions of discrete electrons in a tube. Schottky's theory was revised and experimentally tested at Siemens, General Electric, and AT&T during the 1920s, leading to the discoveries of several other types of noise and an increasing interest in the thermal fluctuations in electronic circuits. In 1925–1928, J.B. Johnson and Harry Nyquist at Bell Labs developed a theory of thermal noise for any electrical resistor at a non-zero temperature. Although these studies were initiated to chart the fundamental performance limit of electronic technology, they ended up assisting the empirical determination of individual electronic components' characteristics.

Communicated by Jed Buchwald.

C.-P. Yeang (✉)
Institute for the History and Philosophy of Science and Technology, University of Toronto,
Room 309, 91 Charles Street West, Toronto, ON M5S 1K7, Canada
e-mail: chenpang.yeang@utoronto.ca

1 Introduction

Very few scientific works at the turn of the twentieth century were as significant as the study of Brownian motion. In both popular and scholarly literatures, Einstein's 1905 paper (and Marian Smoluchowski's independent article a year later) on this subject provided a novel rationale for the existence of atoms, leading to Jean Perrin's provision of experimental evidence for such otherwise conjectural objects. It marked as well the onset of a statistical mechanics that focused on unstable phenomena and sought programmatically to compute the macroscopic effects of microscopic characteristics. And it laid the foundation for the mathematics of stochastic processes that would come to play a crucial part in realms ranging from pricing financial commodities and designing gun-control systems to modeling biological evolution and predicting climate change.[1]

Despite the extensive literature on the history of Brownian motion, one of its important implications has rarely been addressed: the understanding and effects of electronic noise. Brownian motion, after all, was not only about the esoteric behavior of pollen dusts under a microscope, or even solely about the reformulation of statistical mechanics, the provision of evidence for atomism, or the production of new mathematical methods. It had increasingly salient implications as well for the practical world of telecommunications. As thermionic tubes became more and more indispensable to high-fidelity cabled telephone and radio systems, engineers endeavored to improve the performance of such electronic devices and to explore the ultimate limit of their performance. In the 1910s–1930s, researchers in Germany and the U.S. investigated the Brownian-like fluctuations of microscopic charge carriers in electronic tubes and associated circuit components in efforts to understand and to quantify the disturbances such fluctuations produced in the transmission of voice signals, whose frequencies ranged from several hertz to tens of kilohertz. "Schottky" (shot) noise and "Johnson" (thermal) noise were the best known products of these efforts.

Günter Dörfel and Dieter Hoffmann's work is the only detailed historical study of electronic noise. They examined the theoretical reasoning and experimental findings that led to the formulation of Schottky and Johnson noise, and situated such studies in the context of attempts to come to grips with the fundamental limit of measurement.[2] According to Dörfel and Hoffmann, some physicists in the 1900s–1920s considered electronic noise to be intrinsic fluctuations of electrical currents, similar in their view to the Brownian motion of pollen dust. From Einstein's schematic electrostatic electrometer for the amplification of fluctuating electricity and Geertruida de Haas-Lorentz's analysis of the effect of thermal agitation on metrology to Frits Zernike's and Gustav Ising's examinations of galvanometer's sensitivity, electronic noise seemed to impose an ultimate constraint on the accuracy of any electrical measuring devices. The thermal agitations of galvanometers, electrometers, or any other measuring devices at non-zero temperature unavoidably reduced their sensitivity and thus restricted the minimum level of signals they could measure. Before the introduction of the uncertainty principle and the notion of quantum indeterminancy, Brownian

[1] See, for example, Bigg (2008), von Plato (1998, pp. 123–136), Maiocchi (1990), Stachel et al. (1990, pp. 206–222), Bigg (2005), Cohen (2005).

[2] Dörfel and Hoffmann (2005).

noise was widely conceived of as the fundamental barrier to the exactness of measurement.[3]

Electronic noise indeed posed challenges in the early twentieth century as scientists attempted to grapple with its implications for their ways of handling issues concerning accuracy and precision. Although the discussions concerning these challenges constituted an important aspect of noise research during this period, another equally significant aspect was the specific engineering context within which such research proceeded. The discoveries of and early research on electronic noise were done at industrial laboratories—Siemens's "K Laboratory," AT&T Bell Laboratories, and General Electric's Schenectady Laboratory. And the purpose of these endeavors was to improve the quality of the thermionic tubes and associated electronic circuits that were just becoming prevalent in radio and long-range telephonic systems. After the invention of feedback amplifiers in the 1910s, vacuum tube circuits were designed to facilitate significant amplification of minute electrical signals with minor distortion.[4] As weaker and weaker signals could thereby be detected, the previously inconsequential "background noise" became an increasing nuisance, and hence demanded a closer examination. In consequence, the studies of shot noise and thermal noise in the 1910s–1930s concerned either the characteristics of thermionic tubes or the functionality of other electrical devices in tube circuits. And here we find that the engineering scientists were not concerned to single out one noise-producing factor over another as worthy of particular understanding from a fundamental standpoint. To them, no particular noise source was in and of itself considered to be any more "fundamental" than a plethora of other sources of uncertainties that might corrupt signals, for that was their sole concern—the alleviation of disturbances that corrupted signal transmission. And so, in that context, the intrinsically stochastic character of Brownian motion was just one factor—albeit an important one—among the several that generated electronic noise. Other, system-dependent factors, such as the aggregation of electric charge at different degrees of vacuum in thermionic tubes also played significant roles, especially in the case of shot noise.

This article explores precisely that engineering context in the origin of electronic noise research. Like Dörfel and Hoffmann, we will also examine Schottky's, Johnson's, and Nyquist's works (albeit at a finer scale). In contrast with these authors, however, we will also pay close attention to the experimental and technological developments that revealed the particularity of electronic noise. In Section 2, we examine the German physicist Walter Schottky's intellectual route to the theory of the "Schroteffekt," or shot noise. Section 3 gives analyses of the aftermath of Schottky's theory, including empirical testing of the theory and the discovery of a new kind of noise, the flicker effect, as researchers tried to identify the physical conditions that produced shot noise. Section 4 concerns American scientists John B. Johnson and Harry Nyquist's work at the AT&T Bell Labs, work that led to the claim concerning the inevitability of thermal

[3] Ibid., pp. 12–17. Also see Beller (1999, pp. 92–93).

[4] For the invention of feedback amplifiers, see Mindell (2000).

noise in every electrical resistor at a non-zero temperature. Section 5 turns briefly to the general engineering treatment of noise by the 1930s.[5]

2 Walter Schottky's route to the "Schroteffekt"

Engineers had known about electrical circuits' fluctuating performance for some time. Their hardware was never perfect enough. The tube amplifier, a still immature technology in the 1910s, was especially problematic. Fluctuating battery output disturbed a tube's bias voltage. Variations of the filament temperature affected an electronic current's uniformity. A tube wall's bad insulation created charge leakage. Casual design of supporting containers caused unexpected mechanical coupling and resonance. These factors were not essentially different from the ways in which defective devices degraded engineering products. Yet did the defective-device factors exhaust performance uncertainty? Or did tube amplifiers have a fundamental performance limit determined by the basic nature of electrons, not by malfunctions of design and manufacturing? The German physicist Walter Schottky was one of the first to ask these questions.

2.1 Schottky's intellectual background

Son of German mathematician Friedrich Hermann Schottky, Walter Hans was born in Zurich, Switzerland, and moved to Berlin when his father was appointed professor at the Friedrich-Wilhelm University (the University of Berlin). Walter studied physics, took courses with Max Planck and Walter Nernst, and chose Planck to supervise his Ph.D. thesis. His research was on the dynamics of the electron near the relativistic limit, which was consistent with Planck's general interest to reconcile electrodynamics with special relativity. He defended his thesis in 1912.[6]

After receiving his Ph.D., Schottky was tired of theory and switched to experimental physics "as remedy for pure abstraction." He moved to Jena for postgraduate study with Max Wien, who was experimenting on electrons. Schottky's first task as Wien's assistant was to measure electrical currents on light-illuminated metal surfaces. In 1877, when Heinrich Hertz was experimenting with the spark gaps that would lead to the famous discovery of electromagnetic waves, he found that the intensity of the induced spark at the gap of the receiving coil changed with the light influx on the coil. Dubbed as the photoelectric effect, this phenomenon was closely studied in the following decade. By the turn of the century, it was widely believed that the photoelectric effect occurred when light drove the electrons in a piece of metal into motions, which provided sufficient energy for the electrons to escape from the metal and consequently created an electric current in the space surrounding the metal. An important empirical law from this study was so-called Stoletov's law (à la the Russian physicist

[5] Dörfel and Hoffmann examined Schottky's, Johnson's, and Nyquist's works, which we will also do, albeit with further detail. Our primary aim is however to focus on related experimental and technological developments concerning the nature of electronic noise.

[6] Madelung (1986) (also in Schottky Papers), pp. 1–2; Serchinger (2000, p. 172).

Aleksandr Stoletov) that the strength of the photoelectric current was proportional to the intensity of the illuminating light.[7]

Schottky's experimental work with Wien contradicted the Stoletov's law, however: the measured current remained weak regardless of the illuminating light intensity, which should not have been the case. This result implied that the metal's electrons were prevented from leaving the metal surface to the air, and he conjectured that the space charge, the electric charge in air (induced by the emitted electrons) surrounding the metal surface created a potential barrier against further free electronic motions. Based on this hypothesis, he derived his "$U^{3/2}$ law"—the space-charge-limited current was proportional to the 3/2 power of the voltage across the two metallic plates and to the inverse square of the plates' separation, i.e., $I \propto V^{3/2}/d^2$ ("U" was Schottky's notation for voltage). He quickly found that the space-charge effect did not pose a limit to photoelectric currents at room temperature. But it was the dominant factor for currents emitted from heated cathodes, which meant that the $U^{3/2}$ law was applicable to electronic tubes. As a result, he proceeded to investigate vacuum tubes. He measured the tube current and voltage with different cathode materials and filament temperatures. The voltage–current curves from his measurements were consistent with the $U^{3/2}$ law, verifying his theory of the space-charge effect. Unfortunately, Schottky was not the first to discover this. Irving Langmuir at General Electric and Clement Child at Cornell University had independently found and published the same law in 1913–1914. Schottky's paper came months later.[8]

Schottky's experience with the $U^{3/2}$ law shaped his future career on the studies of electronic tubes, semiconductors, and physical chemistry, leading to a conjunction in his work of basic and applied research. In 1914, he returned to Berlin to continue electronic-tube experiments at the University of Berlin. He presented his work on the $U^{3/2}$ law in 1915 at the university's physics colloquium. The presentation impressed a Swedish physicist Ragnar Holm, who was attaché at the university from the *Siemens & Hauske Aktiengesellschaft*—one of Germany's largest electric manufacturers and pioneers of industrial research. (Siemens had sent research staff to scientific conferences and college colloquia to build connections with academia.) Holm invited Schottky to conduct research for Siemens. Like many German physicists at the time, Schottky was eager to help the nation win the ongoing World War with more "useful" work. Working for Siemens was thus preferable to pure scholarly research. Though still at the University of Berlin, he received funding from the company to build a laboratory for vacuum-tube experiments. Reckoning, however, on the increasing demand for military technology, he left the university in 1916 to work directly at Siemens Research and Development. In 1917, he assumed directorship of the "K Laboratory" of Siemens's general research establishment (the *Werner-Werks für Bauelemente*).[9]

Located in Charlottenburg near Berlin, the K Laboratory was Siemens's main facility for the research and development of electrical devices used in signal transmission

[7] Stoletov (1888).

[8] Ibid, 173–174; Madelung (1986, pp. 4–5); Schottky (1962) (also in Schottky Papers). The $U^{3/2}$-law was also known as the Langmuir-Child equation.

