The Transistor, an Emerging Invention:
Bell Labs as a Systems Integrator Rather Than a 'House of Magic'

Florian Metzler, PhD
Massachusetts Institute of Technology
8/12/2020

Working Paper

Abstract

The transistor is one of the most consequential human inventions with dissemination of the eventual MOS-FET design estimated to exceed one quintillion devices. However, the transistor's genesis remains poorly understood. Many received accounts associate transistor invention closely with a small group of Bell Labs scientists during the 1947-1948 period. This paper argues that such a view is too narrow. Rather, the transistor – as a solid-state amplifier – emerged over a period of several decades starting with early observations of anomalous amplification in semiconductor crystals and early device designs during the 1910s and 1920s. Other types of relevant knowledge evolved in the form of advances in solid-state physics and materials processing techniques during the 1930s and early 1940s. Bell Labs identified, absorbed, evaluated, and integrated such diverse but interrelated knowledge streams – making Bell Labs appear much more like a systems integrator than the prototypical closed innovation organization it is often portrayed as. The Bell Labs transistor effort was both mission- and device-oriented with the specific goal of turning existing but imperfect solid-state amplifier designs into reliable substitutes for vacuum tubes – as such the research program is better described as applied industrial research rather than basic research. Through its systems integration activities, Bell Labs catalyzed a qualitative shift in the hitherto fragmented semiconductor field, enabling greater resource allocation and intensified research activity, as reflected in a hike in publication growth rates and the later introduction of marketed products. Thus expanded research activity eventually led to the 1959 MOS-FET design as the transistor's dominant design used in large-scale dissemination such as in modern computer chips. Consequently, I propose to view the transistor as an emerging invention – in contrast to a discrete or singular one – with an emergence period spanning several decades. I propose to distinguish between an exploration phase (~1920-1945), a consolidation phase (1945-1950), and a maturation phase (1950-) whereas the intermediate consolidation phase represents a topological transition as is characteristic of emerging fields. Another emphasis of this article lies on the role of informal knowledge in the invention process. In the transistor case, such informal knowledge included patent specifications with proposed device designs, amateur radio magazine articles with reported anomalies, and oral anecdotes in practitioner circles describing experimental configurations of interest. This research asserts that such kinds of informal knowledge played an important role early on in the invention process as they guided both early research campaigns and managerial decisions, including at Bell Labs.

Contents

A. Introduction

1. Open vs. closed: two narratives of the early Bell Labs

The transistor – as the basic building block of the computer chip – has been regarded widely as one of the most consequential devices ever developed. Its cumulative economic and societal impact is without precedence. Given this backdrop, one may wonder: How did this important invention come about? And can we derive generic lessons from its genesis for future inventions of a similar nature?

Conventional wisdom suggests that the transistor resulted from an intense period of basic research at Bell Labs – the industrial research laboratory of AT&T – during the 1947-1948 period. Bell Labs as a research organization has since assumed somewhat legendary status. In the media and popular literature, it has been described as an "Idea Factory" (Gertner 2012) and a "House of Magic" (Silverman 1947). It was said to have been populated by scientists of "true genius" (Daitch & Hoddeson 2002) who "could effectively see into the deepest recesses of the atom, and could theorize inventions no one had previously deemed possible" (Gertner 2012). In the academic literature, Bell Labs has been presented as the prototypical case of "closed innovation" (Chesbrough 2003; Chesbrough 2006): an industrial research laboratory that sweeps up top talent, creates conditions to maximize their potential, and develops new ideas and technologies "from within" (Chesbrough 2006). In this model, ideas originating from outside the firm are viewed with indifference or even suspicion (see Katz & Allen 1982 on the "Not Invented Here" syndrome).

For the practitioner today, neither of these perspectives are particularly satisfying; after all, they leave few generic lessons to be derived about the process of invention that could be transferred to different contexts. If Bell Labs was indeed a unique place populated by unique people – geniuses even –, then there is little hope to recreate – let alone reengineer – the feats accomplished there, unless a new crop of comparable genius is to be found and assembled. If Bell Labs was a prototypical closed innovation organization which, as an organizational model, is said to be no longer compatible with today's economy (as argued by Chesbrough 2006 and others), then we can only study it as a historical curiosity, with little immediate relevance for today.

However, a closer look suggests that both of these perspectives fall short of actual developments. When studying Bell Labs through the entire 1930s and 1940s period, a remarkable openness to outside ideas comes into view: innovation scouts; study groups; knowledge exchange with international scholars; collaborative research networks involving universities, firms, and government organizations – in short: mechanisms for fielding new ideas from a wide variety of sources. Bell Labs employees were able to tap into a rich ecosystem of formal and informal knowledge relevant to solid-state electronics which had been gradually emerging since the early 20th century. Relevant artifacts ranged from abandoned patents and reports of anomalies in amateur radio magazines to recipes for new metallurgical techniques and academic papers on quantum mechanics. When approached from this perspective, Bell Labs appears as an organization that excelled in identifying, absorbing, and evaluating external ideas – and then combining this activity with an internal capacity for specific types of complementary research. Overall, these are characteristics that correspond much more to the concept of an open innovation organization (Chesbrough 2006) rather than a closed one.

A related misconception pertains to the notion that Bell Labs was primarily a "basic research" or "pure research" operation. In actuality, little basic research was conducted at Bell Labs in the context of transistor development. Rather, Bell Labs was an organization that was optimized – maybe better than any other – to absorb the results of basic research conducted elsewhere and to identify and exploit corresponding application potential through "applied industrial research." Leading Bell Labs executives such as Mervin Kelly were explicit about this orientation (Kelly 1950)[1]. That nevertheless the critical role of "applied research" vis-a-vis "basic research" has been downplayed in many secondary accounts might be owed to normative and narrative biases.

2. The transistor as an emerging invention

Shifting the level of analysis from the organization – Bell Labs – to the technology itself – the transistor – also yields new insights. A close look at the early days of semiconductor history suggests that – rather than having originated from a single eureka moment – the transistor emerged from decade-long research efforts involving numerous institutions and individuals.

Documented predecessors of the transistor date back to the 1910s and 1920s. While such devices did in fact exhibit basic transistor functions – most importantly solid-state amplification – their efficiency was poor, their behavior erratic, and their underlying physics poorly understood. The so-called point-contact transistor design, first encountered in 1947 by the two Bell Labs employees John Bardeen and Walter Brattain, represented a major improvement over such earlier devices, even though it still exhibited major flaws. While much relevant understanding of semiconductors had been gained during the 1930s and 1940s (which could then inform material and device design), some aspects of the point-contract transistor remained still unclear, even after the successful demonstration of prototypes. Moreover, the point-contact transistor was still rather unreliable and unsuited for large scale production. Often wires had to be wiggled to recover the amplification effect. It would take another twelve years until a dominant design for the large-scale production of solid-state amplifiers had finally emerged in the form of the metal-oxide-semiconductor field-effect transistor (MOS-FET). As a result, I propose for transistor development to be viewed as a process rather than an event. In other words, the transistor was an *emerging invention* rather than a singular one.

3. Bell Labs as a systems integrator

The process of transistor emergence, comprising many individual contributions across several decades, somewhat relativizes the role of the three Bell Labs employees that are typically emphasized in origin stories of the transistor. Nevertheless, I argue that Bell Labs – as an institution – did indeed play a critical role in the process of transistor emergence and development. Decision makers at Bell Labs recognized early on the need to closely monitor the emerging fields of solid-state physics and semiconductor materials; they subsequently determined when the timing was right to form teams with the objective to integrate the various kinds of relevant knowledge available. When preliminary results had been

[1] Kelly speaks of Bell Labs as an "applied science laboratory" that aims to draw on "areas of pure science that are significant to the industry". He emphasizes that need for "maintaining a close linkage between the forefront of our applied research and that of pure science" but in Kelly 1950 he generally assumes that this kind of pure science originates elsewhere (mostly at universities) and is absorbed into Bell Labs. The timeliness is emphasized by stressing the "minimum reasonable time lag between an advance in pure science and our realization of its contribution to our reservoir of new knowledge." Also consider Morton's 1958 remark that: "American industry has led the world because we have used existing knowledge with ingenuity, energy and innovation."

achieved, they understood their significance and scaled up both research and outreach activities, signaling sustained long-term commitment to the development of the corresponding technologies and applications. Collectively, these actions had a catalytic effect on the fields of solid-state physics and semiconductor materials, motivating other organizations around the world – public and private – to allocate resources to this domain and join expanding development efforts. This kicked into high gear the dynamics that in the long run lead to detailed understanding and large-scale applications of solid-state amplifiers. As such, I argue that in view of early transistor development, Bell Labs can be seen above all else as a *systems integrator* and a catalyst.

4. The challenges of narrative bias

Sadly, even parts of Bell Labs themselves did not actively identify with this role as a systems integrator and catalyst. Instead, much effort, particularly on the initiative of Bell Labs' influential PR department, went into the creation and promotion of narratives that emphasized individual inventors and invention events while downplaying external linkages, institutional factors, and long-term processes. Such tendencies in story telling have been described as "narrative bias" (Heshmat 2016) and in the specific case of Bell Labs as "corporate mythmaking" (Riordan 1998). It might well be that at the time, the narrative of the closed innovation shop populated by geniuses was the only acceptable one from the perspective of Bell Labs' PR department (since the positively connotated alternative of the "systems integrator" or "innovation architect" was not yet conceived or widely recognized). Unfortunately, many subsequent retellings of transistor development uncritically reproduced such early narratives put forth by Bell Labs themselves as well as by personally invested individuals such as William Shockley. As a result, even secondary accounts continued to carry forward a number of misconceptions and distortions.

To learn about the emergence of new inventions and to derive genetic lessons about the underlying process of invention – some of which may be applicable today – we should look at Bell Labs in a dispassionate, unembellished way and appreciate the richness and complexity of the dynamics that brought forth the modern transistor. Some more expansive and balanced efforts to this end have recently been undertaken by several physicists and historians (Lojek 2007; Van Dormael 2004; Arsov 2013). However, none of the received accounts of transistor invention have been composed from the perspective of an innovation scholar. This is the gap that this article aims to address.

B. Phases of emerging inventions: exploration, consolidation, maturation

"Transistor" is a name devised by a Bell Labs committee in 1948 for what can be more generically called a solid-state amplifier[2]. Developments leading up to the emergence of solid-state amplifiers in turn are intimately intertwined with the history of solid-state physics, and semiconductor physics in particular. The name semiconductor was first introduced around 1910 in the form of the German word "Halbleiter" to describe materials that exhibited unusual conduction properties. Before that (and still for some time

[2] In this context, a controllable amplifier (with a gain of 1) can also be used as a switch, as is commonly the case in integrated circuit i.e. computer chip applications.

afterwards), such materials were widely referred to as crystals, as in the case of the "crystal detector" – an early form of a semiconductor diode.

To obtain a first overview of the early semiconductor literature, I conducted a search of publications related to crystals and semiconductors in English and in German during the 1900-1960 period. The results are shown in Fig. 1 TOP.

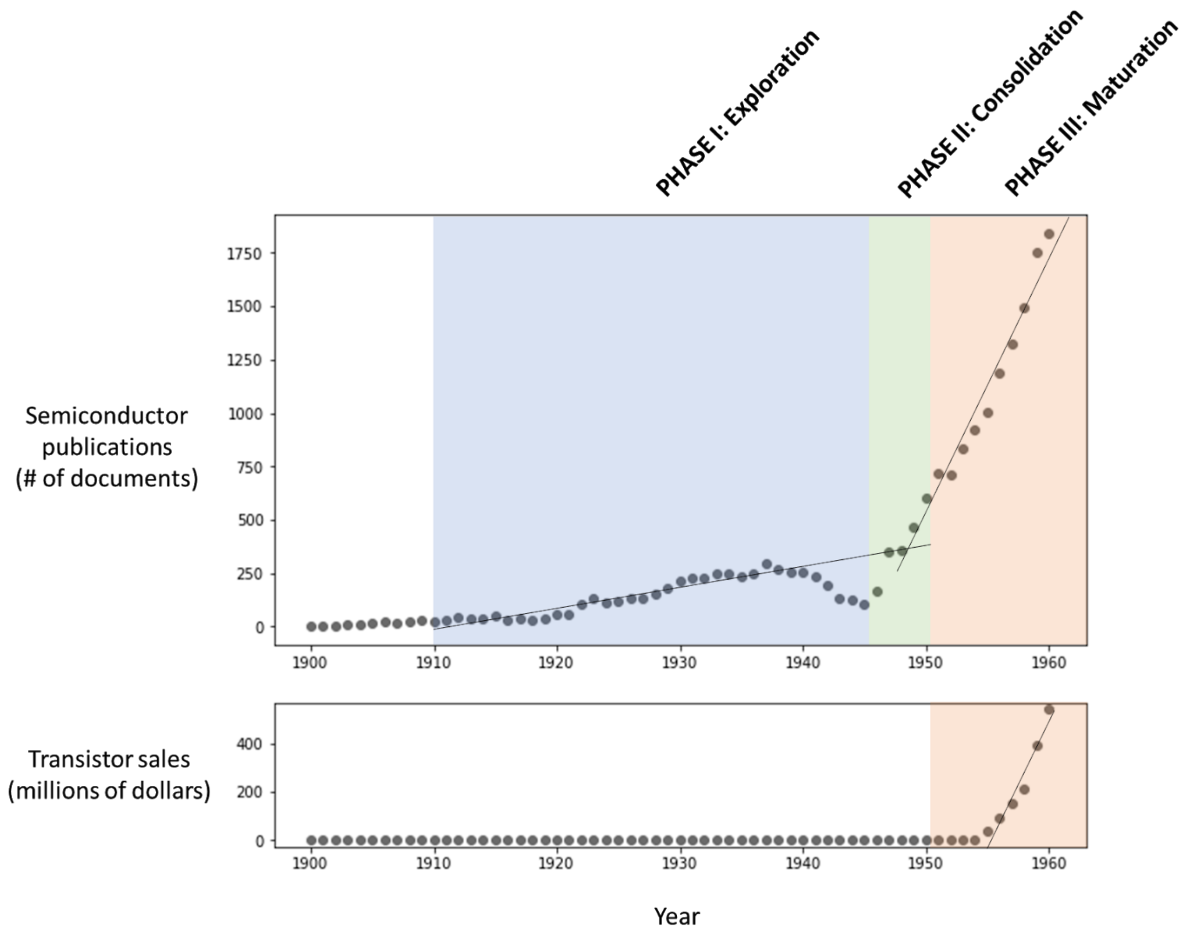


*Figure 1: TOP: Annual number of semiconductor publications 1910-1960. Data obtained via search on SCOPUS with search term: ("crystal" OR "semi*conductor" OR "kristall*" OR "halbleiter*"). BOTTOM: Sales of semiconductor amplifiers (Source: Nelson 1982).*

I added coloration to distinguish between three areas on this graph: an early modest growth phase which I call "exploration phase;" and fast growth phase which I call "maturation phase;" and a phase marking the transition between these two phases which I call "consolidation phase." Regression lines have been added to guide the eye. The growth of publications is fairly consistent during the exploration phase and the maturation phase respectively, with two different growth rates. The 1940-1945 period (World War II) exhibits a dip in publications which does, however, not correspond to a dip in actual research activity (research on semiconductors continued but was considered war-relevant and in large parts prevented from being published during this time).

The consolidation phase, here representing the 1945-1950 period, is then defined as the period that marks the transition between the lower growth rate and the higher growth rate. This is also the period that is most typically associated with the invention of the transistor. It will be discussed in more detail in section D below.

Preceding the consolidation phase is the exploration phase which already exhibits a substantial number of semiconductor publications (more than 5000 cumulatively by 1945). This observation raises the question what role these publications played for subsequent semiconductor development and what underlying research activity they emerged from. More specifically, how did research during the exploration phase affect the subsequent consolidation phase? This will be the subject of section C.