[9] Serchinger (2000, pp. 174–176); Schottky, "Lebenslauf," (undated) and "Übersicht über meine wissenschaftlichen und technischen Untersuchungen," September 1948, in Schottky Papers.

such as loading coils, vacuum tubes, and amplifiers. (K denoted *Kabel*, the German word for *cable*.) At the K Laboratory, Schottky led a team to develop screen-grid tubes, to invent superheterodyne reception (independent of the American inventor Edwin Armstrong who usually receives sole credit for its invention), to experiment on vacuum tubes with multiple grids, and to investigate the properties of crystal detectors' metal-semiconductor contacts. These were critically relevant to improving German military communications.[10]

The motivations for developing screen-grid tubes and heterodyne reception were to increase amplifier gain and to reduce external noise. Both concerns led Schottky to a question—is there a limit of amplification due not to individual device characteristics but to the fundamental physical principles governing the operations of all electronic devices? Building on his background in statistical mechanics, Schottky proposed a theory for the fundamental limit of electronic devices in 1918.[11]

2.2 Shot noise in a resonating tube circuit

Schottky began by noting that two kinds of noise due to electrons' fundamental characteristics posed limits to tube amplification: (i) electrons move randomly at any non-zero temperature; (ii) electrons are discrete. The first feature causes a random fluctuation in a supposedly constant current—"thermal noise" (*Wärmeeffekt*). The second feature implies that a current is a discrete series of surges, not a continuous flow, and thus also causes a fluctuation in an apparently constant current—"shot noise" (*Schroteffekt*).[12] Schottky decided to focus on shot noise. An electron flow's discrete characters are most conspicuous in a high-vacuum diode tube in which all electrons emitted from the cathode arrive at the anode. Schottky modeled the number of emitted electrons within a time duration τ as a random process $n(\tau)$. When the tube current is apparently constant, the average number of electrons per unit time is a constant N. Consequently, the electron-number deviation from its average is $\Delta n(\tau) = n(\tau) - N\tau$, also a random process depending on how electrons are emitted. Schottky assumed that the emitted electrons were uncorrelated with one another. Mathematicians had known that a random process with this property had a Poisson distribution. Hence $\Delta n(\tau)$ must follow a Poisson process with zero mean and standard deviation $N\tau : \langle \Delta n(\tau) \rangle = 0, \langle [\Delta n(\tau)]^2 \rangle = N\tau$.[13]

The number $n(\tau)$ determines the tube's electric current. Since the number of electrons emitted from the cathode equals that arriving at the anode, the tube current within period τ is $i(\tau) = en(\tau)/\tau$ ($e = 1.6 \times 10^{-19}$ C is the charge of electron), and

[10] Serchinger (2000, pp. 176–177); Madelung (1986, pp. 6–7).

[11] Schottky (1918, pp. 541–567).

[12] Ibid., pp. 541–547.

[13] Here the use of the term "random process" may be anachronistic, as the more comprehensive mathematical theory of random processes was not formulated until Andrei Kolmogorov's work in the 1930s. But the point here does not concern rigorous mathematical definitions or properties such as the existence of the probability density function over time. It instead concerns the intuitive characteristics of the random time-varying signal $\Delta n(\tau)$. In fact, that $\Delta n(\tau)$ has a probabilistic distribution at any instant and a correlation function between two instants—the basis of Schottky's theory—are pretty consistent with the attributes of the random processes in today's definition.

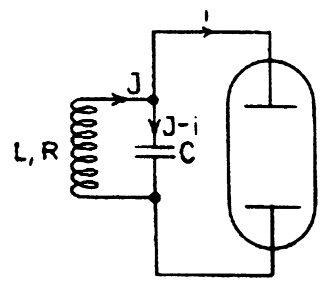

Fig. 1 Schottky's tube circuit for evaluating the shot effect (Schottky 1918, Figure 6)

accordingly $\langle[\Delta i(\tau)]^2\rangle = e^2N/\tau$. The current's instantaneous value obviously fluctuates. An operator connecting an earphone to a vacuum tube with a DC bias would hear a hissing tone caused by the fluctuation—this is shot noise. The strength of shot noise is measured by the current's standard deviation $\langle[\Delta i(\tau)]^2\rangle$. For a DC bias current $i_0 = eN$, the effective "shot current" is $i_S \equiv [\langle[\Delta i(\tau)]^2\rangle]^{1/2} = i_0(e/i_0\tau)^{1/2}$.[14]

The above formula expresses the shot current of a diode vacuum tube alone. In reality, however, a tube is rarely used without being connected to other circuit elements. Moreover, as we will see below, shot noise in Schottky's theory has a flat, frequency-independent "white" power spectrum, which leads to the unphysical outcome of infinite signal power if the bandwidth of the tube circuit is not limited. Therefore, Schottky considered the shot effect of an oscillating circuit with a diode vacuum tube in parallel with a capacitor (capacitance C) and a resistive inductor (inductance L and resistance R.) The RLC circuit forms a resonator that selects the tube's shot current at a specific range of frequencies (Fig. 1).

To evaluate the noise current intensity of the circuit, Schottky used a method based on Fourier analysis. This method resembled a technique his mentor Planck adopted in treating black-body radiation in 1897–1898. The core of such a technique was performing an average over frequency to distinguish the slowly from the fast varying elements of the Fourier series, and attributing physical meanings only to the slowly varying elements (more details below).[15]

2.3 A Fourier analysis of shot noise

Schottky first wrote down the equation linking the current i into the tube with the current J out of the resistive inductor:

$$\frac{\mathrm{d}^2J}{\mathrm{d}t^2} - \rho\frac{\mathrm{d}J}{\mathrm{d}t} + \omega_0^2 J = \omega_0^2 i$$

where $\omega_0^2 = 1/LC$ and $\rho = R/L$. If the tube current i is time harmonic with a single frequency, $i = I\sin(\omega t + \varphi)$, then the solution leads to the circuit's (time averaged) oscillating energy $E_S = (I^2L/2)/[(1 - x^2)^2 + r^2x^2](x = \omega/\omega_0$ and $r = \rho/\omega_0)$.

[14] Ibid., pp. 548–553.

[15] Ibid., pp. 554; Kuhn (1978, pp. 72–84).

In reality, however, the tube current i is the randomly varying shot current, and (like other finite-duration, time-varying, non-monochromatic signals) can be expressed as a Fourier series sampled over a long period T:

$$i = \sum_{k=0}^{\infty} i_k = \sum_{k=0}^{\infty} I_k \sin(\omega_k t + \varphi_k)$$

where $\omega_k = 2\pi k/T$. Here the amplitude I_k and phase φ_k of each oscillating mode at frequency ω_k vary randomly. As a result, the circuit's oscillating energy associated with the shot current is the sum over all the oscillating modes:

$$E_S = \sum_{k=0}^{\infty} E_k = L \sum_{k=0}^{\infty} \frac{I_k^2}{2} \frac{1}{\left(1 - x_k^2\right)^2 + r^2 x_k^2} \tag{1}$$

where $x_k = \omega_k/\omega_0 = 2\pi k/(\omega_0 T)$.

The value of E_S in Eq. 1 depends on the values of I_k^2 for $k = 0$ to ∞. According to Fourier analysis,

$$I_k = \frac{2}{T} \int_0^T i(t) \sin(\omega_k t + \varphi_k) \mathrm{d}t.$$

But since $i(t) = \langle i(t) \rangle + \Delta i(t)$, $\langle i(t) \rangle$ is a constant, and $\int_0^T \mathrm{d}t \sin(\omega_k t + \varphi_k) = 0$, the term corresponding to $\langle i(t) \rangle$ vanishes after the integral. Thus,

$$I_k = \frac{2}{T} \int_0^T \Delta i(t) \sin(\omega_k t + \varphi_k) \mathrm{d}t.$$

Merely plugging this expression of I_k into Eq. 1 helps little compute the oscillating energy E_S, for $\Delta i(t)$ is a randomly varying signal. To facilitate such computation, a certain average is required, which Schottky did.

Note that Schottky's procedure of calculating average E_S differed from today's standard practice. A student of modern probability theory would base her computation on tools like ensemble average and correlation function. To her, the average oscillating energy $\langle E_S \rangle$ is nothing but an ensemble average (with respect to the random configurations of $\{I_k^2\}$). She would move the ensemble average into the summation in Eq. 1 and turn it into $\langle E_S \rangle = L \sum_{k=0}^{\infty} \frac{\langle I_k^2 \rangle}{2} \frac{1}{(1-x_k^2)^2 + r^2 x_k^2}$, express $\langle I_k^2 \rangle$ as

$$\left\langle \left[\frac{2}{T} \int_0^T \Delta i(t) \sin(\omega_k t + \varphi_k) \mathrm{d}t \right]^2 \right\rangle$$
$$= \frac{4}{T^2} \left\langle \int_0^T \int_0^T \Delta i(t) \Delta i(t') \sin(\omega_k t + \varphi_k) \sin(\omega_k t' + \varphi_k) \mathrm{d}t \mathrm{d}t' \right\rangle,$$

exchange the order of ensemble average and integrals in this expression, replace $\langle \Delta i(t) \Delta i(t') \rangle$ with a known correlation function of the random process $\Delta i(t)$ (a Dirac delta function $I\delta(t - t')$ in the case of shot noise), and perform the integrals and summation accordingly. This is *not* what Schottky did. In contrast to the modern probability theory, his averaging procedure was carried out in the frequency domain. In other words, his average over random configurations was manifested in the form of average over Fourier spectral components. Schottky's apparently cumbersome procedure bore striking similarity to his mentor Max Planck's approach to the black-body radiation problem in 1897–1898. To understand better Planck's intellectual legacy to Schottky's 1918 paper, it is worth sketching the mentor's early treatment of black-body radiation, which Thomas Kuhn neatly reconstructed in his 1978 monograph.

2.4 Techniques from Planck's work on black-body radiation

Planck's overarching agenda in his research during 1897–1898 was to demonstrate irreversibility—the central characteristic of thermodynamics—in black-body radiation. To fulfill that aim, he considered a model with a spherical cavity (representing the black-body radiator) and a resonator sitting at the center of the cavity. The resonator served both as a portion of the black-body model and a measuring device. It had a resonating frequency ν_0 and a damping constant ρ. Like any other resonator (such as an RLC circuit), this device had a frequency response resembling a band-pass filter, whose center frequency was ν_0 and bandwidth proportional to ρ. In Planck's choice, ρ was large enough so that the resonator's bandwidth corresponded to the range of average over frequency, which was unavoidable for an actual measuring instrument. Thus, Planck's resonator was a device that generated measurable physical quantities. For this reason, he called it an "analyzing resonator."[16]

When electromagnetic waves entered and reverberated in the cavity, part of their energy was absorbed by the resonator, which then underwent motion and re-radiated secondary electromagnetic waves. Planck expressed the axial component near the resonator of the total (primary plus secondary) electric field in terms of a Fourier integral:

$$E = \int_0^\infty \mathrm{d}\nu C_\nu \cos(2\pi \nu t - \theta_\nu).$$

[16] Kuhn (1978, p. 80).

Then he was able to show that the radiation intensity J_0 (the amount of energy radiated from the entire spherical cavity per unit time) was

$$J_0 = \int \mathrm{d}\mu \left(A^0_\mu \sin 2\pi\mu t + B^0_\mu \cos 2\pi\mu t\right),$$

with

$$A^0_\mu = \frac{2}{\rho\nu_0} \int \mathrm{d}\nu C_{\nu+\mu} C_\nu \sin^2\delta_\nu \sin(\theta_{\nu+\mu} - \theta_\nu),$$
$$B^0_\mu = \frac{2}{\rho\nu_0} \int \mathrm{d}\nu C_{\nu+\mu} C_\nu \sin^2\delta_\nu \cos(\theta_{\nu+\mu} - \theta_\nu),$$

and δ_ν was defined by $\cot\delta_\nu = \pi(\nu_0^2 - \nu^2)/\rho\nu_0\nu$. In this expression, $\sin^2\delta_\nu$ represented the frequency response curve of the analyzing resonator, as it took its maximum value at $\nu = \nu_0$ and diminished quickly beyond the bandwidth 2ρ. Moreover, Planck demonstrated that the energy U_0 of the resonator was

$$U_0 = \int \mathrm{d}\mu (a_\mu \sin 2\pi\mu t + b_\mu \cos 2\pi\mu t),$$

with

$$a_\mu = \frac{3c^3}{16\pi^2\rho\nu_0^3} \int \mathrm{d}\nu C_{\nu+\mu} C_\nu \sin\delta_{\nu+\mu} \sin\delta_\nu \sin(\theta_{\nu+\mu} - \theta_\nu),$$
$$b_\mu = \frac{3c^3}{16\pi^2\rho\nu_0^3} \int \mathrm{d}\nu C_{\nu+\mu} C_\nu \sin\delta_{\nu+\mu} \sin\delta_\nu \cos(\theta_{\nu+\mu} - \theta_\nu).[17]$$

According to Planck, the Fourier components A^0_μ, B^0_μ, a_μ, b_μ in the equations for J_0 and U_0 all corresponded to measurable physical quantities, as they were the frequency components discernable by the analyzing resonator. However, they were expressed in terms of $C_{\nu+\mu} C_\nu \sin(\theta_{\nu+\mu} - \theta_\nu)$ and $C_{\nu+\mu} C_\nu \cos(\theta_{\nu+\mu} - \theta_\nu)$ determined by the amplitudes C_ν and phases θ_ν of the electric field's Fourier components, which did *not* correspond to measurable physical quantities since they fluctuated too rapidly over frequency. As Kuhn commented, the electric field's Fourier components (represented by C_ν and θ_ν) in Planck's model resembled the non-measurable microstates in Ludwig Boltzmann's kinetic theory of gas, like the positions and momenta of individual molecules, while the radiation intensity's and resonator energy's Fourier components A^0_μ, B^0_μ, a_μ, b_μ resembled the measurable macrostates in Boltzmann's theory. Planck wanted to eliminate the "unphysical" C_ν and θ_ν entirely and represented the measurable quantities A^0_μ, B^0_μ, a_μ, b_μ in terms of one another. To achieve that goal, he introduced an important approximation.