Finally, the proposed maturation phase – the phase following the consolidation phase – is characterized by the development of commercial products, as is reflected by the beginning of transistor sales occurring around the mid-1950s (Fig. 1 BOTTOM). Incremental improvements in transistor designs and manufacturing techniques during this period eventually led to the transistor's dominant design in 1959: the so-called MOS-FET. A dominant design is defined as "as the specification (consisting of a single, or a complement, of design features) that defines the product category's architecture" (Srinivasan et al. 2006). In practice, the emergence of a dominant design is typically associated with the ability to scale the deployment of a technology. Suarez & Utterback 1995 describe the MOS-FET as a prototypical case of a dominant design. With the MOS-FET finally satisfying conditions for massive scaling, more than one quintillion (10^{18}) MOS-FET devices have been produced and deployed since then.

C. Phase I: Exploration

While most invention narratives of the transistor focus on the 1945-1950 period – a period which I refer to as the consolidation phase – this section traces back into the preceding period the evolution of different kinds of knowledge relevant to the development of solid-state amplifiers. The kinds of knowledge that turned out to be relevant to that end can be identified in view of what we know today about the subsequent phases.

Concretely, I distinguish here between four knowledge domains and one economic domain – each of which can be viewed as representing a condition where crossing a particular threshold contributed to motivating subsequent consolidation efforts:

1. Proposed concepts
2. Observed anomalies
3. Materials and diagnostics
4. Theory and models
5. Economic demand

The five categories represent a simple framework for analyzing emerging inventions during their exploration phase[3]. The categories populated with critical milestones relevant to the development of solid-state amplifiers is shown in Fig. 2. The remainder of this section will now motivate and discuss the content of this figure in some detail, stepping through each category.

[3] The framework may be general enough to be applicable to other technologies as well.

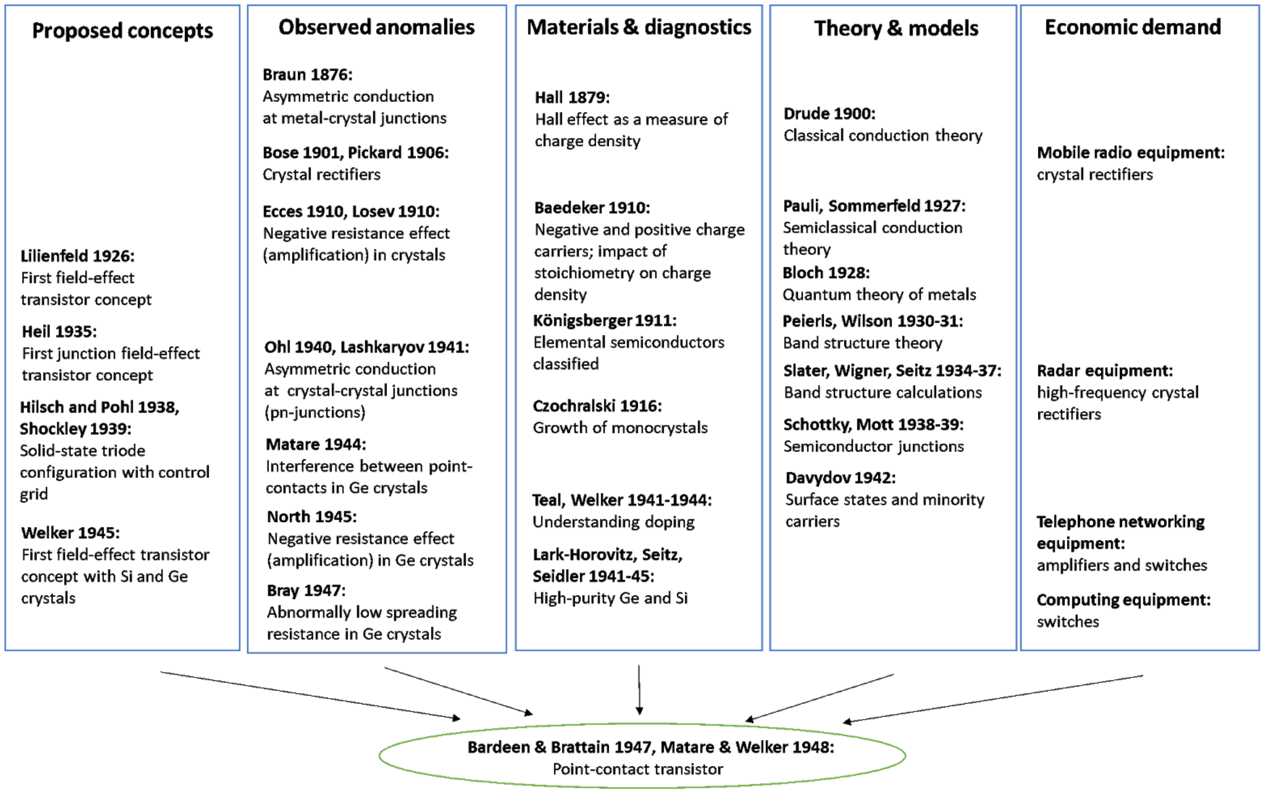


Figure 2: Simple framework to characterize the exploration phase: five factors are proposed whose respective advancements are viewed as preconditions for subsequent integration during the consolidation phase (which resulted in the point-contact transistor design).

1. Proposed concepts

Early explicit designs of controllable solid-state amplifiers can be traced back to the 1920s. Over time, different variants and refinements evolved which became well-known among researchers in the field of applied solid-state physics. Publications occurred mostly in the form of patents and in some cases in the form of journal articles. This section will summarize important milestones in this category before establishing the relevance of this earlier work for the later Bell Labs efforts.

The earliest codified record of the transistor principle is typically attributed to Julius Lilienfeld who filed a patent in 1925 with the title "Electric current control mechanism" (later followed by patents "Device for controlling electric current" and "Amplifier for electric currents"). Lilienfeld held a PhD from the University of Berlin and started his career as a professor at Leipzig University before emigrating to the US in 1921. In the patent specifications, Lilienfeld proposed the principle of "affecting, as by suitable incoming oscillations, a current in an electrically conducting solid or such characteristics that said current will be affected by and respond to electrostatic changes" (Lilienfeld 1930). Effectively, this design represented a semiconductor diode overlaid by a capacitor whose electric field modulates the semiconductor's resistance. As such it represented the first articulation of the Field Effect Transistor (FET) principle which would become the dominant transistor technology from the 1960s onwards (see

Arns 1998, Arsov 2013). However, despite their anticipatory nature, Lilienfeld's patents had a number of shortcomings. First, the semiconductor materials he proposed were compounds such as copper sulfide (one of the most common semiconductor materials at the time). Copper sulfide is indeed a semiconductor and can in principle be used for transistor action; however, its properties are difficult to control and to obtain consistency with, as will be further explained in the sections below. As a result, if devices were built per Lilienfeld's specifications at the time, they were likely unreliable and highly variable. Later attempts to replicate Lilienfeld's designs resulted in devices that did indeed exhibit amplification but were also indeed unreliable (Crawford 1991; Arns 1998). Second, Lilienfeld's theoretical arguments[4] are qualitative in nature due to an incomplete understanding of conduction in semiconductors at the time (see section C4 below). As a result, various dimensions of the proposed amplifier were not precisely specified and required some experimentation to be implemented. For the same reason, an optimal implementation of the design that would maximize efficiency and reliability could not be easily devised.

Another patent that received widespread attention was Oskar Heil's 1935 patent "Improvements in or relating to Electrical Amplifiers and other Control Arrangements and Devices". This patent reflects progress in materials and theory made since Lilienfeld's work and is regarded as the first articulation of the principle of the Junction Field Effect Transistor (JFET). Heil (1935) suggests: "if a semi-conductor be arranged as to form part of a condenser which is subjected to a varying voltage charge the resistance thereof will vary as a function of said varying voltage and according to this invention this phenomenon or effect is utilised for amplifying or other control purposes." Heil held a PhD from Gottingen University based on his research on molecular spectroscopy and later became a collaborator of Ernest Rutherford at Cambridge University's Cavendish Laboratory.

Surprisingly, neither Lilienfeld nor Heil published academic papers corresponding to their solid-state amplifier concepts. A possible explanation is the poor reputation of the field of semiconductor physics which was held in disregard by many leading physicists into the late 1930s due to the unpredictability and inconsistency of experiments in this area (see section C2 for more detail).

The first academic publication describing a controllable solid-state amplifier concept was Rudolf Hilsch and Robert Pohl (1938). Pohl was a physics professor at Gottingen University who Nevill Mott later called "the real father of solid-state physics" (Mott 1976). Mott remarked that many physicists at the time considered research on semiconductor as "dirty physics" and that it was Pohl's systematic approach that "made it 'clean' and a precise branch of science." Hilsch, Pohl's student at the time, later also became a professor at Gottingen University. Their design comprised a potassium bromide single-crystal as a semiconductor into which a platinum control electrode was melted. Their paper with the title "Control of electric current by a three-electrode crystal and model of a depletion layer" is also the first description of a solid-state amplifier accompanied by experimental data. Hilsch and Pohl demonstrated amplification of 20x which they claim could be achieved reliably. The shortcomings of

[4] Lilienfeld does provide a theoretical picture which refers to a prevalent theory of conduction at his time: J.J. Thomson's concept of atoms and electrons in dipole form whose resistance depends on their orientation – which in turn is influenced by external fields. Lilienfeld (1930) explains in the patent: "varying with the electric field at this point; and in this connection it may be assumed that the atoms (or molecules) of a conductor are of the nature of bipoles. In order for an electron, therefore, to travel in the electric field, the bipoles are obliged to become organized in this field substantially with their axes parallel or lying in the field of flow. Any disturbance in this organization, as by heat movement, magnetic field, electro static cross-field, etc., will serve to increase the resistance of the conductor; and in the instant case, the conductivity of the layer is influenced by the electric field."

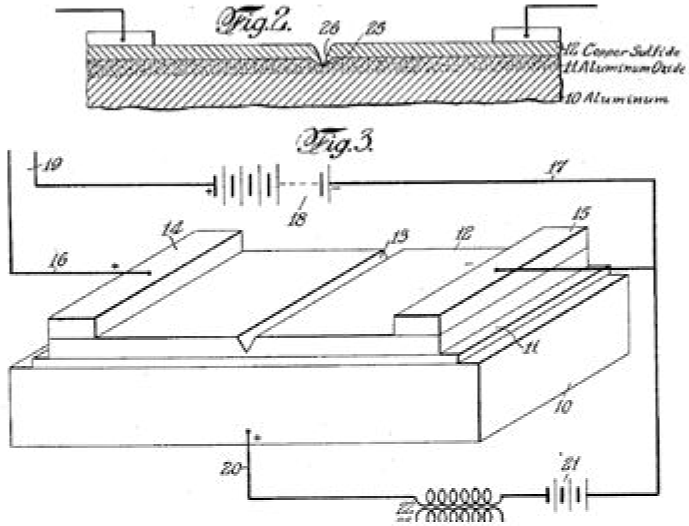
their design – which left it unsuitable for commercial use – was the slow frequency response which was limited to 1 Hz. A patent describing a similar concept was filed already in 1935 by the Dutch industrial conglomerate Philips (Holst and Geel 1939). Shockley at Bell Labs attempted an analog approach in 1939 by using a copper oxide grid as a control electrode in the midst of a copper sulfide crystal. However, Shockley achieved no amplification from his design. It is not known whether Shockley was aware of the publication by Hilsch and Pohl or of the patent by Holst and Geel in previous years.

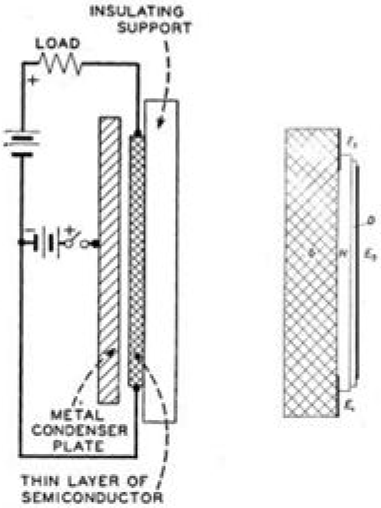
Finally, a patent application for a solid-state amplifier was filed by Heinrich Welker in June 1945 titled "Semiconductor arrangement for the capacitive control of currents in a semiconductor crystal". Welker held a PhD from the University of Munich obtained under Arnold Sommerfeld. This patent by now reflected the full extent of progress that had been achieved in materials and theory (see sections C3 and C4). It explicitly referred to purified germanium and silicon as preferred semiconductor materials and justified design choices based on modern theoretical notions of solid-state physics such as depletion and inversion layers. The patent also discussed junction configurations such as an N-P-N-P configuration (which anticipates what would later become known as a thyristor). However, Welker's patent did not get published until the 1950s and was unlikely to be known to Bell Labs staff in 1947/48.

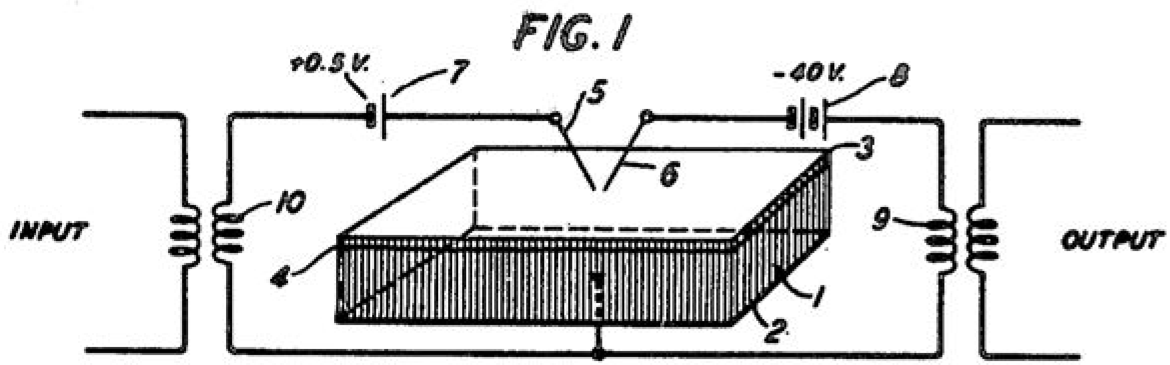
In the later descriptions of their own work on solid-state amplifiers, Bell Labs researchers did not reference any of the patents or publications listed above which prompted some chroniclers to assume that they had been unknown to Bell Labs[5]. However, upon being prompted directly by a journalist about Lilienfeld's patents in 1964, a Bell Labs researcher stated that their semiconductor team was in fact aware of such earlier work and engaged in replication attempts which, however, yielded only negative results. In contrast to this statement, Arns (1998) later found that patent affidavits of 1948 describe attempts by Bell Labs staff Shockley and Pearson to replicate Lilienfeld's designs which did yield positive results. In at least one case, amplification on the order of 11% was obtained[6]. An effect of this magnitude would have not been relevant in view of applications but would have certainly been of value in guiding and motivating the research program. It can thus be concluded that Bell Labs researchers were indeed aware of earlier literature with proposed concepts for solid-state amplifiers, digested such literature to understand corresponding advantages and shortcomings, and built off of it.

[5] Previous work was mentioned somewhat vaguely in the first transistor patent application of 1948 in a way that has been described as acknowledging but downplaying the significance of prior art (Bardeen & Brattain 1950): "Attempts have been made in the past to convert solid rectifiers utilizing selenium, copper sulfide, or other semi-conductive materials into amplifiers" Note that rather than stating that these devices did not work, the patent attorneys that authored the patent specifications chose to write in the following paragraph: "past devices […] do not appear to have been commercially successful."