[17] Ibid., pp. 79–81.

What Planck did was to approximate the fast varying $C_{\nu+\mu}C_\nu \sin(\theta_{\nu+\mu} - \theta_\nu)$ and $C_{\nu+\mu}C_\nu \cos(\theta_{\nu+\mu} - \theta_\nu)$ with the slowly varying A^0_μ, B^0_μ. Specifically, he expressed $C_{\nu+\mu}C_\nu \sin(\theta_{\nu+\mu} - \theta_\nu)$ and $C_{\nu+\mu}C_\nu \cos(\theta_{\nu+\mu} - \theta_\nu)$ as

$$C_{\nu+\mu}C_\nu \sin(\theta_{\nu+\mu} - \theta_\nu) = A^0_\mu + \varepsilon(\nu, \mu),$$
$$C_{\nu+\mu}C_\nu \cos(\theta_{\nu+\mu} - \theta_\nu) = B^0_\mu + \eta(\nu, \mu).$$

The function ε and η captured the fast varying parts of $C_{\nu+\mu}C_\nu \sin(\theta_{\nu+\mu} - \theta_\nu)$ and $C_{\nu+\mu}C_\nu \cos(\theta_{\nu+\mu} - \theta_\nu)$. However, within the (sufficiently large) bandwidth of the resonator's frequency response curve $\sin^2 \delta_\nu$, ε, and η changed so unsystematically that their contributions from different frequency components cancelled one another. That is, $\int \varepsilon \sin^2 \delta_\nu = \int \eta \sin^2 \delta_\nu = 0$. This condition, coined by Planck as "natural radiation," guaranteed that the slowly varying physical measurables A^0_μ and B^0_μ (obtained by averaging, with the resonator's frequency response, over neighboring Fourier components of the actual field) could well approximate the fast varying quantities $C_{\nu+\mu}C_\nu \sin(\theta_{\nu+\mu} - \theta_\nu)$ and $C_{\nu+\mu}C_\nu \cos(\theta_{\nu+\mu} - \theta_\nu)$ corresponding to the Fourier components of the actual field. By replacing $C_{\nu+\mu}C_\nu \sin(\theta_{\nu+\mu} - \theta_\nu)$ and $C_{\nu+\mu}C_\nu \cos(\theta_{\nu+\mu} - \theta_\nu)$ with A^0_μ and B^0_μ in the expressions for a_μ and b_μ and employing the law of energy conservation, Planck established a differential equation relating J_0 and U_0, $dU_0/dt + K\nu_0 U_0 = (P/\nu_0)J_0$, which produced evolutions of U_0 irreversible with time.[18]

While fin-de-siècle physicists' (and Kuhn's) focus with respect to Planck's 1897–1898 papers was on the relationship between his natural radiation and Boltzmann's molecular disorder as well as Planck's clandestine introduction of statistical argument, his early work on black-body radiation nonetheless conveyed a more generic lesson: Instead of treating directly the rapidly varying Fourier components of a complicated physical quantity, we should work whenever we can on their averages over distinct windows of frequencies. In so doing, we actually smooth out the quantity's fine-scale fluctuations and transform spectral analysis into scales that are much coarser and hence more "physical" and much easier to handle.

2.5 Replacing the ensemble with spectral average

Schottky did not cite Planck in his 1918 paper (there was no citation in the article), but the Siemens researcher's approach followed closely the aforementioned lesson. Returning to the calculation of the shot current's energy E_S in the resonating circuit, the summation in Eq. 1 represented a Fourier analysis. The energy spectrum in this expression contained a fast changing factor (with respect to frequency index k, not time) of I_k^2 and a slowly changing factor $L/2[(1 - x_k^2)^2 + r^2x_k^2]$. (The random and independently arriving pulses of the shot current made I_k^2 fluctuate rapidly and spread over the entire spectrum; by contrast, $L/2[(1 - x_k^2)^2 + r^2x_k^2]$ was a smooth band-pass curve with maximum at $x_k = 1$ and bandwidth proportional to r.) Following Planck's

[18] Ibid., pp. 80–84.

generic lesson, Schottky treated these two factors separately by first averaging I_k^2 over finer windows and then performing the summation over a coarser scale. Specifically he partitioned the set of non-negative integers $k = 0$ to ∞ into disjoint and consecutive windows $[k_1(1), k_2(1)]$, $[k_1(2), k_2(2)], \ldots, [k_1(n), k_2(n)], \ldots$ Within the nth window, the factor $L/2[(1-x_k^2)^2+r^2x_k^2]$ varied so little with k that it could be approximated by $L/2[(1-x_{k0(n)}^2)^2+r^2x_{k0(n)}^2]$, where $k0(n)$ was an integer between $k_1(n)$ and $k_2(n)$. (A natural partition of this kind would be $[0, K-1], [K, 2K-1], \ldots, [(n-1)K, nK-1], \ldots.$, while the representative numbers $k0(1), k0(2), \ldots, k0(n), \ldots$ could be the mid-points of the intervals, i.e., $K/2, 3K/2, \ldots, (n-1/2)K, \ldots$) Consequently, Eq. 1 could be approximated as follows:[19]

$$E_S = \frac{L}{2}\sum_{k=0}^{\infty}\frac{I_k^2}{\left(1-x_k^2\right)^2+r^2x_k^2} = \frac{L}{2}\sum_{n=1}^{\infty}\sum_{k=k_1(n)}^{k_2(n)}\frac{I_k^2}{\left(1-x_k^2\right)^2+r^2x_k^2}$$
$$\cong \frac{L}{2}\sum_{n=1}^{\infty}\frac{1}{\left(1-x_{k0(n)}^2\right)^2+r^2x_{k0(n)}^2}\sum_{k=k_1(n)}^{k_2(n)} I_k^2$$

The sum $\sum_{k=k_1(n)}^{k_2(n)} I_k^2$ in this expression was supposed to average out the rapid fluctuations of I_k^2. For this to be effective, the size of the window should be large enough to contain a sufficient number of terms; i.e., $k_2(n)-k_1(n) \gg 1$. On the other hand, the size of the window should be small enough—specifically, $[k_2(n)-k_1(n)]/x_{k0(n)} \ll 1$—so that the function $1/[(1-x_k^2)^2+r^2x_k^2]$ varied little from $k = k_1(n)$ to $k = k_2(n)$.

To evaluate the sum $\sum_{k=k_1(n)}^{k_2(n)} I_k^2$, Schottky plugged in the relation $I_k = \frac{2}{T}\int_0^T \Delta i(t)\sin(\omega_k t+\varphi_k)\mathrm{d}t$ and accordingly expressed it in terms of a double integral:

$$\sum_{k=k_1}^{k_2} I_k^2 = \frac{4}{T^2}\sum_{k=k_1}^{k_2}\int_0^T\int_0^T \Delta i(t)\Delta i(t')\sin(\omega_k t+\varphi_k)\sin(\omega_k t'+\varphi_k)\mathrm{d}t\mathrm{d}t'$$

(Here the explicit denotation of k_1's and k_2's dependence on n is removed for convenience.) Then he approximated the double integral with a discrete double Riemann sum with short time increments Δt and $\Delta t'$. To justify this approximation, he argued that the resonating circuit's frequency response curve $L/2[(1-x_k^2)^2+r^2x_k^2]$ suppressed any high-frequency contribution from the double integrand, while the integrand's low-frequency components behaved like constants within the time windows Δt and $\Delta t'$, as long as they were much shorter than the circuit's resonating period. Thus, the double

[19] Schottky (1918, p. 556).

integral became[20]

$$\int_0^T\int_0^T \Delta i(t)\Delta i(t')\sin(\omega_k t+\varphi_k)\sin(\omega_k t'+\varphi_k)\mathrm{d}t\mathrm{d}t'$$
$$\cong \sum_{t=0}^{T}\sum_{t'=0}^{T}\Delta i(t)\Delta t\Delta i(t')\Delta t'\sin(\omega_k t+\varphi_k)\sin(\omega_k t'+\varphi_k)$$

where t and t' in the Riemann sum took discrete values between 0 and T.

To evaluate the above double Riemann sum, Schottky grouped the terms in the sum into two clusters—those with identical time indices ($t = t'$) and those with different time indices ($t \neq t'$). The first cluster's sum is

$$\sum_{t=0}^{T}[\Delta i(t)\Delta t]^2\sin^2(\omega_k t+\varphi_k).$$

The modern approach to evaluate this sum would be to calculate the ensemble average of the shot current's square to obtain $\langle[\Delta i(t)\Delta t]^2\rangle = ei_0\Delta t$, and to plug the value back into the sum. Schottky did not do this. Instead of averaging over distinct configurations of the shot current, he worked on a single time series extended over a very long period T, and performed average over time based on an implicit ergodic assumption. Schottky observed that in the sum the factor $\sin^2(\omega_k t+\varphi_k)$ is periodic in time. Thus, within the very long period T, each distinct value of $\sin^2(\omega_k t+\varphi_k)$ occurs many times in the sum. Group the terms with an equal value of $\sin^2(\omega_k t+\varphi_k)$(say, $= \sin^2(\omega_k t_0 + \varphi_k)$) together, it leads to $\sin^2(\omega_k t_0+\varphi_k)\{[\Delta i(t_0)]^2+[\Delta i(t_1)]^2+[\Delta i(t_2)]^2+\ldots\}\Delta t^2$, where t_0, t_1, t_2 are the instants at which $\sin^2(\omega_k t+\varphi_k)$ has the same value. Albeit with a coarser resolution, the sum $\{[\Delta i(t_0)]^2+[\Delta i(t_1)]^2+[\Delta i(t_2)]^2+\ldots\}\Delta t^2$ can be considered as a kind of time average. Under the ergodic assumption, therefore, it equals $\langle[\Delta i(t)\Delta t]^2\rangle = ei_0\Delta t$ for shot current. Thus, the first cluster of the Riemann sum becomes

$$\sum_{t=0}^{T}[\Delta i(t)\Delta t]^2\sin^2(\omega_k t+\varphi_k) = ei_0\Delta t\sum_{t=0}^{T}\sin^2(\omega_k t+\varphi_k) = ei_0 T/2.$$

Since this result is independent of k, its contribution to $\sum_{k=k_1(n)}^{k_2(n)} I_k^2$ is $(2/T)(k_2 - k_1)ei_0$.[21]

[20] Ibid., p. 557.

[21] Ibid., p. 558.

The Riemann sums' second cluster is

$$\sum_{t=0}^{T} \sum_{\substack{t'=0 \\ t' \neq t}}^{T} \Delta i(t)\Delta t \Delta i(t')\Delta t' \sin(\omega_k t + \varphi_k)\sin(\omega_k t' + \varphi_k).$$

Since t and t' represent distinct, non-overlapping time intervals, Schottky asserted that the shot current $\Delta i(t)$ and $\Delta i(t')$ within these intervals are uncorrelated to each other, due to the mutually independent nature of randomly arriving shot electrons. Again, the modern approach would plug the correlation $\langle \Delta i(t)\Delta i(t')\rangle = 0$ into the sum and obtain zero result. Although Schottky also aimed to get this result, his reasoning, without resorting to direct ensemble average, was more strenuous. He tackled the problem from the observation that unlike the terms $[\Delta i(t)\Delta t]^2 \sin^2(\omega_k t + \varphi_k)$ in the first cluster, the terms $\Delta i(t)\Delta t \Delta i(t')\Delta t' \sin(\omega_k t + \varphi_k)\sin(\omega_k t' + \varphi_k)$ in the second cluster are incoherent and do not exhibit any clear order. In fact, Schottky argued, because $\Delta i(t)$ and $\Delta i(t')$ are independent of each other, the factor $\Delta i(t)\Delta t \Delta i(t')\Delta t'$ varies randomly in a way so that it could be equally often positive and negative ("ebenso oft positiv wie negativ sein können"). The sum of such terms also randomly varies with the same frequency to be positive and negative, and the multiplication of the factor $\sin(\omega_k t + \varphi_k)\sin(\omega_k t' + \varphi_k)$ would not change this property. Moreover, he claimed, it is "well known" ("bekanntlich") that for p quantities of this kind with order of magnitude a, the order of magnitude of the sum of these p quantities is $\sqrt{p}a$.[22] Since the order of magnitude for $\Delta i(t)\Delta t \Delta i(t')\Delta t'$ is less than $\langle[\Delta i(t)\Delta t]^2\rangle = ei_0\Delta t$, it follows from the stated theorem that the order of magnitude for the double sum corresponding to the second cluster is less than $[(T/\Delta t)(T/\Delta t')]^{1/2} ei_0 \Delta t = Tei_0$. Note that the double sum itself is also a zero-mean, randomly varying quantity over k. Hence we can apply the same theorem to the sum of the second cluster over k from k_1 to k_2, and get an upper bound for the order of magnitude of the second cluster's contribution to $\sum_{k=k_1(n)}^{k_2(n)} I_k^2$, which is $(4/T)(k_2 - k_1)^{1/2} ei_0$. Since $k_2 - k_1 \gg 1$, the contribution from the terms corresponding to $t \neq t'$, $(4/T)(k_2 - k_1)^{1/2} ei_0$ is much smaller than the contribution from the terms corresponding to $t = t'$, $(2/T)(k_2 - k_1)ei_0$. Consequently, $\sum_{k=k_1(n)}^{k_2(n)} I_k^2 \cong (2/T)(k_2 - k_1)ei_0$. This relation suggested that the shot current's average power around frequency ω_k is independent of frequency and hence corresponds to a "white" spectrum; specifically, $\overline{I_k^2} = (2/T)ei_0$.[23]

[22] Ibid., pp. 558–559. Schottky did not prove nor provide citation information for this theorem. But it seems that it is an intermediate result in the proof of the law of large numbers: Suppose there are p identical and independent random variables $x_1, x_2, \ldots, x_p$ with zero mean and variance $\langle x^2\rangle = a^2$. Define $s = x_1 + x_2 + \ldots + x_p$. Then it can be shown easily that s has zero mean and variance $\langle s^2\rangle = pa^2$. The root mean square of s, $[\langle s^2\rangle]^{1/2} = p^{1/2}a$ can be interpreted as the "order of magnitude" of s.

[23] Why Schottky did not use the zero correlation $\langle \Delta i(t)\Delta i(t')\rangle = 0$ to get rid of the second cluster straightforwardly is not obvious in his text. The most likely reason is that he did not want to rule out the possibility that $\langle \Delta i(t)\Delta i(t')\rangle \neq 0$ as $t - t'$ was sufficiently small. Even though the zero correlation did not hold in that case, he might be convinced that he could still use the stated theorem to demonstrate that the order of magnitude for the second cluster was much smaller than that for the first cluster.

2.6 The result

Schottky substituted the shot current's white spectrum into Eq. 1 to calculate the circuit's overall oscillating energy. He approximated the discrete sum with an integral and obtained:

$$E_S \cong \frac{ei_0 L}{T} \cdot \frac{\omega_0 T}{2\pi} \int_0^\infty \frac{dx}{(1-x^2)^2 + r^2x^2}. \tag{2}$$

He gave the value of the integral $\int_0^\infty dx/[(1-x^2)^2 + r^2x^2] = 2\pi/r^2$, leading to the total oscillating energy $E_S = \omega_0^3 Lei_0/\rho^2$. To transform this expression into a form directly comparable to the results from measurements, he argued that the energy $E_S = \omega_0^3 Lei_0/\rho^2$ is equivalent to the energy of the tube oscillator when the current i of the RLC circuit has amplitude $\sqrt{2}i_S$ and frequency that is identical to the oscillator's resonance frequency ω_0, viz., $i = \sqrt{2}i_S \sin(\omega_0 t + \varphi)$. The circuit's energy at its resonance frequency is $E_S = (i^2L/2)/r^2$. Comparing both expressions, he obtained the directly measurable effective shot current

$$i_S = \sqrt{\frac{2\pi ei_0}{\tau}} \tag{3}$$

where $\tau = 2\pi/\omega_0$ is the circuit's resonance period. He also estimated the numerical scale of the shot-noise energy compared to that of the thermal-noise energy at room temperature, and found that shot noise is usually much stronger than thermal noise.[24]

Schottky's theory of shot noise proposed an operable experimental condition to verify its underlying hypothesis and gave a quantitative prediction for a measurable entity. Yet his work was not followed immediately. Schottky's article was published in June 1918, when the war had consumed Germany and an unconditional capitulation seemed inevitable. Before the end of World War I in November 1918, vacuum-tube research was a top military secret not open to public discussions. After the war, Germany's political turmoil and economic depression interrupted normal academic activities and impeded the exchange of scholarly information with other countries. For these reasons, Schottky's work did not reach international scientific and engineering communities and even the German academia outside Siemens for years. Schottky himself did not continue the research on shot noise, either. As the war was over and a scientist's duty to serve the nation ended, he turned back to "pure" physics and left Siemens in 1919 for the University of Würzburg.[25]

[24] Ibid., pp. 555–562.

[25] Madelung (1986, p. 7).

3 Shot noise after Schottky

Despite the mathematical tricks in ensemble and frequency averages, Schottky's theory seemed to convey an elegantly simple idea: the random flow of discrete electrons in a thermionic tube gave rise to a kind of fundamental noise in electronic circuits. However, the situation became notably messier when the theory of shot noise was put into empirical test or more realistic modeling setup. In the experimental and engineering work on shot noise during the 1920s, more complex factors than those in Schottky's model had to be taken into account, and whether the shot noise posed a fundamental performance limit for electronic circuits was far from clear.

3.1 Experiments with Schottky noise

The Siemens researcher C.A. Hartmann made the first attempt to empirically corroborate Schottky's theory of shot noise. In 1920, Hartmann conducted an experiment at the K Laboratory to measure a vacuum tube's shot noise.[26] His basic experimental setup was the same as Schottky's model in Fig. 1: a vacuum tube connected in parallel to a capacitor and an inductor. The goals of his experiment were to verify (i) whether this circuit indeed had a noisy tube current, and (ii) if so, whether the noise intensity was consistent with Schottky's prediction in Eq. 3.

Hartmann proposed a procedure to address the second goal based on the fact that Schottky's quantitative prediction involved the charge of the electron, e. He expressed Eq. 3 as a relation of e with other variables, $e = (i_S^2/i_0\omega_0)$. This relation indicated that the ratio involving the measured shot current i_S, the tube's bias current i_0, and the circuit's resonance frequency ω_0 should be a fundamental physical constant $e = 1.6 \times 10^{-19}$ C.

Though simple in concept, Hartmann's experiment was challenging in several senses. First, the tube had to be at a high vacuum to satisfy Schottky's assumption that all electrons leaving the cathode arrived at the anode. To fulfill this requirement, Hartmann designed a tube continually evacuated by a pump. Second, since the shot noise was tiny, the resonance circuit's quality factor should be very high to prevent signal dissipation. Therefore, he reduced the circuit's resistance so that $\omega_0 L/R > 150$. Third, the tiny shot noise also meant that signals from the oscillating circuit should be considerably amplified. Here he used Siemens's high-gain multi-stage wideband electronic amplifiers.

Making precise measurements of the shot current was also critical. Similar to radio atmospheric noise (atmospherics), shot noise was irregular and difficult to measure by common galvanometers. Thus, Hartmann adopted a comparative method similar to AT&T engineers' in atmospherics measurements. In his design, the vacuum-tube oscillator and a "reference" monotone generator tuned at the oscillator's resonance frequency were connected to a high-gain amplifier with a switch (Fig. 2). Like those in atmospherics measurements, Hartmann switched between the tube oscillator and the monotonic signal generator and adjusted the latter until the intensity from both

[26] Hartmann (1921, pp. 51–78).

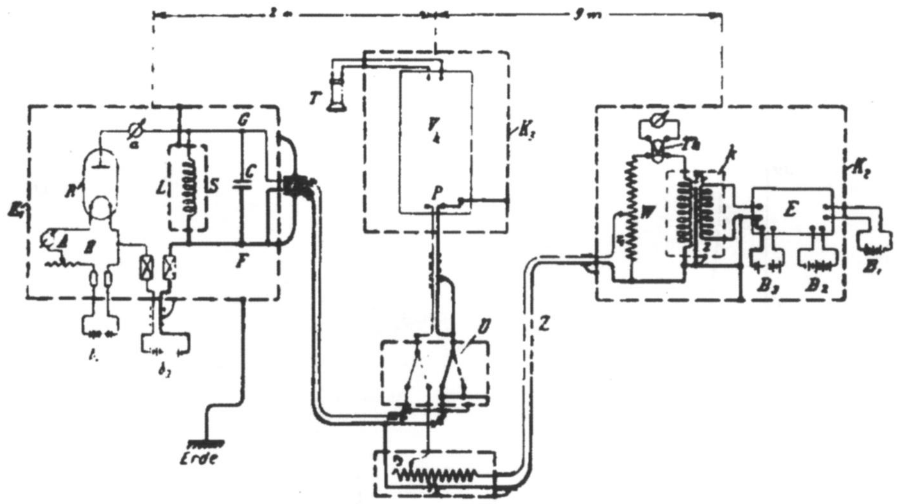

Fig. 2 Hartmann's experimental setup (Hartmann 1921, Figure 8). The block on the *left hand side* is the vacuum-tube oscillator and the block on the *right hand side* is the monotone generator. The block "V_k" is the high-gain amplifier. "U" is the switch connecting the three blocks

sounded identical (to the experimenter) at the output earphone. Then the monotonic signal intensity, easily measured by a galvanometer, was taken as the effective intensity of the shot noise.[27]

Hartmann used this setup to experiment with shot noise. His results were qualitatively satisfactory yet quantitatively problematic. The instrument indeed produced hissing tones at the output earphone. Shot noise existed! He measured the effective shot-noise intensity for various resonance frequencies ω_0 from 238.73 to 2387.33 Hz, with the bias currents i_0 at 2 and 20 milliamps, and used the measured i_S to calculate the charge of the electron e. To his disappointment, however, the experimental results deviated significantly from Schottky's prediction. The problem was twofold: The values of e from the shot-noise measurements were all in the order of 10^{-22} C, 1000 times smaller than its commonly known value 1.6×10^{-19} C from Millikan's oil-drop experiment. Also, the measured e varied with frequency, instead of remaining a constant (Fig. 3).

Hartmann's experimental results were difficult to make sense if Schottky's theory was right. The experimenter's own explanation was that the emission of electrons from the cathode was not really a Poisson process. When an electron was emitted, a part of energy was taken away and the cathode was cooled down for a period. The lower temperature reduced the cathode's ability to further emit electrons. As a result, the actual number of emitted electrons was smaller than what the Poisson-process hypothesis expected. And the measured shot current (and hence the estimated value of e) was smaller than Schottky's prediction.[28] Yet no further research supported this physical picture.

[27] Ibid., pp. 65–67.

[28] Ibid., pp. 74–76.

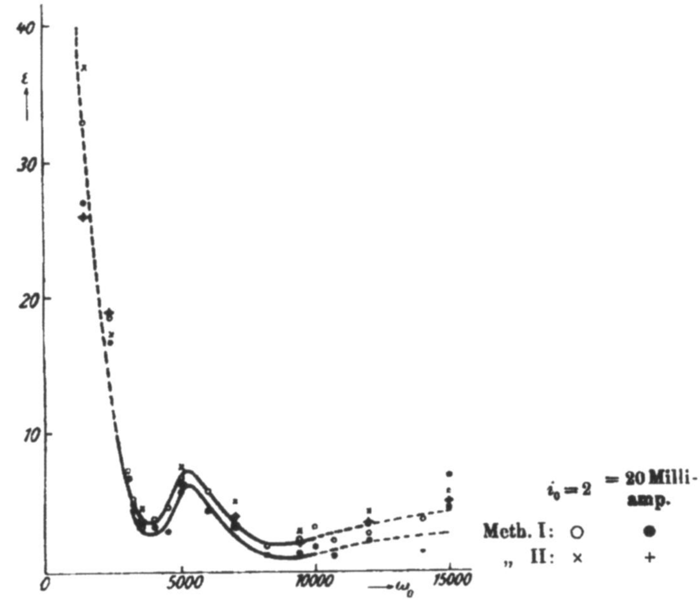

Fig. 3 Hartmann's experimental results. The abscissa is the resonance frequency ω_0 and the ordinate is the value of e from shot-noise measurements (unit 10^{-22} C). The *two curves* correspond to $i_0 = 2$ mA and 20 mA (Hartmann 1921, Figure 9)

The American physicist John Johnson found a more commonly accepted reason for the huge discrepancy between Hartmann's experimental data and Schottky's theoretical prediction: a mathematical error. Born in Sweden, John Bertrand Johnson immigrated to the U.S., received a Ph.D. in physics from Yale University in 1917, and joined AT&T's manufacturing wing Western Electric's Engineering Department in New York. Johnson was among the first-generation of Bell-System researchers with advanced training in physics. His job at Western Electric was to develop the components of AT&T's transcontinental telephony. This brought him to the problems of noise.[29] In 1920, Johnson read Schottky's 1918 article in *Annalen der Physik* (the 1918 issue did not reach America until 1920, owing to Germany's postwar postal delay). He was skeptical of Schottky's integral $\int_0^\infty dx/[(1-x^2)^2+r^2x^2] = 2\pi/r^2$ for Eq. 2. He tried to calculate the integral himself, but could not find a solution from the table of integrals. To solve the problem, he consulted a mathematician L.A. MacColl, who "suggested splitting the Schottky equation [the integrand] into four complex factors, integrating them separately and then recombining them [...] MacColl again looked at the equation and said this was a case for the method of poles and residues and, without putting pencil to paper, read off the correct result."[30] With MacColl's help,

[29] Anonymous (1971, p. 107).