[6] Johnson's sworn affidavit is reported by Arns (1998): "The first [Lilienfeld] patent I undertook to investigate was […] 1,900,018 […]. Prior to my assignment […] two other members of the Laboratories' staff, namely, Wham Shockley and Gerald L. Pearson, had investigated the performance of a structure which is the same in all particulars except one, namely, that the insulating film, instead of being aluminum oxide, was quartz." They report a small but not insignificant effect: "although the modulation index of 11 per cent is not great, […] the useful output power is substantial […] it is in principle operative as an amplifier."







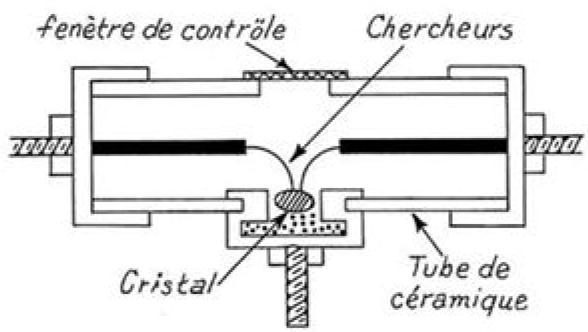


Figure 3: Proposed concepts for solid-state amplification devices: TOP LEFT: Lilienfeld's 1925 field-effect transistor design (Lilienfeld 1930); TOP RIGHT: Shockley's and Welker's 1945 field-effect transistor designs (see Handel 1999). MIDDLE: Bardeen & Brattain's 1947 point-contact transistor design (Bardeen & Brattain 1950); BOTTOM: Matare & Welker's 1948 point-contact transistor design (Matare & Welker 1954). Note the similarities across the upper field-effect designs as well as the similarities across the bottom joint-contact designs.

2. Observed anomalies

The concepts of solid-state amplifiers discussed in the previous section would have seen less traction, had they not been corroborated by a host of reported anomalies surrounding semiconductor materials, anomalies that were in principle encouraging of such concepts. This section describes some of the best known of such anomalies, most of which could not be explained until the late 1930s or early 1940s (when theory had progressed accordingly). The section concludes again by turning to Bell Labs and considering the relevance of observed anomalies to the semiconductor program at Bell Labs.

Among the earliest anomalies surrounding crystalline materials was the recognition by Ferdinand Braun in 1876 that certain metal-crystal connections exhibited "asymmetric conduction" i.e. that electric current could flow through them better in one direction compared to the other. This recognition is typically associated with the discovery of semiconducting properties. The crystal-metal connections for which this behavior was recognized essentially represented the first semiconductor diodes. The asymmetric conduction behavior remained unexplained until the work of Mott (1938; 1939) and Schottky (1938; 1939) – see section C4 on theory.

By the early 1900s, asymmetric conduction was recognized to be useful for rectifying purposes i.e. for turning AC signals into DC signals as is necessary when receiving modulated radio signals. This application motivated Boston-based radio engineer Greenleaf Pickard to test hundreds of materials for their rectifying properties. He found that best performance could be achieved from a group of materials that include zinc oxide, copper sulfide, copper oxide, galena (lead sulfide), and carborundum (silicon carbide) – many of which remained semiconducting material of choice for several decades. Semiconductor rectifiers became known as "crystal detectors" and became widespread among radio operators. In crystal detectors, a thin metal wire (also known as a "cat-whisker") touched the surface of a semiconductor crystal (see Fig. 4 TOP) to form a semiconductor-metal interface (later known as a Schottky junction).

Frequent users of such rectifying crystals – including Pickard – also reported occasional amplification as another kind of unexplained behavior. Amplification was observed in the form of sudden increases of signals passing through an in-line rectifier and circuits bursting into spontaneous oscillation (while drawing more current from the connected power supply). Such in-line amplification came to be referred to as "negative resistance" phenomenon (since a hike in current draw corresponds to a drop in resistance).

In 1924, Oleg Lossev, a Russian radio engineer reported in the US-based magazine Radio News: "Several experimenters have observed that some contacts, such as crystal and metal or crystal and carbon generally employed as detectors may produce undamped oscillations of any frequency, exactly as the vacuum tube oscillator. The same contact may also be utilized as an amplifier." Fig. 4 MIDDLE shows a metal-semiconductor configuration by Lossev that was claimed to exhibit amplification. Fig. 4 MIDDLE also shows corresponding experimental data exhibiting negative resistance. Because of such and similar reports, some authors credit Lossev with the invention of the first solid-state amplifier (see Lee 2004).

In a 1925 article in Radio News, Pickard summarized crystal amplification reports known to him. Specifically, he referenced earlier work by William Eccles, a British physicist to whom the term "diode" is attributed and a collaborator of radio pioneer Marconi. Pickard noted that Eccles already "in May 1910 demonstrated before the Physical Society [of London], a galena crystal combination capable for generating oscillations." Pickard then reported of his own attempts to reproduce Eccles' results: "I repeated the Eccles experiments with the galena detector [..] and then I used silicon, zincite fused zinc oxide and pyrite. In each case I was more or less successful in producing sustained oscillations. In all cases, of course, a condition of oscillation was obtained only after careful adjustment of the contact point on the detector and experimentation with the voltage applied across the crystal. […] the adjustment required by the crystal is extremely fussy as compared with that necessary when it is to be used for detection [as a diode] […] under certain conditions the crystal had the property of negative resistance."

Such reports led the editor of Radio News, Hugo Gernsback, to comment in 1925: "The crystal now actually replaces the vacuum tube. That this is a revolutionary radio invention need be emphasized no further. […] we can not only detect [rectify] with the crystal, but we can also amplify with it. [..] the oscillating crystal also explains now how some radio experimenters have been able to obtain such remarkable long distance records with crystal outfits. It would seem that wherever these records were made, the crystal actually oscillated in one way or another without the user being aware of it." However, Gernsback added: "we must sound a note of caution. It must be understood that for the present, the invention is practically confined to the laboratory. […] as wonderful as the invention is, it still has all the troubles and weakness of the crystal. There is the usual cat-whisker contact and the usual elusive sensitive spot. […] it may take many years for the oscillating crystal to be perfected in such a manner that it will supersede the vacuum tube, but we predict that such a time will come."

Both Pickard and Gernsback emphasize the unreliability of the reported amplification phenomenon. Such erratic behavior was commonplace when working with semiconductor crystals. Radio operators often had to patiently search for a suitable spot on the surface of their crystals to obtain rectification (i.e. diode behavior). According to Pickard, occasional amplification effects were even more difficult to obtain than rectification. From today's perspective, the erratic nature of this behavior can be explained by the polycrystalline nature of these simple early semiconductor materials whereas each crystal grain would be expected to exhibit a slightly different composition of materials and impurities – and therefore different electron concentrations. Each grain – which often spans just micrometers – then represents a different type of Schottky junction when touched by the cat-whisker wire. However, without such hindsight, working with semiconductor crystals was considered more of an art than a science and notoriously difficult to experiment with.

Because of the poor reproducibility of observations and the lack of theoretical understanding, many scientists avoided applied solid-state physics altogether, even questioning the existence of semiconductors as a separate category of materials. Frederick Seitz noted that "such variability, bordering on what seemed the mystical plagued the early history of crystal detectors and caused many of the vacuum tube experts of a later generation to regard the art of crystal rectification as being close to disreputable." (Riordan 1998). Along similar lines, Wolfgang Pauli wrote to his assistant Rudolf Peierls in 1931: "One shouldn't work on semiconductors, that is a mess, who knows whether semiconductors even exist." (see Handel 1999)[7]. John Brattain reflected later: "The difficulty in this period was that copper oxide [is] such a messy type of structure-sensitive thing." (Riordan 1998). Among radio engineers, certain manufacturers, mines, and even batch numbers of materials were known to exhibit the best rectifying behavior – a degree of variability that left many academic physicists uncomfortable about conducting systematic research in this area (see Cuff 1993). Mervin Kelly at Bell Labs later

[7] Referencing the debate whether there can be elemental semiconductors or whether semiconducting behavior would always be a function of additives alone.

remarked (Wolff 1983): "In the Bell System, we used only the copper from a particular mine for fabricating copper oxide telecommunication varistors. It made the best. We did not know why. Processing of selenium rectifiers was largely an art – cookbookery." As late as 1958, a Bell Labs handbook on transistors referred to the practices of dealing with certain semiconductor surfaces as ''Folklore and Witchcraft" (Bridgers & Biondi 1958).

Asymmetric conduction at metal-semiconductor interfaces could finally be explained with Schottky's 1938/1939 publications (Schottky 1938; Schottky 1939 – see section C3). However, soon after, asymmetric conduction was also observed at certain semiconductor-semiconductor interfaces which Schottky's theories could not explain. In 1939, Russell Ohl, an engineer at Bell Labs, noticed that some grain boundaries in polycrystalline silicon also exhibited asymmetric conduction (Ohl 1976). The equivalent observation was made independently by Ukrainian physicist Vadim Lashkaryov, and published in 1941. Such interfaces later came to be known as P-N junctions (see section C4).

Further anomalous observations surrounding semiconductors were noted during the World War II period. H. Q. North, a General Electric engineer involved in the US radar program, repeatedly noted negative resistance behavior in germanium crystals used as crystal detectors for radar receivers (documented in North 1945 and Torrey & Whitmer 1948). Other members of the US radar program – in particular at Purdue University which was responsible for germanium crystal research – observed unexpectedly low resistance from metal-germanium contacts (Bray 1982) – an effect later attributed to hole injection which is central to the operation of the point-contact transistor. Similar observations were made independently in Germany: Herbert Matare was a leading member of Germany's radar program, working at the Telefunken laboratory (a Siemens and AEG joint venture). During his work with high-purity germanium crystals for radar receivers, Matare noticed that two close metal contacts on a germanium crystal were influencing one another. The configuration resembled the later point-contact transistor which Matare went on to co-invent at a Paris-based Westinghouse laboratory in January 1948, independently of the same invention at Bell Labs in December 1947 (Handel 1999; Van Dormael 2004; Riordan 2005; Arsov 2013 – see sections C4 and D1).

This section suggests that a wide range of anomalies surrounding semiconductors had been observed and described in the decades preceding the 1945-1950 period. The team at Bell Labs had knowledge of most, if not all, of the anomalies described in this section. The anomalous behavior of the P-N junction was observed in its own facilities. The negative resistance behavior of germanium crystals was described in reports that circulated among members of the US radar program, which Bell Labs was a part of. The anomalous resistance behavior on germanium crystals were also directly reported to Bell Labs by the researchers at Purdue (Bray 1982). Regarding the asymmetric conduction of certain semiconductors, virtually every electrical engineer and physicist since the early 20th century had been aware of this unexplained behavior of crystal detectors.

No direct documentation is available whether Bell Labs employees were also aware of the early amplification reports in publications such as Radio News. However, a number of points suggest that this was likely the case. Mervin Kelly, the Bell Labs executive who later initiated the solid-state research program was the head of the Bell Labs vacuum tube department during the 1920s. Since radio magazines represented a major outlet for discussion of vacuum tube technology and since Radio News had print runs on the order of 200 000, it is certainly plausible that Kelly may have encountered during the 1920s some of the anecdotal reports of erratic solid-state amplification. What is certain is that the Bell Labs solid-state research group was aware of reported amplification anomalies by the 1930s at latest. Ohl recalled in an interview (Ohl 1976) that he had known early on of reports of anomalous amplification in semiconductor crystals, and even specifically of a configuration with two metal contacts on the crystal surface (i.e. a three-electrode configuration, considering the base contact)[8]. Moreover, on behalf of Ohl, Bell Labs submitted several patent applications during the World War II period describing "negative resistance" solid-state amplifiers (Ohl 1949, 1950; see Fig. 4 BOTTOM). These devices closely resembled the earlier solid-state amplifiers proposed by Lossev (see Fig 4 MIDDLE) although no references were made. Ohl also built a small radio receiver from such devices which Bell Labs executive Kelly arranged to be demonstrated to other Bell Labs staff, especially theorists such as William Shockley. Shockley later recalled (Riordan 1998): "Ohl demonstrated that amplified radio broadcasts could be heard over a small loudspeaker [from a tubeless radio set]." However, Shockley also noticed that "gross instabilities due to thermal effects made this amplifier erratic and unreliable." Still, he considered it "indeed an exciting solid-state development." Shockley's assessment of the shortcomings of this early semiconductor amplifier echo the grievances voiced by Pickard and Gernsback years earlier.

In the context of this demonstration at Bell Labs, the dynamic between Ohl, Shockley, and Kelly is remarkable. Ohl was an older technician and engineer who was known as "a systematic tinkerer like Edison and de Forest" (Riordan 1998) with his finger on the pulse of the radio equipment community. However, Ohl had no PhD and not the kind of state-of-the-art theoretical training as Shockley. Shockley in turn was an ambitious young physics PhD who had been recruited right out of MIT's solid-state physics program by Bell Labs executive Mervin Kelly. Kelly apparently recognized the potential implicit in the semiconductor anomalies but also understood that without a thorough understanding of underlying mechanism, industrial scale exploitation of such principles would not be possible[9]. As a result, Kelly brought together Ohl with his informal, empirical knowledge and Shockley with his formal, theoretical knowledge to guide and motivate the work of Shockley's semiconductor group over the following years.

---

[8] Ohl 1976: "[A colleague] gave me a copy that he had of, I think it was *Electrician*. It was a British magazine, one of these big paged things, you know. In it was a translation from a Russian paper in which they had used carborundum [silicon carbide] with two contacts and a battery supplying one of the contacts and had gotten a power gain of ten times. And this was way back in the 1910s, so the fact that you could get a power gain had been known, but it was never put on a controlled basis. I knew about it because an operator of the Signal Corps back in 1919 had told me that some of the operators used carborundum as oscillators for receiving. When I had seen this article that Curtis gave me, I was not astounded because I had known about this before I ever saw the article. I had heard about it. I knew a former first sergeant in the Signal Corps who had lived in, the boarding house that I lived and he was an expert radio operator. He told me a great deal about the use of crystal detectors on ships. He told me that professional operators carried two crystal detectors with them. One of them was made of carborundum and one of them was something like galena [lead sulfide] or something of that sort. He said the carborundum was used for two purposes. They used it in the harbor when they were close to a transmitter to prevent burnout. They also used it at long distances with two points. One point was excited with a battery and they were able to get long wave oscillations out of it."

[9] This is reflected in this excerpt in Kelly 1950: "Until the beginning of this century, the work in applying new scientific knowledge to new facilities and instrumentalities for society was quite a hit and miss process. The inventor, having little or no direct contact with pure science, took the first steps whose end product was the inventor's model. Then the engineer, who at that time was largely a graduate from the drafting board or from the shop, reduced the inventor's model to the design of a new product for manufacture. While progress was made, the procedure was slow, inefficient, and the intervals of time were quite long between the availability of new scientific knowledge and the appearance of new products made possible by it. The break from this pattern began at the turn of the century with the appearance in industry of men trained in the scientific method of research. They were the pioneers of industrial research [..]."

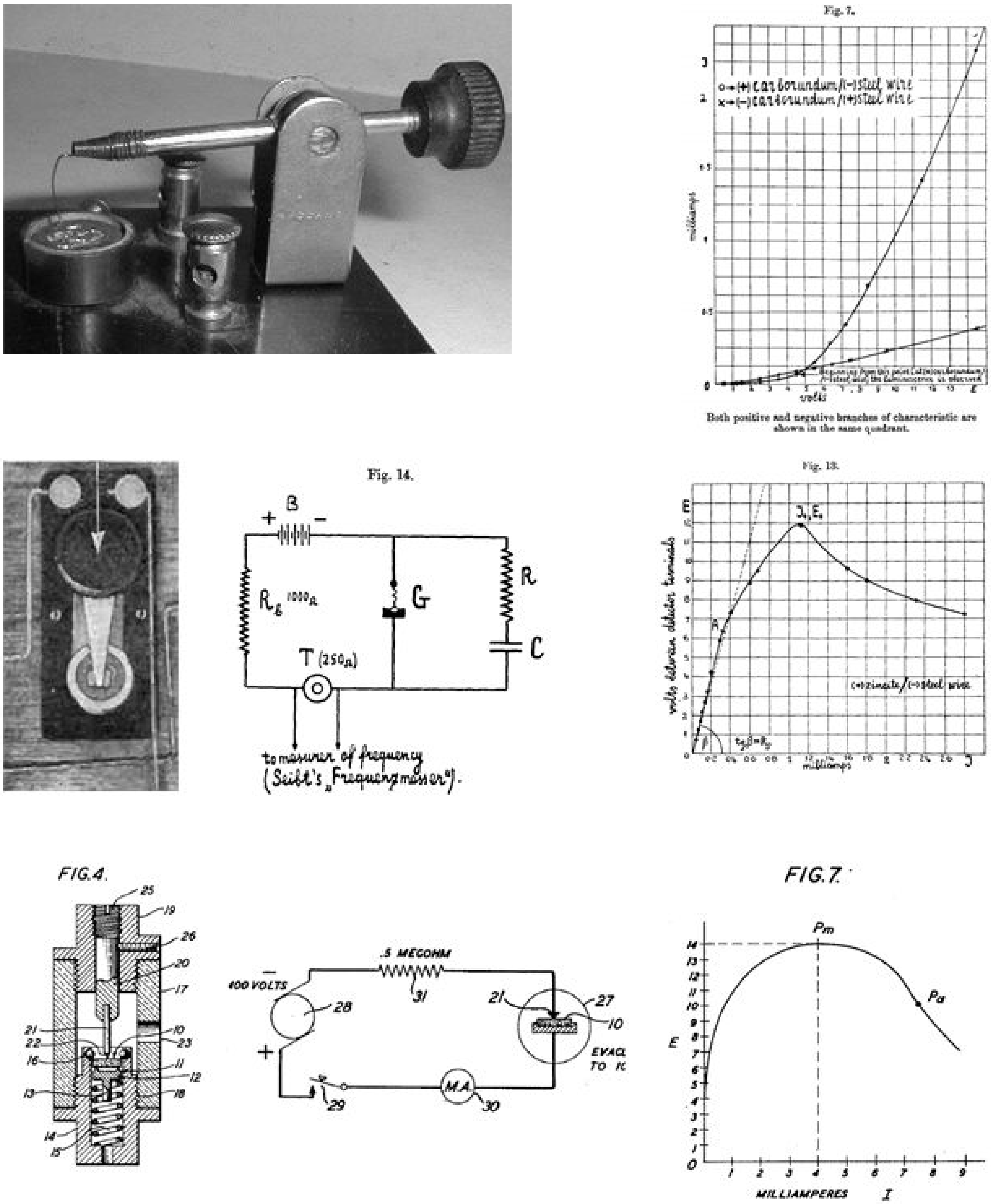


Figure 4: Experimental configurations that exhibited anomalous behavior (left and middle) and corresponding data (right). TOP: A typical cat-whisker crystal detector as commonly used for rectification of incoming signals in radio receivers. The metal wire typically had to be adjusted on the surface of the underlying semiconductor crystal to find a spot that exhibited the desired asymmetric conduction behavior. Asymmetric conduction is shown in the data on the right where the current is much higher in one direction than the other. MIDDLE: Lossev's 1924 solid-state amplifier (known as Crystodyne) which exhibited negative resistance behavior i.e. amplification. The data on the right shows a current drop despite rising voltage. BOTTOM: Ohl's 1945 solid-state amplifier (known as "point-contact negative resistance device" which also exhibited negative resistance behavior i.e.

amplification. Note the similarities between Lossev's and Ohl's configuration. Nevertheless, Lossev was not cited in Ohl's Bell Labs patent which might have been due to prior art concerns of Bell Labs lawyers.

3. Theory and models

The observed anomalies discussed in the previous section remained anomalies until new theoretical explanations could account for them. In many cases, such anomalies provided the original impetus for corresponding theory development; in other cases, explanations for anomalies followed from new theory incidentally. The area of physics most closely associated with semiconductors and crystals was solid-state physics, and specifically conduction theory. Moreover, the prediction of properties and behavior of certain real-life materials via computational techniques represented a bridge between theory and experiment. To match theory and experiment, materials had to be as simple as possible – ideally monocrystalline and without impurities. Modern conduction theory resulted from the application of quantum theory to solids, a process pioneered by German and Swiss university groups during the early 1930s. By the mid-1930s, many of the respective pioneers left Europe in response to Hitler's rise and the US became a new center for solid-state physics. Bell Labs researchers followed the development of modern conduction theory closely, often via seminars and published materials and in some cases via direct interaction or training with leading researchers of the field.

Around 1900, the established notion of electric conduction viewed solids as a lattice of atoms between which electrons move like a gas (Drude 1900) – a classical picture. However, this picture could not explain certain empirical observations such as the temperature dependence of semiconductors (lower resistance at higher temperatures) which was inverse to metals. First attempts to incorporate quantum theory into models of solids were made by Arnold Sommerfeld (1927) who extended the Drude model to give the electron gas quantum behavior (which resulted in the so-called free electron model).

Much subsequent refinement of conduction theory resulted from Sommerfeld's first- and second-generation students. Werner Heisenberg – a Sommerfeld protege turned professor – advised Felix Bloch, Rudolf Peierls, and Alan Wilson at Leipzig University. Bloch first applied quantum theory to a periodic crystal lattice which was viewed as a first approximation of a solid (Bloch 1929). This resulted in calculations for energy states of electrons which, depending on the lattice and its composition, turned out to bunch together in some areas of the energy spectrum and not in others. Such clustering of electron energy levels led to the notion of "energy bands" as well as "band gaps," as developed in Peierls (1932) and Wilson (1931). In this "band structure" picture, some electrons were associated with holding together the lattice while others with electric conduction. Differences across these groups provided a basis for distinguishing between metals, semiconductors, and insulators – and for reconciling some of the unexplained semiconductor behavior. Wilson's work also implied that materials can differ in their "electron density" and that impurities added to a lattice can affect electron density and thus conduction properties.

With a generic quantum model of solids and conduction behavior now available, several follow-on efforts ensued. Such efforts considered how the proposed models would be affected by more realistic conditions, going beyond the simplifying assumptions of an infinite lattice that Bloch had adopted. This included the consideration of interfaces, surfaces, and defects in materials. A key development represented the work on interfaces between metals and semiconductors by Nevill Mott at the University of Bristol (1938, 1939) and Walter Schottky at the Siemens Research Laboratories (1938, 1939).

Metal-semiconductor interfaces was exactly what the anomalous crystal detectors with their asymmetric conduction behavior represented. Following band structure theory, here the differing band structures of two adjacent materials meet and a picture was needed for the detailed behavior of such interfaces – which in turn would determine macroscopic conduction. This led to the notion of a depletion layer at interfaces between different materials which could now finally explain asymmetric conduction – around 60 years after Braun's first reports. Metal-semiconductor junctions such as the early crystal detectors have since been called Schottky diodes.

A related problem concerned interfaces between two different types of semiconductors. In a series of papers published in English and German during the late 1930s, Ukrainian physicist Boris Davydov (1937, 1938, 1939) laid out an advanced theory of semiconductor interfaces that was more general than Schottky's. Davydov proposed that the copper oxide semiconductors in crystal detectors comprise a double layer with an area of excess electrons (later called N-type) and an adjacent area of lacking electrons (later called P-type). Lacking electrons were in later models referred to as "holes." Davydov's model also predicted the movement of holes through excess electron areas as well as the movement of electrons through lacking electron areas. In these cases, electrons and holes can be described as "minority carriers" of charge since, in the respective materials, charge carriers of the opposite type dominate. The possibility of minority charge carriers had previously been ignored under the assumption that electrons and holes would quickly recombine and therefore neutralize. However, Davydov suggested that minority and majority carriers could co-exist long enough to become macroscopically relevant. From Davydov's work followed that monocrystalline semiconductors are advantageous for minority carrier conduction as such crystals allow for farther movement of minority charge carriers (electrons or holes) before recombination.

Other important work of the 1930s includes research on the behavior of crystal lattices with defects (see Frenkel 1931, Schottky 1939) as well as the behavior of crystals at surfaces. Crystal defects explain conduction dynamics relevant to the photovoltaic behavior of semiconductors (but less relevant to early solid-state amplifiers). Like the study of interfaces, the study of semiconductor surfaces was motivated by the desire to describe semiconductors in more realistic and less idealized forms. At a real-life surface, the uniformity of a regularly spaced lattice with its atoms and electrons is broken. This leads to an expected modification of electron bands within the top few atomic layers of the material (sometimes referred to as "band bending" or "surface states"). The Ukrainian physicist Igor Tamm[10] introduced the notion of surface states in his 1932 publication "On the possible bound states of electrons on a crystal surface" and discussed their impact and relevance.

A further line of research concerned the application of the new quantum physics-based theoretical models to concrete materials, the computation of expected macroscopic behavior, and the comparison with experimental results. Two centers for such work emerged in the US around John Slater at MIT and Eugene Wigner at Princeton University (who had migrated from the University of Berlin). PhD students of Slater and Wigner included William Shockley and John Bardeen who were tasked to carry out early band structure calculations for sodium chloride (Shockley 1936) and sodium (Bardeen 1936).

Bell Labs was able to absorb much of the state-of-the-art theoretical knowledge in solid-state physics in a number of ways. One means was recruitment: Bell Labs manager Mervin Kelly recruited freshly minted PhDs from MIT and Princeton as well as Caltech: this includes William Shockley, James Fisk, John

[10] Tamm received the 1958 Nobel Prize in physics although for a different contribution.

Bardeen, Charles Town, Dean Woolridge, and John Pierce. Shockley and Bardeen in particular were intimately familiar with much of the literature and developments discussed in this section, having been trained at two of the leading centers for solid-state physics in the US. Other Bell Labs employees of the later semiconductor team such as experimental physicist Walter Brattain had their own share of early exposure to the global solid-state physics community. Brattain attended a summer school with Arnold Sommerfeld in 1931 at the University of Michigan while already a Bell Labs employee. Other Bell Labs personnel also had several contact points with centers of solid-state physics in Europe. Mervin Kelly and Karl Darrow, another Bell Labs scientist, frequently visited universities in Europe, particularly in Germany, Switzerland, and England to interact with leading scientists (Kelly 1976; Darrow 1964). Both Kelly and Darrow had obtained physics PhDs under Milikan in Chicago. Another means of exposure to recent theoretical work was through literature and study sessions: Darrow regularly wrote summary and review articles on new developments in European physics for US colleagues, some of which were published in nationwide journals and others in Bell Labs' internal journal. German-speaking members of the wider Bell Labs network such as Darrow and Purdue University's Lark-Horowitz monitored German publications and occasionally translated or summarized relevant articles (Van Dormael 2004). Through much of the 1930s and 1940s, Bell Labs personnel organized weekly study groups where relevant papers and text books in solid-state physics were identified and discussed.

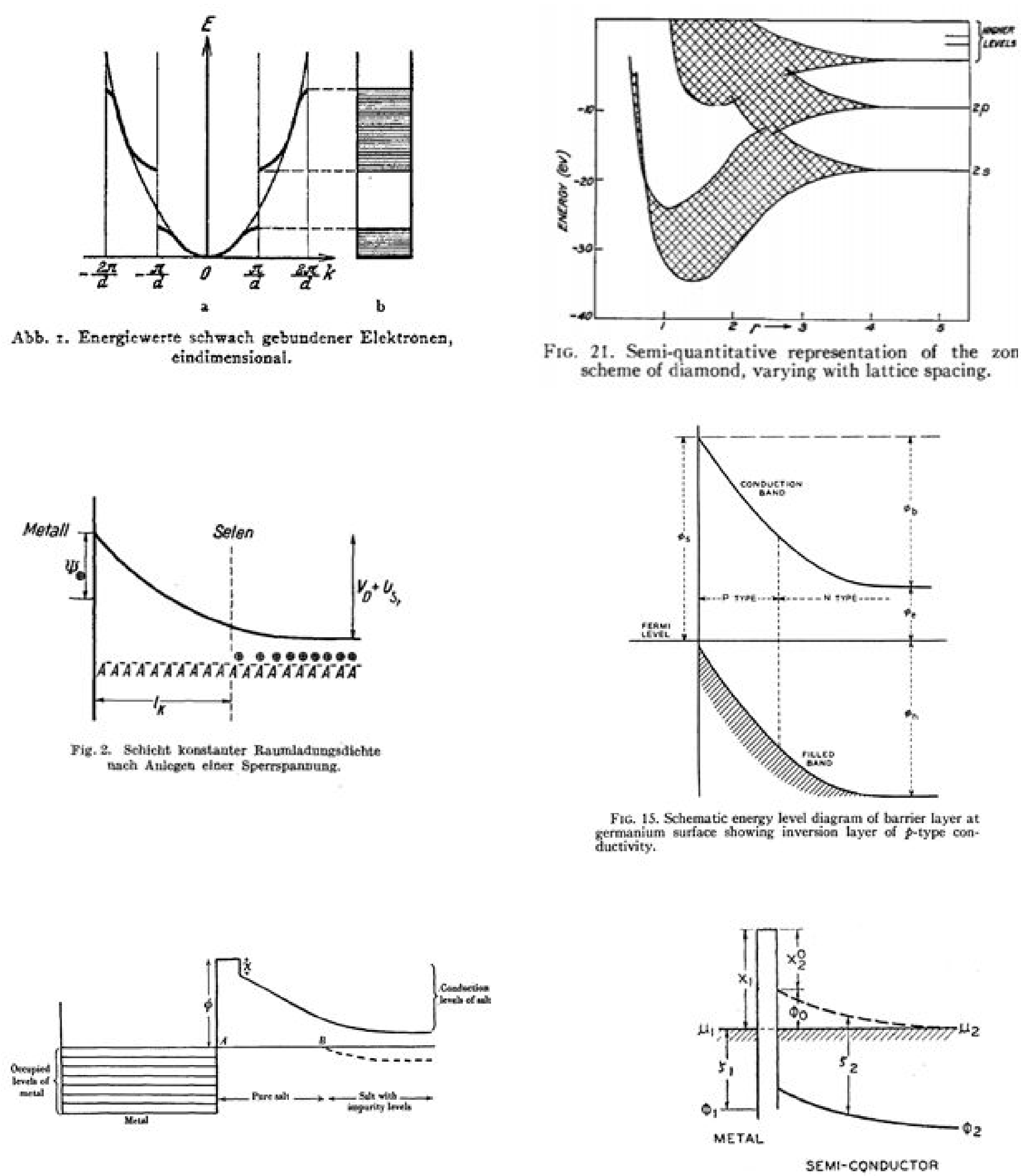


Figure 5: Theoretical models on semiconductor electronic band structure in bulk and at interfaces. TOP LEFT: Peierls' 1932 derivation of electronic bands in an infinite lattice of regularly arranged atoms. TOP RIGHT: Seitz and Johnson's 1937 semiquantitative estimates of electronic band structure of a diamond crystal as a function of lattice spacing (note the discrete energy levels at large distances which are equivalent to those of single atoms); also note Slater 1934; MIDDLE LEFT: Schottky's 1942 model for band bending at a metal-semiconductor interfaces (here with selenium as semiconductor). MIDDLE RIGHT: Illustration from Bardeen & Brattain's 1949 paper on transistor actions describing band bending at a semiconductor-semiconductor interface (note the similarities with Schottky 1942). BOTTOM LEFT: Mott's 1938 model of surface states on a semiconductor surface. BOTTOM RIGHT: Illustration from Bardeen's 1947 paper on surface states on a semiconductor surface (note the similarities with Mott 1938).