[30] Johnson (1971, p. 42).

he obtained the correct value of the integral $\int_0^\infty dx/[(1-x^2)^2 + r^2x^2] = \pi/2r$. He immediately wrote to Schottky and published this new result.[31]

Johnson's mathematical exercise bridged the gap between theory and experiment. After receiving Johnson's letter, Schottky recalculated the integral with the help of his mathematician father and found Johnson right. Then he took Johnson's correction to recalculate the shot-noise energy E_S, and thus the effective shot current i_S:

$$i_S = \sqrt{\frac{ei_0R}{2L}} \tag{4}$$

Consequently, $e = (2Li_S^2/i_0R)$ instead of $e = (i_S^2/i_0\omega_0)$. For the same i_S, the new result $e = (2Li_S^2/i_0R)$ gave an estimate of the e value approximately 1000 times larger than the calculation from the old result $e = (i_S^2/i_0\omega_0)$. After the correction, therefore, Hartmann's shot-noise data led to an estimate of e much closer to its actual value—the order of magnitude was right.[32]

Nevertheless, the problem of e's frequency dependency still remained. The corrected data still yielded e values that varied with ω_0. In 1922, a Czech physicist Reinhold Fürth at the University of Prague, expert on Brownian motion,[33] explained this anomaly in terms of the "physiological" (or cognitive) nature of the experimental method.[34] Fürth argued that one could not take the results of noise measurements from the comparative method at their face value, for the method relied on the mediation of experimenters' aural cognition, a frequency-dependent feature. The comparative method required an experimenter to identify when the noise sound intensity equaled the monotonic sound intensity and to take the monotonic current as the measure of the noise current. Yet the two signals causing the same degree of aural perception in human ears had equal intensity if and only if the strength of human aural perception was proportional to the signal intensity (or, strength of stimulation) and the signals were at the same frequency. Both conditions were absent in Hartmann's experiment.

The Weber-Fechner law in physiology states that the strength of (aural) perception s is in a logarithmic rather than a linear relation with the strength of stimulation V (signal intensity): $s = c \cdot \log(V^2/V_0^2)$, where V_0 is a perceptive threshold (note $s = 0$ when $V < V_0$.) While the reference signal in the comparative method is monotonic with a definite intensity, the noise to be measured has an extended spectrum with a random intensity. Accordingly, Fürth revised the Weber-Fechner law for noise in a statistical-average form:

$$s_1 = \frac{2c\int_{V_0}^\infty dV \log(V/V_0)W(V)}{\int_{V_0}^\infty dV\, W(V)}$$

[31] Johnson (1922, pp. 154–156).

[32] Schottky (1922, pp. 157–176).

[33] Weinstein (1922, p. 403).

[34] Fürth (1922, pp. 354–362).

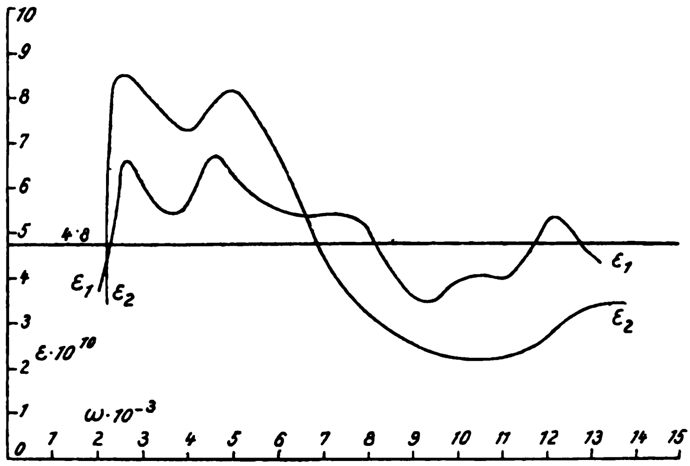

Fig. 4 Fürth's correction of Hartmann's experimental results (Fürth 1922, Figure 6). The ε_1 curve corresponds to DC bias current 2mA, the ε_2 curve corresponds to DC bias current 20 mA

where s_1 is the strength of the aural perception stimulated by noise, $W(V)\mathrm{d}V$ is the probability that the noise intensity is between V and $V+\mathrm{d}V$. Under Fürth's assumption, the random noise intensity has a Rayleigh distribution $W(V) = A \cdot \exp(-V^2/2\langle V^2 \rangle)$. Meanwhile, the aural perception s_2 stimulated by a monotonic reference signal V' follows the original law $s_2 = 2c\log(V'/V_0)$. In the comparative method, the experimenters establish that $s_1 = s_2$ by equalizing the noise and reference sound. But unlike what Hartmann had assumed, $s_1 = s_2$ does not entail that the mean noise intensity equals the reference signal intensity $[\langle V^2 \rangle]^{1/2} = V'$. Instead, Fürth's formulae for s_1 and s_2 shows that $[\langle V^2 \rangle]^{1/2}/V'$ is a function of V'/V_0. Moreover, the threshold V_0 is not a constant, either—it is a function of frequency depending on the frequency responses of both the human aural perception and the amplifying-telephonic circuit.

Fürth obtained the empirical values of V_0 at different frequencies from Hartmann's setup, and computed and plotted $[\langle V^2 \rangle]^{1/2}/V'$ as a function of V'/V_0. From both, he constructed a relation between $[\langle V^2 \rangle]^{1/2}$ and V', which he used to retrieve the noise intensity $[\langle V^2 \rangle]^{1/2}$ from Hartmann's data for the reference signal intensity V'. Reinterpreting Hartmann's data led to new estimates for the value of e with a much slighter variation with frequency than the original estimates'—Fürth's values deviated from 1.6×10^{-19} C within 100% (Fig. 4).

Albert W. Hull and N.H. Williams at General Electric's Schenectady Research Laboratory in New York further improved the accuracy of the shot-noise experiment in 1924. As GE's expert on vacuum-tube circuits, Hull had developed multi-stage high-gain amplifiers using shielded-grid tubes in the early 1920s.[35] These devices soon played an important part in the work of shot noise, in which he and his assistant Williams became interested. They attacked the shot-noise problem from the perspective of precise experiment, especially the determination of e from accurate

[35] Two years later, Hull's vacuum-tube amplifiers would achieve a gain 2000000 below 1 MHz. See Hull (1926, pp. 439–454).

Table 1 Shot-effect of temperature-limited electron current i_0 in a U.V. 199 radiotron

i_0 (mA)	γ (mA)	l (cm)	v_1 (μV)	R (Ω)	F	$\bar{v}_0^2$ (μV)	$\bar{J}^2$ (μA)	e (C)
1	.443	9.0	65.0	3.045	.776	73.8	.204	1.541×10^{-19}
2	.925	6.0	89.4	3.37	.763	102.2	.282	1.640
2	.602	9.0	88.3	3.37	.763	101.0	.279	1.603
3	1.05	6.0	102.7	3.65	.750	118.5	.327	1.595
3	.700	9.0	102.7	3.65	.750	118.5	.327	1.595
4	.580	12.0	113.4	3.85	.740	131.8	.364	1.570
4	.780	9.0	114.4	3.85	.740	133.1	.367	1.595
5	.835	9.0	122.5	4.06	.727	143.8	.397	1.566
5	.625	12.0	122.2	4.06	.727	143.5	.396	1.556
							Mean	1.586×10^{-19}

Hull and Williams's determination of e value from shot-noise measurements (Hull and Williams 1925, Table 2). The last column represents the estimates of e from measurements

noise measurements. With the high-gain amplifiers, they could get rid of Hartmann's comparative method: They connected the amplified shot-noise current—now strong enough to be measured with more common means—directly to a crystal detector and a DC current meter that measured the mean square value of the shot current.[36]

Hull and Williams used the new experimental setup to determine the value of e from the shot-current measurements. Working on a single frequency at 725 kHz, they obtained the experimental data leading to values of e extremely close to its canonically recognized value. Their estimated e at $i_0 = 1 - 5$ mA had an error range of less than 3% with respect to 1.6×10^{-19} C (Table 1).

Hull and Williams did more than improve the experiment in order to come closer to the known value of e, however. They found a new factor that qualified Schottky's theory of shot noise—the space-charge effect. W.L. Carlson at GE's Radio Department had noted that a tube's noise intensity fell as the tube was more strongly charged. This phenomenon caught Hull and Williams's attention. At first, they believed that it was an artifact due to the reducing tube resistance. But they changed their mind when they still observed consistently low noise values after stabilizing the tube resistance. Now they thought that the reduction of the shot noise was caused by the accumulated space charge between the cathode and the anode; such space charge created a potential barrier against the flow of electrons. If the space charge inside a tube reached saturation after a sufficiently long time (the "space-charge limited" case), then all the relative motions between the electrons inside the tube disappeared and the electrons moved regularly in a uniform stream. In this case, the lack of randomness and discontinuity eliminated the shot noise. Schottky's theory of shot noise was valid when a tube's electronic current was limited only by its filament temperature—the filament's ability to emit electrons (the "temperature limited" case). When the current was limited by the tube's space charge, the shot noise was greatly reduced and Schottky's theory no longer applied.[37]

36 Hull and Williams (1925, pp. 148–150).

37 Ibid., pp. 166–170.

3.2 Theoretical revision

As Hartmann et al. worked on the shot-noise experiments, efforts were made to examine and to refine Schottky's theory. Schottky's theory of shot noise was built on statistical reasoning. Translated into mathematics, his problem was to solve a differential equation (the circuit equation) with a random source (the tube's shot noise) whose statistics were partially known. He solved the problem via a Fourier analysis. But his approach had shortcomings—its calculations were complicated, the physical meaning of each Fourier component was unclear, the assumption underlying the frequency average was questionable, and his spectrum of random noise lacked a rigorous definition. In the 1920s, physicists and engineers attempted to solve the shot-noise problem without resorting to the Fourier analysis.

The Dutch physicists L.S. Ornstein and H.C. Burger at Ryks University (Utrecht) were the earliest to revise Schottky's approach. In 1923, they claimed to solve the random differential equation in a more direct way. Key to their solution was to conflate the statistical average with the time average. In so doing, they considerably simplified the differential equation using integration by parts and obtained the mean square of the current J out of the resistive inductor: $\langle J^2\rangle = e^2n^2 + e^2n\omega_0^2/\rho$. In this expression, the first term corresponded to the DC current and the second term corresponded to the current fluctuation, equivalent to Schottky's formula of E_S.[38]

The British physicist Norman Campbell made another attempt to revise the formulation. Aiming to explain both thermoionic and photoelectric emissions, he formulated a generalized principle that considered a series of identical events randomly occurring in time. An indicating instrument measures the effect of these events. Suppose $\theta = f(t)$ gives the instrument reading at time t after an event occurs at time 0, $\langle\Delta n^2\rangle$ gives the standard deviation of the number of events in a unit time. Then the standard deviation of θ is $\langle\Delta\theta^2\rangle = \langle\Delta n^2\rangle \int_0^\infty \mathrm{d}t\, f^2(t)$. This principle implied that the overall shot effect can be decomposed into a factor describing the statistics of the *collection* of electrons and a factor related to the waveform excited by an *individual* electron.[39]

Thornton C. Fry at AT&T pursued further Ornstein and Campbell's reasoning further. A mathematician, Fry had taught at the University of Wisconsin and MIT before joining AT&T. In 1925, he assumed directorship of Bell Lab's Mathematical Research Department, where he built a staff of computing and mathematics to serve the firm's industrial research. A specialist in engineering probability,[40] Fry thought that the major problem of Schottky's theory was the lack of a mathematically rigorous treatment of probability and randomness, and he was prepared to offer one of his own.