4. Materials and diagnostics

Progress in the category of semiconductor materials and diagnostics went hand in hand with the evolution of theory, the observation of anomalies, and the technological demands of new inventive concepts. Much early work in this category revolved around the characterization and classification of semiconductor materials. As theoretical insights deepened, materials researchers responded to such insights, for instance by striving for single-crystal production techniques, high-purity crystals and crystals with well-defined doping content (i.e. controlled addition of impurities). Bell Labs had its own metallurgists who absorbed and contributed to knowledge in this domain. Bell Labs teams were also part of the cross-institutional US radar effort during World War II, a major driver for semiconductor materials research. Through participation in this network, Bell Labs gained direct access to state-of-the-art germanium crystal production techniques as well as high-purity germanium samples – which would prove critical for solid-state amplifier advancement.

Early characterization of the electric properties of materials included resistance measurements under varying conditions such as different temperatures. Here, semiconductors differ from metals in that their conductance increases with temperature. This led to the classification of some materials as so-called “Heissleiter” (hot-conductors), a term adjusted around 1910 to “Halbleiter” (semi-conductors). Up to the late 1930s, some of the most widely used semiconductor materials included copper oxide and copper sulfide, iron sulfide, lead sulfide (known as galena), and silicon carbide (known as carborundum).

Karl Baedeker, a physics professor at Leipzig University, discovered around 1907 that the conduction of a compound semiconductor – he used copper iodide – could be changed by adjusting its iodide content (i.e. by changing its stoichiometry). This provided a first explanation for the large variation in the reported behavior of seemingly similar materials. Baedeker also noticed that copper iodide exhibited a positive Hall voltage[11] and cadmium oxide a negative Hall voltage, deducing that the charge carriers in the former material were positive and the charge carriers in the latter material were negative (which later came to be understood to be due to electrons carrying charge in N-type materials and holes in P-type materials). For this reason, Hall voltage measurements became an important diagnostic for characterizing semiconductors.

Originally, only compound materials were identified as semiconductors. In 1906, Radio engineer Pickard also proposed the use of pure silicon. However, debates persisted into the 1930s as to whether elemental materials such as pure silicon were indeed semiconductors, or whether observed effects were due to the inadvertent formation of compounds. Materials like elementary silicon were particularly difficult to experiment with and their behavior was considered notoriously irreproducible. Only the later emergence of a sufficient theoretical picture (conduction theory and band structures; see section C3) helped to make sense of the erratic behavior: in materials like silicon, charge density – and thus conduction behavior – is highly sensitive to even slightest impurities. Subsequent to this insight, efforts could be motivated toward achieving greater purification of materials like silicon. This, in turn, led to greater consistency of experiments. For similar reasons, germanium was not even recognized as a semiconductor until the 1930s but became popular once sufficient purification became feasible.

Physicists also came to understand that polycrystallinity was another major source of irregularity in semiconductor experiments as well as in the applied use of crystal detectors. This resulted in efforts to

[11] A transverse voltage across a sample in a magnetic field with an applied current. The Hall voltage is characteristic of the charge carrier population.

grow homogeneous, large single-crystals. In 1916, Polish Chemist Jan Czochralski, working at the AEG industrial research laboratory, developed a crystal growth method which formed the basis for much subsequent semiconductor research and manufacturing. The so-called Czochralski method involves pulling a "crystal seed" slowly from molten base material (e.g. molten silicon). The material then grows as a single-crystal through gradual solidification. Adding small amounts of other materials to the molten base allowed for the controlled addition of impurities (later known as doping).

During the late 1930s, semiconductor crystal detectors came to be used as rectifying diodes in radar receivers, as vacuum tube diodes could not keep up with the higher frequencies required. The resulting demand provided a major impetus for further semiconductor research. Military organizations on both sides of the Atlantic required much higher levels of reliability than the early cat-whisker crystal detectors had exhibited (which still required the manual search for a suitable spot on the crystal surface before each use). By now, it had become widely understood that greater consistency would result from (a) using elemental semiconductors instead of compounds; (b) achieving high levels of purity; and (c) using single-crystals instead of polycrystalline material. As a result, research focused on silicon and germanium, purification methods, as well as crystal growth methods. Purification of silicon and germanium could be achieved by repeated melting and recrystallization. Impurities would then aggregate in dedicated parts of the ingot which could be cut off. Silicon was initially preferred over germanium since it was more readily available. However, the lower melting point of germanium meant that it was easier to work with and once germanium procurement could be increased, focus switched to germanium both in the US and in Germany. Wartime research also led to a better understanding of the role of deliberate impurities (dopants) in these elemental semiconductors. Both US and German researchers identified suitable dopant materials such as boron (for P-type doping) and phosphorus (for N-type doping) and empirically characterized the effect of impurity content on charge carrier concentration (see Fig. 6 RIGHT).

In the US, Purdue University and the University of Pennsylvania (UPenn) emerged as major academic centers for semiconductor materials research with DuPont, Sylvania, GE and Bell Labs as industry partners. Leading personnel included Karl Lark-Horovitz, physics professor and department head at Purdue, and Frederick Seitz, physics professor at UPenn (like Bardeen, also a former student of Wigner). In Germany, the centers for wartime semiconductor research were the Telefunken laboratories near Breslau and the German radar research institute near Munich with Siemens being the main industry partner and with ties to the University of Munich. Leading personnel included Herbert Matare who had joined Telefunken right after his PhD and Heinrich Welker (a former PhD student of Heisenberg's and Sommerfeld's) who was poised to take over Sommerfeld's chair at the University of Munich until political turmoil forced him into industrial research.

When considering the 1940-1945 period, much technical progress in the quality of semiconductor production was achieved. For example: before the war, "commercial grade" silicon typically exhibited 99.8% purity (Riordan 1998); however, by 1945, DuPont was able to produce batches of silicon with 99.999% purity. Progress on purification techniques also affected germanium production. Ringer and Welker (1948) report germanium available to them with purity on the order of 99.99%. After the war, germanium had become the semiconductor material of choice for a number of reasons, including its low melting point, its advantageous electrical properties (high back-voltage), and the observations of several anomalies (see section C2). Researchers on both sides of the Atlantic were now eager to explore the newly available materials and their behavior. Handel (1999) describes the post-war knowledge levels between US and German semiconductor groups as comparable. Lark-Horovitz at Purdue continued to work on germanium crystals and collaborated closely with Bell Labs which he supplied with germanium samples. In Europe, both Matare and Welker joined a Westinghouse subsidiary near Paris where they were tasked with developing a semiconductor research program. Both groups initially focused on producing semiconductor diodes (essentially high-quality crystal detectors) from doped high-purity germanium crystals.

It was subsequent work with this type of material that led both the team in the US and the team in Europe to the observation of controlled solid-state amplification i.e. the so-called point-contact transistor (in December 1947 in New Jersey and in January 1948 in Paris – see Handel 1999; Van Dormael 2004; Riordan 2005). In each case, two metal wires ("point contacts") were placed close to one another (less than 100 microns) on the surface of a high-purity germanium crystal, resembling early crystal detectors in a two cat-whisker configuration.

Overall, Bell Labs had much exposure to and, in some cases, direct involvement in the process of semiconductor materials, diagnostics, and tools development. Within its ranks, Bell Labs employed a number of metallurgists such as Foster Nix, Dean Woolridge, Jack Scaff, and Henry Theurer. Bell Labs metallurgists were themselves involved in developing methods for silicon and germanium purification and crystal growth (building on earlier work such as Czochralski's). This internal capability for metallurgy also allowed Bell Labs to take full advantage of its participation in the wartime radar research network. Bell Labs metallurgists collaborated particularly closely with researchers at Purdue and UPenn which involved frequent visits and the exchange of technical reports and samples (Nix 1975). These external linkages were critical for Bell Labs' later research on solid-state amplifiers: precisely controlled materials would turn out to be key for achieving high efficiency, reliability, and reproducibility of the sought-after effects.

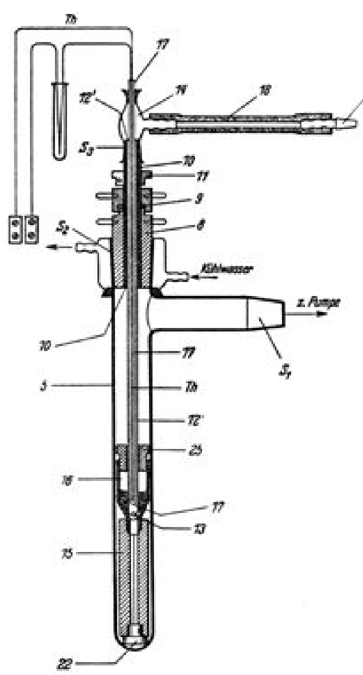


Abb. 1. Schmelz- und Gießapparatur.

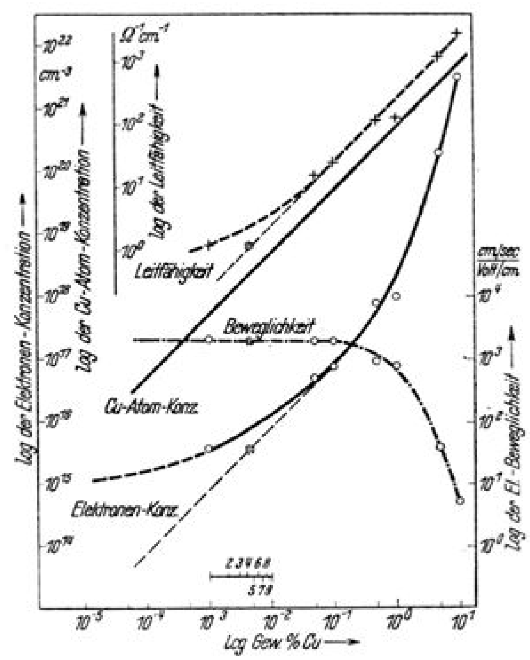


Abb. 7. Elektronenkonzentration und Elektronenbeweglichkeit als Funktion des Gew.-%-Gehaltes an Kupfer.

Figure 6: LEFT: Apparatus for creating high-purity doped germanium crystals, as described in Ringer and Welker 1948. RIGHT:

Experimental data presented in Ringer and Welker 1948 that shows electron concentration and electron mobility as a function of copper alloying percentage in germanium (i.e. copper doping).

5. Economic demand

The previous four sections discussed diverse technical categories that each progressed to finally enable the possibility of consolidation and maturation of solid-state amplifier technology. Whereas this category is economic in nature, it also represents a partial condition that impacted the feasibility of the technology's advancement: that is the anticipation of substantial economic demand for solid-state amplifiers. The anticipation of demand was needed to justify the cost of sustained research and development. This is a condition that was not met until after World War II, when the US experienced both an overall economic upswing and rapidly growing demand for amplifiers and switches. Bell Labs executives recognized these dynamics early on, not least because Bell Labs' own business represented a major source of demand for solid-state electronics.

The discussion of early solid-state amplifier patents in section C1 forces the question why respective patent holders such as Lilienfeld did not manage to turn their inventions into economic benefit. After all, filing and maintaining a patent incurs costs and is ideally only pursued with a reasonable expectation of return. Inventors like Lilienfeld may indeed have held such expectations at the time of patent filings. However, clearly, they did not materialize. Considering the discussions in Crawford 1991 and Arns 1998, it seems likely that Lilienfeld built some prototypes of his proposed devices and that they worked at least some of the time and to some extent. Still, when considering the high requirements for materials and the importance of theory to guide precise design decisions, it seems equally likely that a lot more research was needed to turn such prototypes into reliable and consistent devices suitable for industrial use. The extent of effort, associated cost, and the difficulty of obtaining investment might have well been underestimated by Lilienfeld. Arns (1998) further argues that Lilienfeld likely failed to obtain support from an industrial sponsor for two reasons: 1) the Great Depression hit the world economy hard during the late 1920s and created a pessimistic economic climate through much of the 1930s; 2) the market for solid-state amplifiers and switches was limited and well served by the incumbent vacuum tube diodes and triodes. Under these circumstances, it seems unlikely that an industrial sponsor would have agreed to take on a risky multi-year research and development program with highly uncertain outcomes. Elaborating on the second point, Arns (1998) points out that the number and specifications of vacuum tubes were even seen as a sign of quality for certain consumer products such as radio sets (see Fig. 7 LEFT). At the same time, industrial demand for amplifiers and switches was still limited.

The development of the vacuum tube diode is typically attributed to John Fleming in 1904 i.e. around the same time as solid-state diodes emerged which became known as crystal detectors. For stationary radio systems, vacuum tube diodes quickly became more popular than crystal detectors since they did not require adjustment ("searching for a hot spot") before each use. As a result, the crystal detector ended up being mostly used by specialists and radio enthusiasts through much of the 1920s and 1930s.

However, crystal detectors – and therefore semiconductors more generally – saw an uptake in demand during the 1930s as their relevance for radar receivers became apparent. Radar systems evolved towards higher and higher frequencies and – in contrast to solid-state diodes – vacuum tube diodes had comparatively low cut-off frequencies. The allocation of funds (largely through military programs) toward improved crystal detectors based on better controlled semiconductor materials was a direct result of this change in demand.

Additionally, World War II catalyzed a massive uptake in the development and use of electronics more generally. The most prominent manifestations of this trend were the massive vacuum tube computers built in the US and UK (see Fig. 7 RIGHT). But advanced electronics also became more commonplace in less visible applications such as telephone networking. As the number of required and deployed amplifiers and switches grew rapidly, the short-comings of vacuum tubes became ever more obvious: high power consumption, long warm-up times, and limited operating lifetimes. Some vacuum tube computers required replacement of defective valves as frequently as every 15 minutes.

Bell Labs was ideally positioned to assess the expected demand for advanced electronics. First, its mother company AT&T represented a major source of demand for telephone network equipment. Equipment itself became more sophisticated over time and required more and better electronics. Additionally, subscriber numbers were expected to rise steeply after the war which boosted demand even more. Second, Bell Labs employees had been involved in a large number of advanced wartime activities. Many such activities involved optimization tasks which increasingly relied on computers. This gave Bell Labs staff first-hand insights into another major source of growing demand for advanced electronics.

That this rapidly growing demand for advanced electronics could over time hardly be fulfilled by vacuum tubes must have been especially clear to director of research Mervin Kelly. Kelly was intimately familiar with vacuum tubes, their potential and their constraints, as he led Bell Labs' vacuum tube department until 1936. In this capacity, Kelly managed to extend the lifetimes of typical vacuum tubes by orders of magnitude. Nevertheless, this line of work must have also demonstrated to Kelly the intrinsic limitations of vacuum tubes and the need for alternatives.

Figure 7: LEFT: 1932 Radio ad emphasizing the large number of vacuum tube amplifiers as a sign of quality. RIGHT: Early computers such as the UK's Harwell computer shown here contained thousands of valves that had to be regularly maintained and replaced (Source: The National Museum of Computing).

D. Phase II: Consolidation

The previous sections discussed technical and economic factors that I argue had to sufficiently advance before maturation and subsequent dissemination of solid-state amplifier technology became feasible. The corresponding "exploration phase" covered the period between roughly 1905 and 1945. This present section will focus on the subsequent period from 1945 to 1948 which I call, the "consolidation phase." This is the period that is most typically associated with the invention of the transistor.

This section is structured as follows:

I will first present a short summary of the commonly received narrative of transistor invention, a narrative typically centered on a small number of Bell Labs employees and events in late 1947/early 1948 that are presented as extraordinary in the way they transpired and reverberated. Following nomenclature introduced by Epstein (1926) for inventor-centered narratives, this perspective can be referred to as the "heroic" or "individual" narrative. I will then critique this narrative and question some of its assumptions.

This is followed by the proposition of an alternative narrative: in this second narrative, emphasis shifts from individual inventors to overall institutions and research programs. Again following nomenclature introduced by Epstein (1926), this narrative can be referred to as the "systematic" or the "institutional" narrative.

1. The individual narrative

Many descriptions of transistor invention begin in 1945 and narrowly focus on the three Bell Labs employees who were later awarded a Nobel Prize "for transistor action": William Shockley, John Bardeen and Walter Brattain. The corresponding narrative can be summarized as follows:

> In 1945, Bardeen was recruited into Bell Labs and joined a newly formed solid-state amplifier team under the leadership of Shockley. Shockley then "formulated a theory on something he called the 'field effect'" (Gertner 2012) which resulted in a series of experiments based on designs that resembled the future field effect transistor. However, these designs failed to produce the desired amplification effect. Shockley then tasked Bardeen with finding a theoretical explanation for the lack of amplification achieved, prompting him to contemplate the role of surface states. Bardeen's shift of attention is then often described as a transition from applied research to basic research – which is then often argued to have enabled the eventual transistor invention (see, for instance, Wolff 1983). Subsequent experimentation led the team to notice interferences between two closely spaced contacts on a high-purity germanium crystal in November 1947. A few smaller adjustments such as well-defined spacing between the two contacts led to a configuration which yielded reasonably reproducible effects. The above-described process is then viewed as the genesis of the point-contact transistor as the first transistor. An additional contribution is typically associated with Shockley's interpretation of the underlying mechanism: Shockley proposed that the effect is caused by a flow of minority carriers through the germanium crystal i.e. the flow of holes through a crystal region with excess electrons (N-type region). The minority carrier interpretation then led Shockley in January 1948 to the conception of a so-called junction transistor (a transistor comprised of three alternating N-type and P-type layers). Therefore, within two "magic months" (Shockley in Essers & Rabinow

> 1974) – December 1947 and January 1948 – the trio is said to have conceived and explained the solid-state amplifier. Their individual genius and their unwavering dedication to “basic science” are often presented as central, if not causal, to their success (Hoddeson & Daitch 2002; Gertner 2012).

As a standalone origin story, this perspective appears consistent and plausible, somewhat poetic even. However, a closer look raises questions as to the selectivity and interpretation of this narrative:

First of all, the mentioned trio was part of a much larger group at Bell Labs that provided motivation and input for the undertaken research. This included the semiconductor group’s co-leader Stanley Morgan and other group members such as Gerald Pearson, Robert Gibney, and Hilbert Moore. However, as the sections above indicated, contributions from researchers beyond the semiconductor group were critical. This includes engineers such as Russel Ohl who demonstrated early solid-state amplifiers and physicists such as Purdue’s Lark Horovitz who provided the crucial high-purity germanium materials. While Bardeen and Brattain advocated for including other contributors in promotional photographs, Shockley insisted that only the three of them would be on photographs associated with transistor invention.

Some authors such as Gertner 2012 imply that the field effect concept that motivated early experiments of the semiconductor group was conceived by Shockley himself. However, the 1948 Johnson affidavit (see Arns 1998) suggests that the field effect experiments were simply replication attempts of the earlier Lilienfeld patents.

The emphasis on basic science as having been key to the future invention of the point-contact transistor is also misleading. This notion that the point-contact transistor emerged from “basic science” or “pure science” has its origins in Bell Labs' own interpretation of events. During the press conference that presented the point-contact transistor to the public, Bell Labs manager Ralph Bown presented it as a reflection of “the value of basic research.” However, in actuality, Bardeen’s surface state theory and the associated inversion layers had little to do with the point-contact transistor design and its underlying mechanism. Even Shockley himself later admitted that “inversion layers were probably not important for the point-contact transistors [..] nor for the junction transistors.” (Shockley in Essers & Rabinow 1974). In other words, Bardeen’s work on surface state theory merely coincided with Brattain’s various permutations of common experimental configurations; but it did not drive the experimental work toward the design of the point-contact transistor in any deliberate way. For narrative effect, those two aspects were later often causally linked, adding an element of control and volition to the otherwise fairly serendipitous experimental process.

Moreover, Bardeen’s surface state theory was not as groundbreaking as sometimes presented. Already in 1932, the future Nobel prize winner Igor Tamm introduced the notion of surface states which was then further developed by Boris Davydov (1938, 1939). Bardeen effectively expanded on Tamm’s and Davydov’s work (which he was aware of, as is clear from his own citations of Davydov e.g. in Bardeen 1947). Bardeen’s work on surface states certainly represented an important contribution but an incremental one rather than a fundamentally novel one.

The inventive process that led to the point-contact transistor prototype is sometimes presented as deliberate and purposive. In actuality, it was fairly serendipitous and unplanned. For instance, the experimental campaign in 1947 included experiments where a carefully prepared oxide layer got accidentally washed off which led to new incremental insights about semiconductor surfaces. When Bardeen and Brattain finally considered the successful configuration with two closely spaced contact wires, they had actually predicted the opposite behavior from what was observed. Even once the point-contact transistor functioned reasonably reliably, the interpretation of the underlying mechanism was still flawed and had to be repeatedly adjusted due to earlier misconceptions. For instance, only a follow-on experiment by a younger colleague, John Shive, led to the insight that minority carriers played a central role in the observed effect i.e. that holes injected through one of the contact wires were able to travel through the N-type bulk of the crystal, despite the presence of excess electrons[12] (see Fig 8 BOTTOM). In 1962, Nelson commented that even then "there still is no really adequate quantitative theory explaining the working of the point contact transistor."

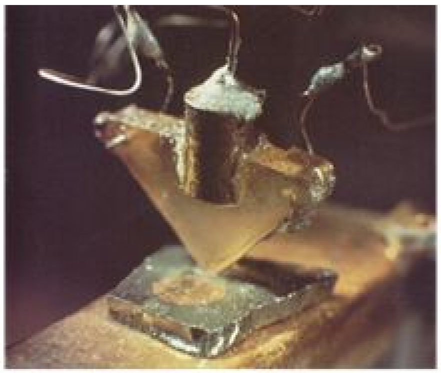
When considering the experimental configuration itself, the main difference from earlier configurations – such as the two-cat-whisker crystal detectors – was the high-purity single-crystal germanium slab that was provided by the Purdue collaborators. The new material enabled comparatively long lifetimes of injected charge carriers and consistency across the crystal surface[13]. Bassett (2002) confirms: "the point-contact transistor represented an experimental discovery, utterly dependent on the current state of materials and experimental techniques." The trio knew that with the proper material available, the same kind of discovery would be inevitable in other groups – it was going to be only a matter of time until another scientist would try out the two-cat-whisker configuration. This acute concern is clearly reflected in the increased nervousness that permeated group members during the lockdown period from January and July 1948. During this time, results of the point-contact transistor prototypes were to be kept secret such as to allow for patent filings and related formalities. During the first part of 1948, Bell Labs scientists were particularly concerned that groups at Purdue or at Westinghouse in Paris might make similar announcements ahead of them, stealing the spotlight. For instance, at the APS March conference in 1948, a researcher from the Purdue team suggested to Brattain to try an experiment with two closely spaced contact wires – i.e. a two cat-whisker configuration –- inspired by their own observations of resistance anomalies on the surface of their high-quality germanium crystals. Having already done that exact experiment (which was in fact the point-contact transistor experiment) and knowing the important result, Brattain simply said "Yes, this might be a good experiment", turned around, and walked away. In terms of experimental work, the Purdue group has been described as "6 months behind" the Bell Labs group. In summer 1948, a Bell Labs delegation learned about the ongoing Westinghouse work in Paris and subsequently accelerated their publication timeline. The Westinghouse team had independently found the same configuration (see Fig. 8 RIGHT) in January 1948 and was now getting ready to announce their results – which took place in August 1948, one month after the Bell Labs announcements in July.

Similarly, Shockley's junction transistor design was considered to be obvious once the feasibility of minority carriers was verified – which was the case after the experiments by John Shive. The obviousness of the design is illustrated by the fact that Shockley felt the need to quickly and spontaneously disclose his junction transistor ideas to colleagues right after Shive had first presented his experimental result in support of minority carriers. Shockley later described that he improvised this presentation out of concern that his colleagues would have recognized the junction transistor design possibility within hours of Shive's presentation – which could have deprived him of taking credit for the idea.

---

[12] Which were originally expected to neutralize the holes in short order.

[13] Even so, sometimes the contact wires still had to be slightly wiggled whence the amplifier stopped working.



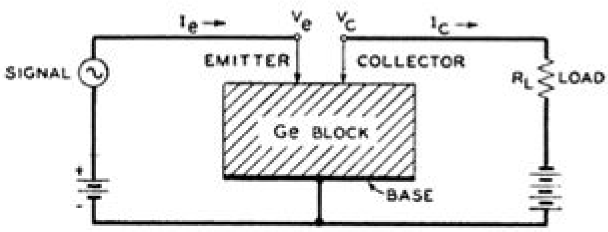
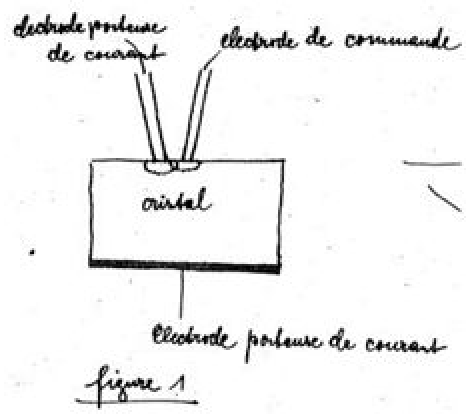


FIG. 1. Schematic of semi-conductor triode.



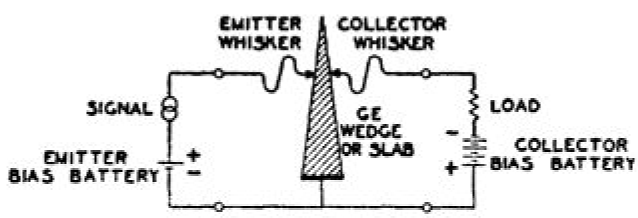


FIG. 1. Electrode geometry and circuit connections for a double-surface transistor.

Figure 8: Early point-contact transistor prototypes and designs. LEFT: US-based Bell Labs design of the point-contact transistor as devised in December 1947 by Bardeen and Brattain (Bardeen & Brattain 1948). RIGHT: France-based Westinghouse design of the point-contact transistor as devised in January 1948 by Matare and Welker (Van Dormael 2004). BOTTOM: Configuration of Shive's double-surface experiment which confirmed the role of minority carriers in observed conduction (Shive 1949).

Several authors recently argued that the emphasis on the creation of the point-contact transistor prototype and the “magic months” is exaggerated and disproportionate to other contributions (citation). The central point of such arguments is not to downplay the talents and contributions of Shockley, Bardeen and Brattain – but rather to come closer, in one’s descriptions, to the reality of scientific research. As pointed out by Allchin 2003, research is often complex and confuse, fraught with uncertainty and conflicting results. This is contrast to what he calls “idealized classroom-histories” where one step leads to another and where the overall narrative is often crafted from the perspective of the final result to which earlier steps inevitably led. At the same time, the messiness and relative unpredictability of research does not mean that no steps can be taken to increase the odds of success. While the Bell Labs team still relied on serendipity, their knowledge allowed the researchers to create conditions where productive serendipity could occur – akin to what some philosophers have called “engineered serendipity.” Moreover, the researchers possessed the kinds of knowledge that were required for recognizing novel behavior once encountered, to interpret it, form hypotheses, and derive conditions for future iterations of experiments. None of this is trivial – rather, it requires very specific types of knowledge. The question is whether such knowledge is characteristic of individuals – perhaps restricted to individual genius even – or characteristic of institutions. In other words, had Brattain or Shockley left the Bell Labs semiconductor group early, would the discovery of the point-contact transistor still have occurred? Brattain himself indirectly puts forth an answer to this question in the form of a comment about his colleagues: “one could not have accomplished the work he had done without them, and that it was really only a stroke of luck that it was he and not one of them” (Murty 2008). Even more generally speaking, the scientists considered here were embedded in a research program with a specific orientation, size, composition, and timeline; embedded in a research organization with a specific mission, culture, and set of priorities; and ultimately embedded in a global community of scientists with certain shared data sets, academic literatures, and educational backgrounds. The following section will now propose an institutional narrative as an alternative to the individual narrative presented and critiqued here.

2. The institutional narrative

An alternative narrative focuses on the programmatic and organizational aspects of the process of transistor invention. In fact, it was Bell Labs director of research Melvin Kelly himself who pushed back against the “magical” narrative and instead encouraged an institutional perspective: “There is nothing magical about science. Our research people are following a straight plan as a part of a system and there is no magic about it.” A corresponding narrative could be developed as follows:

> While vacuum tubes played a central role in the Bell Labs of the 1920s, researchers and management began to realize their limitations. Around the same time, reports began to surface in popular radio magazines of unexpected amplification effects in semiconductor materials. However, such effects were erratic and not explainable. Patents began to surface that proposed concepts of solid-state amplification, however, their specifications lacked precision and the absence of commercial products confirmed their immaturity. Simultaneously, solid-state physics made great strides and semiconductor materials became better understood. Bell Labs instituted a first solid-state group in 1936: an interdisciplinary composition of researchers, ranging from metallurgists and chemists to electrical engineers and physicists. A critical activity comprised the

absorption and assessment of existing semiconductor related knowledge of all kinds: Ohl monitored the radio equipment community; Darrow and Kelly maintained close contact to European academics; new recruits such as Shockley and Fisk drew on their networks in US academia; metallurgists such as Woolridge and Theurer adopted and improved techniques for handling semiconductor materials, thereby also developing the capacity to quickly absorb new techniques from elsewhere. By the late 1930s, executives such as Kelly assessed that several relevant categories had evolved to such an extent that a push for integration could be attempted: Kelly explicitly encouraged members of the solid-state group to consider developing reliable and scalable devices that draw on solid-state amplification – effectively defining the mission orientation for an applied industrial research program. However, this process got interrupted in short order with the start of World War II. Wartime research still led to further progress in semiconductor materials, especially with respect to the precise control of material properties. This kind of progress set up ideal conditions for the post-war resumption of the solid-state research program. The now available high-purity single-crystal germanium and silicon samples allowed for a better match between models and experiment; and for more systematic empirical research. Moreover, the need for finding a vacuum tube replacement became ever more urgent with skyrocketing demand for amplifying and switching electronics. Allocating resources with the goal of developing reliable solid-state amplifiers became all the more justifiable. Kelly recruited more PhDs to the solid-state research group and intensified interactions between theorists, materials experts, and engineers familiar with reported anomalies. The team caught up on recent literature and continued to take advantage of the former radar research network. With materials, theoretical models, and design concepts aligning, a reliable solid-state amplifier configuration was encountered in short order. The corresponding prototypes were in many ways still flawed and unsuited for industrial use. However, they highlighted the potential for continued progress and represented a major step of de-risking research in this area. As a result, IP was secured, a PR and marketing strategy prepared, and more resources allocated, resulting in substantial growth of the solid-state research team. A sustained and steadily scaled research effort over the following 12 years (with changing team members but institutional continuity) eventually led to the emergence of the transistor's dominant design in 1959: the MOS-FET. Almost half a century after first reports of solid-state amplification, the MOS-FET finally represented a solid-state amplifier that was reliable, well understood, easily manufacturable – and therefore highly scalable.