Like Campbell, Fry began by considering a general case in which a measuring device detected current I and voltage E excited by independently arriving electrons in a long period T. Since the number of electrons within T is uncertain, the device's mean instantaneous power is $\langle EI\rangle = \sum_{n=0}^{\infty} p(n)\langle(EI)_n\rangle$, where $p(n)$ is the probability that n electrons arrives within T and $\langle(EI)_n\rangle$ is the mean instantaneous power of n

38 Ornstein and Burger (1923, pp. 622–624).

39 Campbell (1925, pp. 81–86).

40 Mindell (1996, pp. 193–194).

electrons. That an electron's arrival is independent of others implies that $p(n)$ follows a Poisson distribution $p(n) = \exp(-\upsilon T)(\upsilon T)^n/n!$ (υ is the average number of electrons arriving in a unit period). Also, $\langle (EI)_n \rangle$ can be calculated from the assumption that the current and voltage excitation from the arriving electrons are additive. The additive formula can be expressed in terms of an iterative relation between $\langle (EI)_n \rangle$ and $\langle (EI)_{n-1} \rangle$. Repeating the iteration n times led to $\langle (EI)_n \rangle = \langle (EI)_1 \rangle + n(n-1)\langle E_1 \rangle \langle I_1 \rangle$. Substituting the expressions of $\langle (EI)_n \rangle$ and $p(n)$ into $\langle EI \rangle$, Fry obtained $\langle EI \rangle = (\upsilon T)\langle (EI)_1 \rangle + (\upsilon T)^2 \langle E_1 \rangle \langle I_1 \rangle$. He interpreted $\langle EI \rangle$ as the system's overall power, $(\upsilon T)\langle (EI)_1 \rangle$ as the power of shot noise, and $(\upsilon T)^2 \langle E_1 \rangle \langle I_1 \rangle$ as the power of the DC component. This implies that the shot-noise energy $E_S = \upsilon \langle w_1 \rangle$, where w_1 is an electron's energy dissipated in the measuring device.

The quantity $\langle w_1 \rangle$ is the average energy generated in the measuring device by an electron traveling from the cathode to the anode of a vacuum tube. To facilitate the calculations, Fry modeled the tube as a parallel-plate condenser. A unit-charge particle's movement from one plate to another induces a time-variant voltage $v(t)$ and current $i(t)$ across the plates. Consequently, $\langle w_1 \rangle = e^2 \int_0^\infty \mathrm{d}t\, v(t) i(t)$. The quantities $v(t)$ and $i(t)$ depend exclusively on the circuit. In Schottky's theory, the circuit is a vacuum tube parallel to an RLC resonance network (Fig. 1). Fry modeled the tube as a capacitor, obtained $v(t)$ and $i(t)$, and substituted them into his formulae for E_S. He found that Schottky's prediction for the shot-noise energy should be modified by a multiplicative factor $[1 + R^2(C + C_{\mathrm{tube}})/L]$, where R, L, C are the resonance circuit's resistance, inductance, and capacitance, and C_{tube} is the tube's effective capacitance.[41]

3.3 Flicker noise

Hull and Williams's experiment and Fry's theory inspired J.B. Johnson to research further into shot noise. In 1925, Johnson performed an experiment at Bell Labs to measure shot noise in the Schottky circuit using Hull and Williams's method (high-gain amplifier, crystal detector, and direct current meter). He compared his experimental results with Fry's revised prediction. His goal, relevant to AT&T in particular and to the electronics industry in general, was to find vacuum-tube circuits' performance limit imposed by electronic noise. Johnson measured the noise strength of the triode tubes from about 100 commercial electronic amplifiers of various kinds. His results confirmed Hull and Williams's finding: the vacuum tubes of amplifying circuits operated at the space-charge-limited condition, so their shot-noise intensity was much lower than Schottky's and Fry's predictions.[42]

In addition to that, Johnson discovered a new phenomenon. To test the general applicability of Schottky's or Fry's theory, he measured the noise of the diode tubes operating at the temperature-limited condition (i.e., without the space-charge effect). In the measurements, he changed the circuit's resonance frequency from 0 to 10 kHz to observe the variation of the measured noise intensity with frequency. This arrangement led to a novel effect: at high frequencies, the measured noise intensity was close

41 Fry (1925, pp. 203–220).

42 Johnson (1925, pp. 81–83).

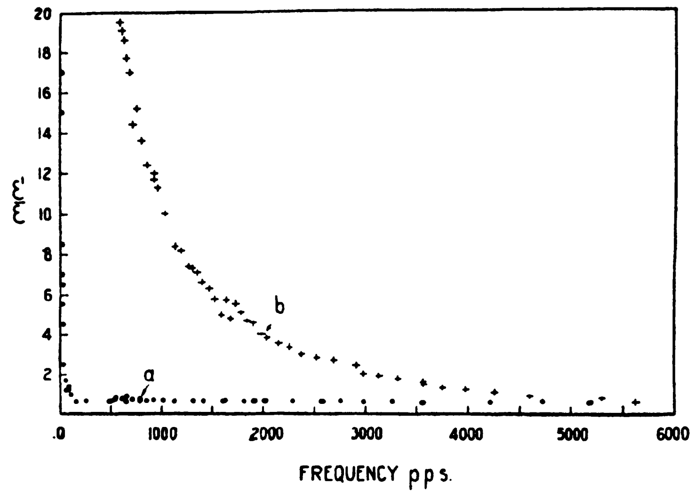

Fig. 5 Johnson's flicker noise versus frequency (Johnson 1925, Figure 7); *a* tungsten filaments, *b* oxide-coated filaments

to Fry's prediction; but at low frequencies, the noise intensity increased rapidly with a decreasing frequency. This low-frequency deviation changed with the filament material. Between tungsten and oxide coated filaments, the latter had much higher noise intensity. With a tungsten filament, the ratio of the *e* value obtained from the noise data to the canonical one of 1.6×10^{-19} C was 0.7 for frequencies above 200 Hz, but increased to 50 at 10 Hz. With an oxide coated filament, the ratio increased from 1 at 5 kHz to 100 at 100 Hz (Fig. 5). It seemed that at low frequencies certain much stronger noise *of a different kind* superseded ordinary shot noise. Johnson named this new kind of noise the "flicker effect" for its special sound pattern.[43]

Flicker noise, Johnson argued, differed fundamentally from shot noise. Its much higher magnitude implied that it was not due to random and mutually independent emission of discrete electrons. Since the flicker effect depended upon the filament material, its cause was related to the activities at the filament. A cathode surface changed continually with evaporation, diffusion, chemical actions, structural rearrangements, and ion bombardment. These factors altered the filament's rate of electronic emission, and the changing rate caused current fluctuations, the source of this noise.[44]

Fry's revision of Schottky's theory, Hull and Williams's findings of the space-charge effect, and Johnson's discovery of the flicker effect at low frequencies opened up new directions for the shot-noise research in the rest of the 1920s. Following Fry's approach, Stuart Ballantine of the Radio Frequency Laboratory in Boonton (New Jersey) calculated in a more precise way the voltage and current excited by an electron between two parallel plates to determine shot noise's power spectrum in the

[43] Ibid., pp. 76–80.

[44] Ibid., p. 85.

temperature-limited case.[45] After leaving General Electric for the Physics Department at the University of Michigan, Williams collaborated with his colleagues to improve the methods of noise measurements to determine more accurately the charge of electron. In the space-charge-limited condition, they also used the experimental techniques to measure the charge of positive thermions and to examine the filament material's effect on charge emission.[46] Schottky developed a theory to explain the flicker effect's frequency dependence. He assumed that the large noise intensity at low frequencies was caused by the fluctuation of the filament's electron-emitting capability owing to the coming and going of "foreign" atoms on the filament surface. Yet a foreign atom, once reached the filament surface, did not depart immediately; instead, it stayed there for a short period of time t_a. Within this short period, therefore, the numbers of foreign atoms at the filament surface at two distinct instants were close to each other; while such numbers fluctuated significantly with respect to each other if the separation between the two instants far exceeded t_a. In other words, the correlation function between the numbers of foreign atoms (which determined the correlation function of the filament's electron-emitting capability) at two instants separated by time t was not a simple impulse at $t = 0$ and zero otherwise, like the shot noise. Rather, it had a flat top within t_a and fell off to zero (quickly or slowly, depending on the nature of atomic attachment and detachment) outside that duration. This non-trivial correlation function led to flicker noise's frequency dependence: $\langle I^2_{\text{flicker}} \rangle \propto f^{-n}$ (n was between 1 and 2).[47]

These diverse studies revealed that noise was much more complicated than Schottky's original theory in 1918 had expected. A vacuum tube was a microcosm with various physical mechanisms. Random emissions of discrete electrons could only partially explain the microcosm's fluctuations in a limited sense. Schottky's theory failed in the space-charge-limited case at which tube amplifiers operated. In the temperature-limited case, the flicker effect dominated over the shot effect at low frequencies. Emission of ions, generation or recombination of ions and electrons, and lumping of charged particles all affected tube noise. Schottky's theory was a proper concept to start grasping electronic noise, but was by no means a useful framework to determine the fundamental limit of vacuum-tube circuits' amplification that engineers wanted.

Then what was that limit?

4 Thermal noise in resistors

When Johnson conducted the shot-noise experiment in 1925 to seek the fundamental limit of tube amplification, he measured the noise intensity and the gain of about a hundred commercially available triode tube amplifiers. He plotted the results in a Cartesian coordinates system (the ordinate was the gain and the abscissa was the noise

[45] Ballantine (1928, pp. 159–167). This approach culminated in a classic textbook on random noise written in 1958 by MIT Lincoln Laboratory researchers Wilbur Davenport and William Root. See Davenport and Root (1958, pp. 112–144).

[46] Williams and Vincent (1926, pp. 1250–1264), Williams and Huxford (1929, pp. 773–788), Kozanowski and Williams (1930, pp. 1314–1329), Donal (1930, pp. 1172–1189).

[47] Schottky (1926, pp. 74–103).

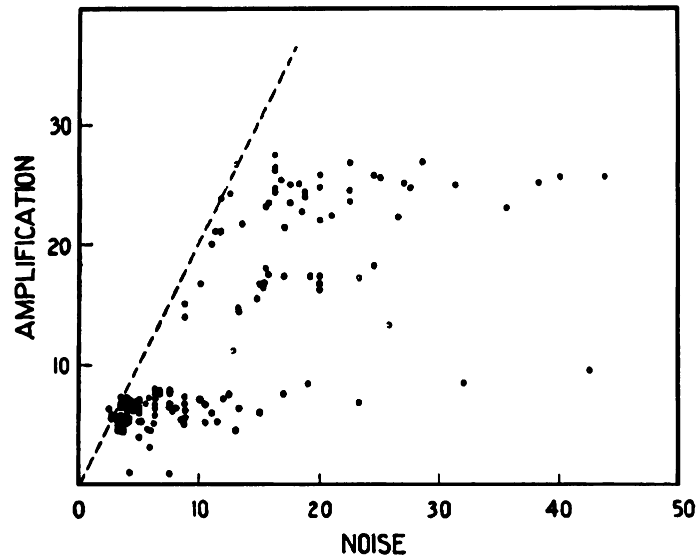

Fig. 6 Amplification as a function of noise in triode tubes (Johnson 1925, Figure 13). The *dots* represent measurements taken from about 100 commercial tube amplifiers

strength) and found an interesting pattern (Fig. 6). The points representing various amplifiers' gains and noise levels distributed to the right of a sloped straight line passing the origin. That is, for a fixed amplification, the noise intensity was always higher than the value given by the sloped line, meaning that the line represented the lower bound of the amplifiers' noise intensity. This minimum noise intensity was the residual tube noise when the disturbing factors were suppressed to the largest extent.[48]

4.1 Johnson's inquiry

What was the physical significance of the minimum noise represented by the sloped line in Fig. 6? At first, Johnson thought it was Schottky's shot noise without the space-charge effect and the flicker effect. But he changed his mind after noting a fact: since a sloped line passing the origin represented a proportional relation, the minimum output tube noise strength in Fig. 6 was proportional to the tube amplifier's gain. This implied that a tube's input noise strength was a *constant* when its output noise was a minimum. If so, then the minimum residual noise had nothing to do with the vacuum-tube amplifier itself. Rather, it was determined only by the amplifier's input condition.