Instead of a small number of discrete events (for instance during the "magic months"), this alternative narrative focuses on the overall process in which such events were embedded. In as much as this narrative involves individuals, it emphasizes their relationship to the process (including, for instance, the creation of boundary conditions).

In this regard, one person stands out in particular: Bell Labs executive Mervin Kelly. Kelly was Bell Labs' Director of Research from 1936-1944 and Executive Vice President and President from 1944-59. Before 1936, Kelly was the head of Bell Labs' vacuum tube department. In other words, Kelly held leadership positions at Bell Labs from the 1920s through the late 1950s. The tenures of most of the scientists he oversaw were much shorter: Shockley and Bardeen were recruited by Kelly in 1936 and 1945 respectively and both left during the early 1950s. Kelly's time at Bell Labs happens to coincide largely with the overall emergence of solid-state amplifiers, as discussed in the sections above: starting with first reports of anomalies in the 1910s and 1920s, and concluding with the MOS-FET in 1959. Having witnessed the development of this technology over the long term may have aided in assessing progress, potential, as well as remaining limitations and risks. Kelly showed an interest in solid-state electronics as early as the 1920s, created the first solid-state program in 1936, and reconstituted it in 1945. He initiated the scale-up of the program in 1948 and oversaw the 1950s developments that led to the MOS-FET. According to his colleague John Pearce, Kelly “understood that a complicated technological process lies between discovery and use.” Throughout this time Kelly oversaw the recruitment of specialists, the formation of teams, the articulation of objectives, and the allocation of resources. Wolff credited Kelly with “asking the right questions and finding the best people to answer them.” Kelly himself explained his management philosophy during a conference presentation: "The first, and perhaps the most important, factor is the program itself. What shall it contain? What can be discarded at once, and what shall be eliminated after limited exploration? How can comprehensive coverage with freedom from gaps be assured?”

The Bell Labs solid-state research effort showed clear signs of mission orientation. Kelly was unambiguous toward Shockley in what he wanted him to achieve: “Kelly arranged for me to have an initial indoctrination experience with high frequency vacuum tubes. This included spending some months of 1937 in the Vacuum Tube Department. During that time Kelly gave me an eloquent pep talk—one that had a long-lasting influence on my own motivations. He pointed out that relays in telephone exchanges caused problems and were expensive to maintain. He felt that electronics should contribute to telephone exchanges in addition to making long distance transmission possible.” (Shockley in Essers & Rabinow 1974). Rather than constituting “basic research” without concrete applications in mind, the program appears to have been quintessentially “applied industrial research” – as Kelly himself called it (Kelly 1950). This was consistent with general observations on Bell Labs research such as Nelson who observed (1962): “research scientists are encouraged to be device conscious.” The notion that key breakthroughs resulted from individual researchers liberally pursuing their own interests seems to be contradicted by actual practices (Nelson 1962): “If the work does not prove of interest to the Laboratories, eventually the individual in question will be requested to return to the fold, or leave. It is hoped that pressure can be informal. […] even top-flight people [..] have been asked to change their line of research.” Whereas in some communities, “basic research” is given a normative preference over “applied research”, the very mission orientation may have actually been a major advantage of the Bell Labs research program, as will be discussed in more detail below.

The above paragraphs support the notion that the quest for a solid-state amplifier was a long-term, applied research program driven by a clear practical objective. That being said, the practical implementation of such a program involves a number of critical aspects which will be discussed next. Here, I distinguish between three types of institutional capabilities, each of which I argue was present in the Bell Labs case: 1) Sensing; 2) Processing; and 3) Acting.

Sensing

The term sensing here denotes an institution’s ability to identify and field relevant ideas from a variety of external sources. This includes the subsequent ingestion of such knowledge into the organization. Metzler 2019 describes this practice as an organization’s “situational awareness.”

In the case of Bell Labs, this capability manifested in a number of ways. A central aspect was the prevalent recruiting practice. Bell Labs recruitment was broad, cutting across academic fields as well as communities of practices. New hires ranged from freshly minted PhDs to technicians with decades of bench experience. As a result, Bell Labs personnel included individuals with a wide range of external linkages in the form of interests, prior exposure, and relationships. As mentioned in section C2, Russell Ohl, for instance, retained close ties to the radio engineering community. He described his own role at Bell Labs as follows: "I studied and reported what the situation was – the radio equipment situation [..] I kept the company knowledgeable with regard to the art." Similarly, Karl Darrows, trained as a PhD physicist and a speaker of multiple languages, maintained personal relationships with a large number of leading academic physicists in Europe and the United States. European-trained researchers fulfilled similar tasks such as Foster Nix who received his graduate training in Berlin and was bilingual in English and German (Nix 1975). Relevant scientific news that Bell Labs employees such as Darrows and colleagues might have missed, were often brought to their attention by collaborators in Bell Labs' extended network. An example is Karl Lark-Horovitz, physics department head at Purdue and an Austrian immigrant, who regularly monitored German-language publications and drew attention to relevant new developments in interactions with Bell Labs collaborators (Van Dormael 2004). Still other Bell Labs scientists such as metallurgists Theurer and Scaff liaised with industry partners Dupont and Sylvania which included site visits and the exchange of technical reports as well as informal knowledge related to materials processing techniques. With respect to materials, the partnership with Purdue was also critical as it meant access to their high-purity, single-crystal germanium samples as well as related experimental results and production processes. Relationships to universities were close beyond those specific research networks – this included Columbia University and other NYC-based universities as well as the various alma maters of Bell Labs' scientists. Nelson (1962) described these linkages as follows: "Because a scientist at the Laboratories is not forced to abandon the traditions of the scientific community, Bell scientists tend to maintain very strong links with the academic world. Many Bell scientists have taught at universities, and many Bell scientists are actively sought by university faculties." The academic culture also manifested in the form of regular seminars and study sessions where national and international speakers were invited, literature reviews were conducted, or recent academic papers discussed. Bell Labs scientists also participated in training programs such as Brattain's attendance of the 1931 summer school with Sommerfeld in Michigan.

Effectively, Bell Labs combined advantages of universities and firms: It benefited from the openness of knowledge flows across organizational boundaries as well as from the participation in research networks. At the same time, Bell Labs was not constrained to certain types of formal knowledge only (as many universities were at the time, and to some extent still today, due to certain cultural or habitual constraints). Rather, as a firm, Bell Labs could practice what from a university perspective might have been perceived as a type of unorthodoxy: that is, pragmatically engaging with diverse knowledge sources, including informal, anecdotal, or purely "practical" ones (types of knowledge that have been described as "clinical knowledge" as opposed to "formal knowledge"; see Lester & Piore 2009).

Processing

Besides fielding ideas from a variety of sources, Bell Labs as an institution also exhibited the ability to digest and process corresponding insights. This included the ability to recognize opportunities and assess remaining knowledge gaps.

Overall, this process required the habitual exchange of ideas between a diverse group of staff that could contribute different perspectives on a topic or identify relevant cross-connections. At Bell Labs this was indeed the case, as it was common to have interdisciplinary teams and close collaboration between academics and practitioners. Bell Labs employees were encouraged to work with office doors open and to reach out to colleagues across the organization in case of questions. By company policy, it was not permitted to turn away a colleague seeking advice, regardless of seniority. This desire to stimulate interaction and exchange was also reflected in the architecture of the Murray Hill campus which contained a large corridor that spanned the entire campus. This main artery thus facilitated frequent serendipitous encounters.

Moreover, managers and executives often had multi-disciplinary backgrounds, comprising training in science, engineering, and business domains. Kelly, himself holder of a PhD in physics, stated: “Leaders or managers must be technologically trained and technologically competent. Only thus can decisions be based on insight and understanding rather than on salesmanship and hearsay.” (quoted in Pierce 1975). The process of identifying cross-connections between different types of knowledge and assessing corresponding opportunities as well as remaining risks is maybe best exemplified by Kelly’s efforts to expose theoretical physicists such as Shockley to phenomenological knowledge from practitioners such as Ohl. Shockley recalled Ohl’s demonstration of crude negative resistance solid-state amplifiers as follows: “Kelly had arranged a demonstration of a radio set lacking vacuum tubes. In this radio the amplification was accomplished by point contact detectors. These semiconductor devices acted as negative resistances, These devices indicated that semiconductors held exciting potentials but they themselves had many shortcomings.” (Shockey in Essers & Rabinow 1974).

Again, a key advantage over traditional university research was the ability to engage academics as well as practitioners to collaborate formally and informally across disciplines, knowledge domains, practices and standards.

Acting

The identification of opportunities at the intersection of different knowledge fields needed to be followed by decisions over the allocation of resources and corresponding long-term strategy making. This involved questions over what size and composition of teams was appropriate for a given topical area at a given point in time.

When Kelly became director of research in 1936, he formed a small solid-state physics team as one of his first major actions. This suggests that already during his tenure as the head of the vacuum tube division, he considered the potential of solid-state electronics. Kelly’s recruiting at the time focused on physicists (such as Shockley) and metallurgists (such as Fisk, Woolridge, and Townes). The solid-state team had to break up during World War II but was reconstituted in 1945. The initial budget then allowed for a team of 20-30 scientists, with about half of them focusing on semiconductors (Nelson 1962). By now, Kelly was very explicit about the perceived opportunity and timeliness of research in this area: "Employing the new theoretical methods of solid-state quantum physics and the corresponding advances in experimental techniques, a unified approach to all of our solid-state problems offers great promise. Hence, all of the research activity in the area of solids is now being consolidated." The sense of urgency can be explained by the circumstances at the time: Already by the early 1940s, the Bell Labs solid-state physics team had accumulated much knowledge about semiconductors. This included study of the academic literature as well as patents, a series of inhouse experiments, inhouse experience in the processing of semiconductor materials, participation in the research network on silicon and germanium purification and crystallization. By 1945, the time seemed ripe for integrating such branches of knowledge and making a push toward device development.

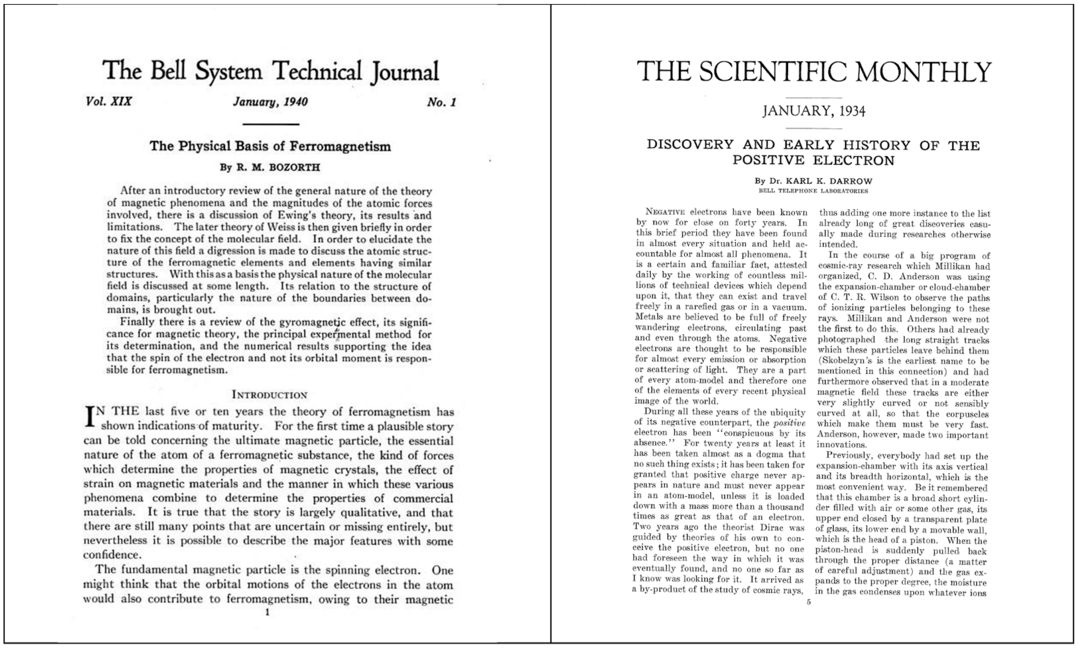
The fact that the point-contact transistor prototype was encountered less than two years into the new program can be explained from the perspective of these aligning circumstances (see section C). Similarly, such circumstances imply that others, too, were coming closer to integrating newly available knowledge towards industrial application. That was indeed the case, as exemplified by the almost simultaneous recognition of the point-contact design in combination with high-purity germanium by the team at Westinghouse in Paris. In many ways analog to Bell Labs, their scientists – especially Matare and Welker – benefitted from their close interactions with leading academics (such as Sommerfeld, Schottky, Heisenberg), state-of-the-art materials (in the context of the German radar research program), awareness of previously proposed concepts (such as the designs of Lilienfeld, Heil, and Hilsch & Pohl), as well as the observation of anomalies (such as Matare's wartime recognition of interference between two closely spaced contact wires on a high-purity germanium crystal).

A critical aspect – and a major difference to the French counterpart – is how Bell Labs responded to the first point-contact transistor prototypes. Upon learning about the new results – which de facto represented a de-risking of applied semiconductor research – Kelly reacted with a major scale-up: "we immediately formed a closely associated fundamental development group to acquire that body of technological knowledge essential to the development and design of transistors for the many specific communications applications that would certainly follow." In Paris, no such response took place. When Matare and Welker first encountered the point-contact transistor configuration, they immediately informed their superiors. However, it took several weeks until a lab visit by executives could be arranged. Even then, the significance of semiconductor research and the potential of solid-state amplifiers was not recognized. In the absence of an expansion of resources, the pioneering French semiconductor program collapsed in short order.

Finally, not to be underestimated is the role of Bell Labs' PR, legal, and marketing activities. The scale-up of research in 1948 was accompanied by a carefully planned campaign comprising synchronized patent filings, press conferences, and academic publications. This was later followed by workshops and continued media interactions. A central piece in this process was the decision to give the solid-state amplifier a dedicated name – transistor – chosen by Bell Labs management and used to clearly demarcate Bell Labs efforts from previous contributions. Bell Labs could now claim to have "invented the transistor" instead of limiting itself to having made an important contribution to the long evolution of solid-state amplifiers.

This section presented an "institutional" narrative of transistor invention – which might in fact be rather called transistor emergence or transistor evolution. The presented narrative emphasized the role of institutional capabilities especially with respect to the management of knowledge and resources. It also emphasized the relationships between individuals, organizations, and large-scale trends and developments. The narrative implies that transistor invention ought to be viewed as a decade-long process rather than a short series of events such as during two "magic months."

So far, most discussions focused on the period between roughly 1900 and 1948. However, transistor development did not stop there. In fact, it was not until 1959 that a transistor design suitable for large scale dissemination emerged with the MOS-FET design. The following section will discuss this final part of transistor evolution in more detail.



The Bell System Technical Journal

Vol. XIX *January, 1940* *No. 1*

The Physical Basis of Ferromagnetism

By R. M. BOZORTH

After an introductory review of the general nature of the theory of magnetic phenomena and the magnitudes of the atomic forces involved, there is a discussion of Ewing's theory, its results and limitations. The later theory of Weiss is then given briefly in order to fix the concept of the molecular field. In order to elucidate the nature of this field a digression is made to discuss the atomic structure of the ferromagnetic elements and elements having similar structures. With this as a basis the physical nature of the molecular field is discussed at some length. Its relation to the structure of domains, particularly the nature of the boundaries between domains, is brought out.

Finally there is a review of the gyromagnetic effect, its significance for magnetic theory, the principal experimental method for its determination, and the numerical results supporting the idea that the spin of the electron and not its orbital moment is responsible for ferromagnetism.