To testify this hypothesis, he put a resistor at the input of an amplifying circuit and changed its resistance to observe whether the output noise strength varied accordingly. It did, and the measured noise intensity was proportional to the input resistance. This resistor-dependent noise reminded Johnson of a less discussed aspect of Schottky's 1918 article—thermal noise. Was the residual noise a result of electrons' thermal

[48] Johnson (1925, pp. 83–85).

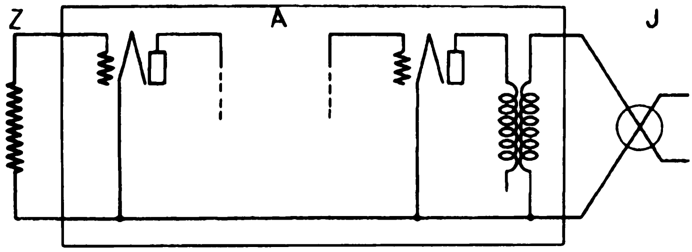

Fig. 7 Johnson's experimental apparatus on thermal noise (Johnson 1928, Figure 1)

agitation at the amplifier's input resistor? Informed by this assumption, Johnson measured the noise of the least noisy amplifiers with input resistors different in temperature, size, or material. The measured noise intensity was proportional to the input resistor's temperature only, not size or material. This strongly suggested that the minimum residual noise was a fundamental physical effect related to the input resistor's thermal agitation.[49]

In 1926, Johnson performed a systematic experiment on thermal noise at Bell Labs with his assistant J.H. Rohrbaugh. His focus now shifted from tubes to resistors. His circuit consisted of a resistor connected to a six-stage tube amplifier with the output coupled to a thermocouple for current measurements (Fig. 7). The high-gain amplifier suppressed the shot noise, the flicker noise, and other fluctuating effects to the maximum extent, and was shielded against external electric, magnetic, acoustic, and mechanical shocks. The input resistor was in a constant-temperature bath. The aim of the experiment was to determine the mean square voltage $\langle V_{\mathrm{m}}^2 \rangle$ of the thermal noise at the input resistor from the mean square noisy current $\langle I^2 \rangle$ measured at the amplifier's output via the relation $\langle I^2 \rangle = Y^2 \langle V_{\mathrm{m}}^2 \rangle$($Y$ was the ratio of the amplifier's output current to its input voltage).[50]

Johnson's experiments in 1926 reached the same but more expressive conclusion: The ratio of thermal noise's mean square voltage to the input resistance $\langle V_{\mathrm{m}}^2 \rangle / R$ was independent of the resistor's shape and material. Johnson tried different materials with identical resistance including metal wires, graphite, thin metal films, films of drawing ink, and electrolytes such as NaCl, $CuSO_4$, K_2CrO_4, and $Ca(NO_3)_2$. All yielded the same measured values. Also, the ratio $\langle V_{\mathrm{m}}^2 \rangle / R$ was proportional to the resistor's absolute temperature. At room temperature, a 5000-Ω resistor had $\langle V_{\mathrm{m}}^2 \rangle / R \cong 10^{-18}$ W. He obtained thereby an empirical formula for the thermal noise: $\langle V_{\mathrm{m}}^2 \rangle / R = \langle I^2 \rangle / (Y^2 R) = KT$ (K was a proportionality constant).[51]

As Johnson continued the experiment, he was obliged to revise the empirical formula for thermal noise $\langle I^2 \rangle = KTRY^2$, since both the band-pass amplification Y and the input resistance R were functions of frequency. To incorporate the frequency-dependent feature, he suggested that $\langle I^2 \rangle = KT \int_0^{\infty} \mathrm{d}\omega R(\omega) \, |Y(\omega)|^2$. Then he compared the measured data with the empirical formula. He determined the amplification

[49] Johnson (1971, pp. 43–44).

[50] Johnson (1928, pp. 98–101).

[51] Johnson (1927, pp. 50–51).

$|Y(\omega)|^2$ by measuring the amplifier's input–output characteristics. The input resistance $R(\omega)$ was the real part of the amplifier's input impedance, which was modeled as a resistor R_0 connected in parallel to a capacitor C. By substituting the measured $|Y(\omega)|^2$ and the modeled $R(\omega)$ into the empirical formula, he obtained an estimate of $\langle I^2 \rangle$ from numerical integration. The measured $\langle I^2 \rangle$ varied in the same pattern as the estimated $\langle I^2 \rangle$ had predicted. Johnson's formula seemed right.[52]

4.2 Nyquist's theory

That the formula was empirically adequate did not satisfy Johnson; he was curious about the theoretical foundation of his formula. In 1927, he brought this issue to his colleague Harry Nyquist at AT&T Development and Research, another Swedish immigrant with a Ph.D. in physics from Yale.[53] To explain Johnson's formula, Nyquist spent a month developing a theory based on electric circuit analysis, thermodynamics, and statistical mechanics.[54]

According to Schottky, thermal noise is fluctuations generated by electrons' random thermal motions in a conductor. Its simple relation with temperature indicates that it is a thermodynamic phenomenon. Thus, Nyquist's theory begins with a simple scenario at thermal equilibrium—two identical conductors (I and II), each with resistance R, connected together by perfectly conducting wires and at temperature T. Electrons in the two conductors undergo continual thermal agitation as long as T is not zero. The thermal agitation in conductor I induces an electromotive force on the entire circuit, including conductor II. The electromotive force yields a current around the circuit. The current heats conductor II, and hence transfers power from conductor I to conductor II. In the same manner, power is transferred from conductor II to conductor I. At thermal equilibrium, power transferred from I to II equals that from II to I.[55]

Nyquist contended that at thermal equilibrium, not only the total power but also the power at any frequency exchanged between conductors I and II equals each other.[56] This shows that the thermally agitated electromotive force is a universal function of frequency, temperature, and resistance—like other thermodynamic functions. Its form does not change with physical setups not involving frequency, temperature, and resistance. Thus, Nyquist considered a different setup with the two conductors connected

52 Johnson (1928, pp. 102–103).

53 Mindell (1996, p. 207).

54 Nyquist (1927, p. 614).

55 Nyquist (1928, p. 110).

56 If not, a contradiction ensues. Suppose that conductor I sends more power to conductor I than it receives from conductor II in a frequency band. Then it sends less power to II than it receives from II in the rest of the frequency spectrum, for the total power from I to II equals that from II to I. When a resonance circuit blocking the energy transfer in the frequency band is set between the two conductors, the total power transferred from I to II equals the power from I to II without the resonance circuit less the power in the frequency band. Similarly, the total power from II to I equals the power from II to I without the resonance circuit less the power in the frequency band. With the resonance circuit, therefore, the total power from I to II is less than that from II to I, a contradiction to thermal equilibrium. So the assumption that the power transfer in the frequency band is unbalanced is wrong.

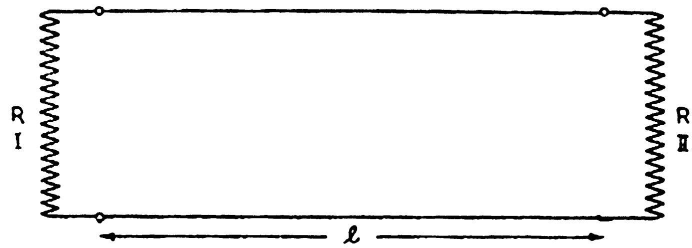

Fig. 8 Nyquist's scenario for thermal-noise calculation (Nyquist 1928, Figure 3)

by a long non-dissipative transmission line with length l (Fig. 8). The transmission line has inductance L and capacitance C per unit length so that its characteristic impedance $(L/C)^{1/2}$ equals R, meaning no reflection at both ends. At thermal equilibrium, two trains of energy traverse the transmission line—one from left to right being the power of the thermal agitation from I to II and the other from right to left being the power of the thermal agitation from II to I.

To calculate the transferred power, Nyquist isolated the transmission line from the conductors at $t = 0$ by short-circuiting its two ends, trapping the energy on the line. The transmission line containing the energy transferred from both conductors resembles Planck's black-body radiator. It is a resonator with modes of vibration corresponding to stationary waves. The modes of vibration have frequencies $nv/2l$ (v is the speed of the energy transfer and n is a positive integer). Since l is long, the number of vibrating modes (degrees of freedom) is huge and can be represented as an approximately continuous function of frequency: the number of modes between f and $f + \mathrm{d}f$ is $2l\mathrm{d}f/v$. For so many vibrating modes, the energy distribution follows Boltzmann's law: each degree of freedom has the mean energy kT ($k = 1.372 \times 10^{-24}$ J/K is Boltzmann's constant). Therefore, the transmission line's mean energy between f and $f + \mathrm{d}f$ is $2lkT\mathrm{d}f/v$. This is the mean energy within $(f, f + \mathrm{d}f)$ transmitted from the two conductors to the line during the transit time l/v. This implies that the mean power each conductor transferred within $(f, f+\mathrm{d}f)$ is $P(f)\mathrm{d}f = (2lkT\mathrm{d}f/v)/(2l/v) = kT\mathrm{d}f$.

From the power spectrum $P(f) = kT$, Nyquist obtained the electromotive-force spectrum $E(f)$: The electromotive force $E(f)\mathrm{d}f$ generates a current $I(f)\mathrm{d}f = [E(f)/2R]\mathrm{d}f$ in the circuit comprising conductors I and II. The power spectrum is therefore $P(f) = I^2(f)R = [E^2(f)/4R]$. This relation, along with $P(f) = kT$, implies $E^2(f) = 4RkT$, which expresses the electromotive force induced by the thermal agitation of a resistor with resistance R and temperature T. Nyquist extended this relation by considering the thermal effect of a passive network with a frequency-dependent impedance $Z(f) = R(f) + iX(f)$. From circuit theory, he showed that $E^2(f) = 4R(f)kT$.

In Johnson's experiment, $E^2(f)\mathrm{d}f$ is the mean square thermal voltage between f and $f + \mathrm{d}f$ at the amplifier's input. That is, the mean square thermal current between f and $f + \mathrm{d}f$ at the amplifier's output is $E_f^2(f)|Y(f)|^2\mathrm{d}f$ ($Y(f)$ is the ratio of the amplifier's output current to its input voltage at frequency f) and the overall output mean-square current is[57]:

$$\left\langle I^2 \right\rangle = \int_0^\infty E^2(f)\,|Y(f)|^2\,\mathrm{d}f = \frac{2kT}{\pi}\int_0^\infty R^2(\omega)\,|Y(\omega)|^2\,\mathrm{d}\omega \qquad (5)$$

Equation 5 is identical to Johnson's empirical formula $\langle I^2 \rangle = KT\int_0^\infty \mathrm{d}\omega R(\omega)|Y(\omega)|^2$ except that the proportionality constant K is now replaced by a fundamental constant $2k/\pi$.

Nyquist tested his formula with Johnson's experimental results. Key to his test was Boltzmann's constant k. Johnson had data for $R(\omega)$, $Y(\omega)$, T, and $\langle I^2 \rangle$. Nyquist calculated k from this data using Eq. 5 and compared the results with its standard value $k = 1.372 \times 10^{-24}$ J/K. The outcome was satisfactory: the estimated value of k from the thermal-noise measurements was only about 7.5% below the standard value. Johnson attributed this discrepancy to the inaccuracy of the amplifiers' gain.[58]

Johnson and Nyquist's theory of thermal noise had practical implications for electronic circuit design. Equation 5 showed that the thermal noise increased with the input resistance, the input resistor's temperature, and the amplifier's bandwidth. As a consequence, Johnson argued, there were three practical means to suppress an amplifier's thermal noise—to reduce the device's temperature, to restrict the amplifier's input resistance (the resistance between the first-stage tube's grid and filament), and to make the amplifier's bandwidth no greater than needed.[59]

5 Conclusion: engineering treatment of electronic noise

By the 1930s, radio engineers had known two kinds of electronic noise—thermionic tubes' shot noise (including the flicker effect) and resistors' thermal noise. Although device defect or malfunction still mattered, such electronic noise became more and more critical in determining the performance limit of electronic devices as the device quality improved. Also, the originally tiny shot and thermal noise was considerably magnified as the tube amplifiers' gains increased significantly. For cutting-edge radio receivers free from most quality problems, shot and thermal noise seemed to pose the "ultimate" limit of signal amplification.

Schottky's, Hartmann's, Hull and Williams's, Johnson's, Nyquist's, and others' scientific research into shot and thermal noise provided engineers with the means to push forward the envelope of the "ultimate" performance limit of electronic devices. Shot noise and flicker noise occurred only when a thermionic tube had zero or little space charge. As the tube's filament temperature was high enough to saturate itself with space charge, the tube current became coherent and the random emission of discrete electrons did not cause current fluctuations. Thermal noise was the result of the thermal

[57] Ibid., p. 113.

[58] Johnson (1928, pp. 104–105).

[59] Ibid., pp. 106–107.

agitation in an amplifier's input resistor. It was saliently suppressed if the amplifier's input stage had a low resistance and was maintained at a low temperature.

Yet engineers found the guidelines from the scientific theories of noise inadequate. According to the theories, a tube amplifier with zero input resistance and high filament temperature had zero output noise. In 1930, F.B. Llewellyn at Bell Labs measured the noise of tube amplifiers with zero input resistance and high filament temperature. Under this almost perfect condition, however, the measured output noise intensity was still conspicuous. The discrepancy between theory and experiment made Llewellyn think about the *real* limit of amplification. He proposed three possible causes for the output noise with zero input resistance and space-charge-limited tubes. First, the shot effect still existed at space-charge saturation, since the space charge only diminished rather than eliminated the emitting fluctuation's influence. Second, even at zero input resistance, electrons traversing a tube still had thermal agitation, and their noisy effect could be represented by an internal resistance at the tube's plate. Third, the traversing electrons ionized air in the tube, and the secondary charged particles from ionization bombarded the tube's electrodes to create some output current fluctuation. Unfortunately, the effects of shot noise at the space-charge-limited condition, thermal noise of the plate resistance, and noise from ionization were difficult to analyze theoretically or quantitatively.[60]

Llewellyn's discovery showed that electronic noise was more complex than Schottky et al.'s theories could capture. The shot-noise and thermal-noise models derived from fundamental electrodynamics, statistical mechanics, and thermodynamics were replaced by pictures of tangled interactions between charged particles and electronic tubes' physical setups. Research into the field in the 1930s indicated that a more accurate understanding of noise could only be achieved by studying the complicated details of electronic physics in thermionic tubes.

Engineers could not wait until physicists confidently grasped the phenomena. Even though physicists could obtain a theoretical understanding of the noise, as Johnson and Llewellyn remarked in 1935, "the greater part of the noise in practical tubes is caused by things that have not been included in theory and that are still in a state of flux."[61] More than a fundamental limit of noise in ideal tubes, engineers needed to know the noise strength in real tubes for designs and operations. And they needed a practical, systematic method to obtain such information. Measurement was perhaps the only feasible way. Built on Johnson and Llewellyn's work, engineers at Bell Labs developed a scheme to experimentally characterize the tube noise strength. They measured a tube's noise intensity under various operating conditions and, most importantly, when the tube's grid (input) was short-circuited to ground. Under the short-circuit condition, the tube's input resistance was zero and the measured values were treated as the tube's intrinsic noise levels. This measure, called the tube's "noise figure," was represented by the thermal noise of an effective resistor at the tube's input. The noise-figure scheme offered a systematic method to rate off-the-shelf thermionic-tube devices. Throughout the 1930s, American, British, and German radio engineers studied, measured, rated,

[60] Llewellyn (1930, pp. 243–265).

[61] Johnson and Llewellyn (1935, p. 92).

and published the noise figures of commercial electronic tubes.[62] Systematic measurements inspired by, but independent of, noise physics became the canonical engineering treatment of electronic noise.

Physicists' research into electronic noises contributed to engineers' endeavors to tame uncertainties in electronic circuit designs and operations, but not necessarily in the way they had anticipated. The discoveries of the shot, flicker, and resistive thermal effects helped engineers to understand noise's fundamental nature. The theories of these effects provided noise-reduction guidelines (increasing the filament temperature, reducing the input resistance, etc.) engineers often used. Yet Schottky's and Johnson's original hope to find the ultimate limit of amplification posed by noise diminished after the more complex physical processes inside electronic tubes were found. The theories of shot noise and thermal noise could not predict electronic noise. To obtain the electronic-noise data useful to radio engineering, systematic measurement was still the only viable method in the 1930s. This does not mean, however, that noise physics was irrelevant as engineers chose to characterize electronic noise with measurements—a practical art—instead of with predictions. Rather, the theories shaped the measuring methods. The noise-figure scheme, for example, was built on the theory of thermal noise at both the operational level (short-circuiting the input resistor) and the representational level (representing the noise with effective resistance).

The electronic-noise studies sought a new kind of engineering knowledge—a technology's fundamental limitation—from the assistance of physics. To grapple with electronic noise, one of the most common uncertainties in electrical devices, the German and American physicists and engineers aimed to characterize the *upper bound* of *ideal* devices' performances, not the actual performances of *real* devices. Conceptually, they focused on the ideal-type devices free from "human factors"—flaws, defects, unsatisfied qualities—so that the performance imperfection was caused by the devices' fundamental working principles and hence posed an intrinsic limit to the real devices of the same kind. Using statistical mechanics, thermodynamics, circuit theory, and electrodynamics, they constructed theories to predict such an upper bound. Underneath these theories was the conviction that the random Brownian-like electronic motions generated fluctuations of performance at electrical devices. Therefore, the physical quantities (such as voltage and current) and the abstract engineering entities (such as signals) were represented by stochastic variables and the measurable noise effect was the result of the variables' statistical behavior. In so doing, the engineers and physicists converted the technological systems they were dealing with to models of stochastic systems.

The noise studies of the 1920s–1930s did not fulfill their goal. Experiments showed that neither shot noise nor thermal noise posed the ultimate limit to vacuum-tube amplifiers. Other kinds of noise were present even when the Schottky effect and the Johnson effect were null. A thermionic tube, a complex microcosm with many types of electron-material interactions, did not achieve its optimum state simply by eliminating shot noise and the input resistor's thermal noise. Failing to fulfill their goal, however, did not make the theories of noise useless to engineers. The theories might not be adequate

62 Ibid., pp. 92–94; Johnson (1971, p. 46).

in explaining or predicting. But they gave engineers a *language* to talk about noise. From the theories, engineers gained intellectual tools to make sense of, to quantify, and to define electronic noise.

References

Archives

Walter Schottky Papers (NL 100), Box 009, Folder 4.1.1, "Persönliche Papiers," Deutsches Museum Archives, Munich, Germany.

Published Sources

Anonymous. 1971. John B. Johnson, recent Sarnoff Award winner (obituary). *IEEE Spectrum* 8:1 (pp. 107).

Ballantine, Stuart. 1928. Schrot-effect in high-frequency circuits. *Journal of the Franklin Institute*, 206: 159–167.

Beller, Mara. 1999. *Quantum dialogue: The making of a revolution*. Chicago: University of Chicago Press.

Bigg, Charlotte. 2005. Brownian motion, In *Albert Einstein: Chief Engineer of the Universe*, ed. Jürgen Renn, pp 120–123. Weinheim: Wiley-VCH.

Bigg, Charlotte. 2008. Evident atoms: Visuality in Jean Perrin's Brownian motion research. *Studies in the History and Philosophy of Science* 39(3):312–322.

Campbell, Norman. 1925. The theory of the 'Schrot-effect'. *Philosophical Magazine* (*Series 6*) 50:81–86.

Cohen, Leon. 2005. The history of noise: On the 100th anniversary of its birth. *IEEE Signal Processing Magazine* 22(6):20–45.

Davenport, Wilbur B., Jr., and William L. Root. 1958. *An Introduction to the Theory of Random Signals and Noise*. New York: McGraw-Hill.

Donal, John S., Jr. (1930). Abnormal shot effect of ions of tungstous and tungstic oxide. *Physical Review* 36:1172–1189.

Dörfel, Günter, and Dieter Hoffmann. 2005. Von Albert Einstein bis Norbert Wiener—frühe Ansichten und spate Einsichten zum Phänomen des elektronischen Rauschens, Preprint 301, Berlin: Max Planck Institute for the History of Science.

Fry, Thornton C. 1925. The theory of the Schroteffekt. *Journal of the Franklin Institute* 199:203–220.

Fürth, Reinhold. 1922. Die Bestimmung der Elektronenladung aus dem Schroteffekt an Glühkathodenröhren. *Physikalische Zeitschrift* 23:354–362.

Hartmann, C.A. 1921. Über die Bestimmung des elektrischen Elementarquantums aus dem Schroteffekt. *Annalen der Physik (Series 4)* 65:51–78.

Hull, Albert W. 1926. Measurements of high frequency amplification with shielded-grid pliotrons. *Physical Review* 27:439–454.

Hull, Albert W., and N. H. Williams. 1925. Determination of elementary charge *e* from measurements of shot-effect. *Physical Review* 25:147–173.

Johnson, John B. 1922. Bemerkung zur Bestimmung des elektrischen Elementarquantums aus dem Schroteffekt. *Annalen der Physik (Series 4)* 67:154–156.

Johnson, John B. 1925. The Schottky effect in low frequency circuit. *Physical Review* 26:71–85.

Johnson, John B. 1927. Thermal agitation of electricity in conductors. *Nature* 119:50–51.

Johnson, John B. 1928. Thermal agitation of electricity in conductors. *Physical Review* 32:97–109.

Johnson, John B. 1971. Electronic noise: the first two decades. *IEEE Spectrum* 8(1):42–46.

Johnson, John B., and F.B. Llewellyn. 1935. Limits to amplification. *Bell System Technical Journal* 14: 92–94.

Kozanowski, H.N., and N. H. Williams. 1930. Shot effect of the emission from oxide cathodes. *Physical Review* 36:1314–1329.

Kuhn, Thomas. 1978. *Black-body theory and the quantum discontinuity, 1894–1912*. New York: Oxford University Press.

Llewellyn, F.B. 1930. A study of noise in vacuum tubes and attached circuits. *Proceedings of the Institute of Radio Engineers* 18(2):243–265.
Maiocchi, Roberto. 1990. The case of Brownian Motion. *The British Journal for the History of Science* 23(3):257–283.
Madelung, Otfried. 1986. Walter Schottky (1886–1976). *Festkörperprobleme* 26:1–5.
Mindell, David. 1996. "Datum for its Own Annihilation:" Feedback, Control, and Computing, 1916–1945. PhD dissertation, Cambridge: MIT.
Mindell, David. 2000. Opening Black's box: rethinking feedback's myth of origin. *Technology and Culture* 41(3):405–434.
Nyquist, Harry. 1927. Thermal agitation in conductors. *Physical Review* 29:614.
Nyquist, Harry. 1928. Thermal agitation of electric charge in conductors. *Physical Review* 32:110–113.
Ornstein, L.S., and H.C. Burger. 1923. Zur Theorie des Schroteffektes. *Annalen der Physik (Series 4)* 70:622–624.
Schottky, Walter. 1918. Über spontane Stromschwankungen in verschiedenen Elektrizitätsleitern. *Annalen der Physik (Series 2)* 57:541–567.
Schottky, Walter. 1922. Zur Berechnung und Beurteilung des Schroteffektes. *Annalen der Physik (Series 2)* 68:157–176.
Schottky, Walter. 1926. Small-shot effect and flicker effect. *Physical Review* 28:74–103.
Schottky, Walter. 1962. Jena 1912 und das $U^{3/2}$-Gesetz: Eine Reminiszenz aus der Vorzeit der Elektronenröhren. In *50 Jahre Entwicklung und Fertigung von Elektronenröhren im Hause Siemens*, reprint from *Siemens-Zeitschrift* 36(2).
Serchinger, Reinhard. 2000. Walter Schottky und die Forschung bei Siemens. In *Oszillationen: Naturwissenschaftler und Ingenieure zwischen Forschung und Markt*, ed. Ivo Schneider, Helmuth Trischler, and Ulrich Wengenroth, 167–209. Munich: R. Oldenbourg Verlag and Deutsches Museum.
Stachel, John, David Cassidy, Jürgen Renn, and Robert Schulmann. 1990. Einstein on Brownian Motion. In *The collected papers of Albert Einstein, Volume 2: The Swiss years: writings, 1900–1909*, ed. Stachel et al., 206–222. Princeton: Princeton University Press.
Stoletov, Aleksandr. 1888. Sur une sorte de courants electriques provoques par les rayons ultraviolets. *Comptes Rendus*. 106:1149
Von Plato, Jan. 1998. *Creating modern probability: Its mathematics, physics, and philosophy in historical perspective*. Cambridge: Cambridge University Press.
Weinstein, P. 1922. *J. C. Poggendorff's Biographisch-Literarisches Handwörterbuch*, vol. 5, 403. Berlin: Verlag Chemie.
Williams, N.H., and H. B. Vincent. 1926. Determination of electronic charge from measurements of shot-effect in aperiodic circuits. *Physical Review* 28:1250–1264.
Williams, N.H., and S. Huxford. 1929. Determination of the charge of positive thermions from measurements of shot effect. *Physical Review* 33:773–788.