INTRODUCTION

IN THE last five or ten years the theory of ferromagnetism has shown indications of maturity. For the first time a plausible story can be told concerning the ultimate magnetic particle, the essential nature of the atom of a ferromagnetic substance, the kind of forces which determine the properties of magnetic crystals, the effect of strain on magnetic materials and the manner in which these various phenomena combine to determine the properties of commercial materials. It is true that the story is largely qualitative, and that there are still many points that are uncertain or missing entirely, but nevertheless it is possible to describe the major features with some confidence.

The fundamental magnetic particle is the spinning electron. One might think that the orbital motions of the electrons in the atom would also contribute to ferromagnetism, owing to their magnetic

1

THE SCIENTIFIC MONTHLY

JANUARY, 1934

DISCOVERY AND EARLY HISTORY OF THE POSITIVE ELECTRON

By Dr. KARL K. DARROW
BELL TELEPHONE LABORATORIES

NEGATIVE electrons have been known by now for close on forty years. In this brief period they have been found in almost every situation and held accountable for almost all phenomena. It is a certain and familiar fact, attested daily by the working of countless millions of technical devices which depend upon it, that they can exist and travel freely in a rarefied gas or in a vacuum. Metals are believed to be full of freely wandering electrons, circulating past and even through the atoms. Negative electrons are thought to be responsible for almost every emission or absorption or scattering of light. They are a part of every atom-model and therefore one of the elements of every recent physical image of the world.

During all these years of the ubiquity of its negative counterpart, the *positive* electron has been "conspicuous by its absence." For twenty years at least it has been taken almost as a dogma that no such thing exists; it has been taken for granted that positive charge never appears in nature and must never appear in an atom-model, unless it is loaded down with a mass more than a thousand times as great as that of an electron. Two years ago the theorist Dirac was guided by theories of his own to conceive the positive electron, but no one had foreseen the way in which it was eventually found, and no one so far as I know was looking for it. It arrived as a by-product of the study of cosmic rays, thus adding one more instance to the list already long of great discoveries casually made during researches otherwise intended.

In the course of a big program of cosmic-ray research which Millikan had organized, C. D. Anderson was using the expansion-chamber or cloud-chamber of C. T. R. Wilson to observe the paths of ionizing particles belonging to these rays. Millikan and Anderson were not the first to do this. Others had already photographed the long straight tracks which these particles leave behind them (Skobelzyn's is the earliest name to be mentioned in this connection) and had furthermore observed that in a moderate magnetic field these tracks are either very slightly curved or not sensibly curved at all, so that the corpuscles which make them must be very fast. Anderson, however, made two important innovations.

Previously, everybody had set up the expansion-chamber with its axis vertical and its breadth horizontal, which is the most convenient way. Be it remembered that this chamber is a broad short cylinder filled with air or some other gas, its upper end closed by a transparent plate of glass, its lower end by a movable wall, which is the head of a piston. When the piston-head is suddenly pulled back through the proper distance (a matter of careful adjustment) and the gas expands to the proper degree, the moisture in the gas condenses upon whatever ions

5

Figure 9: Exemplary cover pages to illustrate The Bell Systems Technical Journal as well as articles published by Bell Labs personnel in academic journals.

E. Phase III: Maturation

The recognition of the point-contact configuration as a solid-state amplifier design during the 1947/1948 period represented an important milestone in the emergence of transistor technology. As described in section D2, several measures followed which collectively catalyzed a sort of phase change in the field of semiconductor research. This is reflected in the increase of research activity and corresponding publications, as captured in Fig. 1. Nevertheless, the point-contact design still had major flaws which prevented it from being adopted for large-scale dissemination.

The point-contact configuration represented a design that worked comparatively reliably and it was comparatively well understood. In contrast to earlier designs, it was capable of routing high frequency signals. Small batches were produced as early as 1949 in the US and in Europe and used in niche applications such as hearing aids. However, customer dissatisfaction was high as early point-contact transistors were sensitive to mechanical disturbances as well as humidity and temperature changes. The two wires that touched the germanium crystal were extremely delicate and could easily be disrupted. For similar reasons, manufacturing was difficult. Having two wires come in from the top and touch the surface of the crystal in just the right way was not easily and reliably automatable

Both of these problems were only satisfactorily solved with the development of the MOS-FET design in 1959. The MOS-FET would come to be recognized as the dominant design of transistor technology and its simplicity and manufacturability led to its mass scale deployment in the form of integrated circuits. As a result, today more than one quintillion (10^{18}) MOS-FETs have been produced and deployed.

I describe as "maturation period" the period between the development of the point-contact transistor with its associated shift in semiconductor research activity and the development of the MOS-FET as the transistor's dominant design. Key milestones of the maturation period are summarized in Fig. 10 and will be discussed in the remainder of this section.

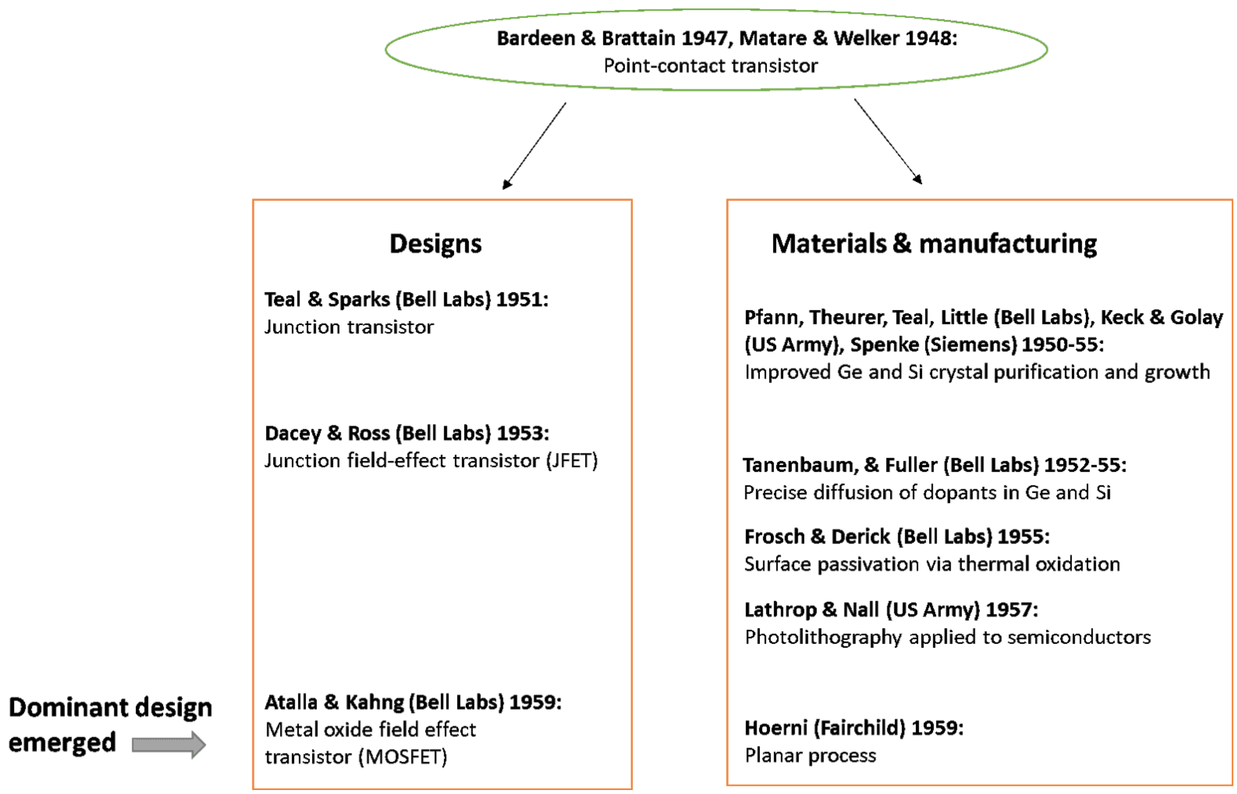


Figure 10: Milestones in designs and materials/manufacturing categories after the creation of first point-contact transistor prototypes. The 1959 creation of MOS-FET prototypes alongside the matching planar production process represents the final emergence of the transistor's dominant design.

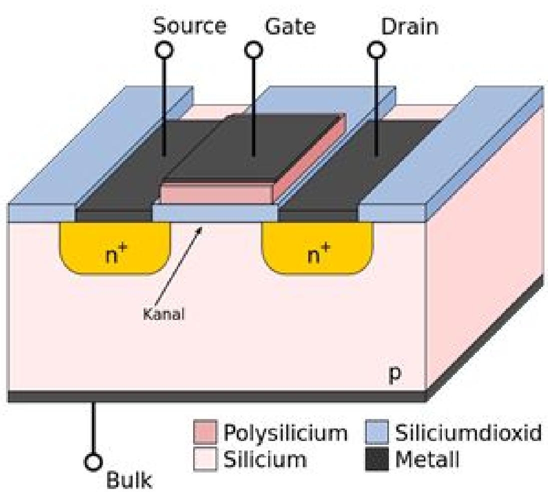
The junction transistor design – first produced in 1950 by Gordon Teal and Morgan Sparks at Bell Labs (Bassett 20020) – represented a more robust design compared to the point-contact transistor. In the junction transistor, the contact wires do not come in from the top but rather from the side onto sandwiched layers of N-type and P-type semiconductor materials. Effectively, this changed the design from a 3D design to a 1D design with a single axis along which different elements of the device were stacked. This eliminated the fragile point contacts and improved manufacturability. However, challenges still remained when considering the effective integration of several such transistors into a single unit.

The question of how to best produce at once entire circuits consisting of several components rather than simply producing individual components had been considered all along the development of 20th century electronics. As for transistors, the first formal proposal to integrate several modules into a single "integrated circuit" is typically attributed to a 1949 patent by Werner Jacobi at Siemens. However, Jacobi's patent remained at the conceptual level and did not discuss technical details as to how such an integration would be accomplished.

Several subsequent discoveries contributed to that end. During the early 1950s, Morris Tanenbaum and Calvin Fuller at Bell Labs developed techniques to create N-type and P-type regions in situ i.e. at specific sites on an already made crystal. This was possible through deposition of impurities onto the crystal surface and subsequent diffusion into the bulk via precisely controlled heating. This technique represented an improvement over doping techniques in the 1940s where impurities were added to entire crystals uniformly during crystal production.

Another major advance resulted from a Bell Labs group managed by Mohamed Atalla that initially aimed at making junction transistors more robust to their environment through surface coverings. By that time silicon was used again in semiconductor research alongside germanium. Atalla's team members Carl Frosch and Lincoln Derick accidentally grew a silicon dioxide layer on a silicon crystal and noticed that it was suitable as a masking material to selectively apply diffusion-based doping. In the course of exploring the properties of such silicon dioxide layers, the researchers also noticed that they reduced the field-blocking effect of the semiconductor's surface layers – a phenomenon now called "surface passivation." This meant that the early field effect transistor designs suddenly became attractive again.

The work by Attalla's group on oxide layers was complemented by further refinements in manufacturing techniques. These include the demonstration of photolithography techniques for applying masked patterns to semiconductors, a development attributed to a team at a US Army-affiliated research lab (Lathrop & Nall). In 1959, Jean Hoerni integrated the above-mentioned techniques into a proposed production process for a large number of field effect transistors on a single semiconductor crystal, arranged in a pre-defined pattern and created layer by layer. This production process (later known as planar process) combined with Atalla's refinement of the field effect transistor design with an oxide layer (later known as Metal-Oxide-Semiconductor Field-Effect Transistor, i.e. MOS-FET) represented the final major step toward transistor scalability. The MOS-FET design was first implemented in a working device in November 1959 and published in June 1960.



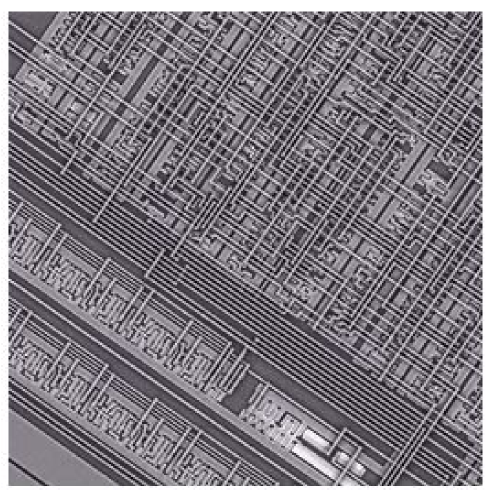


Figure 11: LEFT: Schematic of 1959 MOS-FET design (compare with Lilienfeld's 1925 FET design in Fig. 3). RIGHT: Thousands of MOS-FET transistors arranged in an integrated circuit (image taken via electron microscopy).

F. Discussion

Several issues raised in this article warrant further discussion:

1. Invention narratives and narrative bias

Analog to the physical sciences, historical analysis requires reduction – ideally reduction to essential features of the observed system or phenomenon. In the case of transistor invention, the question then is: What are essential features of the invention process that ought to be included in an invention narrative? Science historian Allchin (2003) describes this challenge and its implications as follows: "History may seem transparent. For some, it may seem only a matter of describing plain historical facts. Yet selective use of such facts, like selective use of facts in science, can mislead."

In this article, I presented two alternative narratives of transistor invention: a first one, modeled on much of the received literature, which I labeled the "individual narrative"; and a second one, largely novel, which I labeled the "institutional narrative." The first narrative strongly focuses on the 1945-1948 period, and in particular on the two "magic months" in late 1947/early 1948. It also focuses heavily on a small group of key people whose individual actions and even personalities are depicted as critical. The alternative narrative views transistor invention as a process that spans a much wider period – stretching from at least the early 1920s to the late 1950s – and that involves many different participating groups and many different types of relevant knowledge. Many of the activities that the institutional narrative rests on took place in parallel. Allchin refers to the complexity and interdependence of such processes as "a web of history" rather than a singular timeline. Moreover, this alternative narrative has been structured into three phases – exploration, consolidation, and maturation – based on the growth of publications and, eventually, commercial products.

Epstein (1928) already cautioned that different approaches to invention narratives can distort the perception of the invention process as such. He distinguished between two major types of invention narratives: one which he called the "heroic" narrative and another which he called the "systematic" narrative. More recently, sociologist Tufekci (2019) presented a similar distinction between narrative approaches. She refers to the first narrative approach as "psychological" or "hero/antihero" narrative and the second one as a "sociological" or "institutional" narrative. Tufekci then cautions: "Whether we tell our stories primarily from a sociological or psychological point of view has great consequences for how we deal with our world and the problems we encounter." Tufekci also stresses that we are culturally susceptible to follow the "psychological" (i.e. individual) narrative by default: "Hollywood mostly knows how to tell psychological, individualized stories. They do not have the right tools for sociological stories, nor do they even seem to understand the job."[14] Defaulting to psychological story-telling has also been referred to as "narrative bias" (Heshmat 2016).

Institutional narratives do not imply that individuals play no roles in them. Rather, as Tufekci suggests: "In sociological storytelling, the characters have personal stories and agency, of course, but those are

[14] Maybe the most notorious example that Tufekci provides for the contrast between sociological and psychological storytelling is the first and last season of Game of Thrones. Tufekci (2019): "The appeal of a show [like the early seasons of Game of Thrones] that routinely kills major characters signals a different kind of storytelling, where a single charismatic and/or powerful individual, along with his or her internal dynamics, doesn't carry the whole narrative and explanatory burden." Other authors have associated Isaac Asimov's Foundation novels to sociological storytelling in contrast to George Lucas' Star Wars (Hayden 2019).

also greatly shaped by institutions and events around them. The incentives for characters' behavior come noticeably from these external forces." Similarly, science historian Allchin (2003) warns of what he calls "the architecture of scientific myths" that includes biases towards "larger than life" characters, idealization, personal drama, and ex post sanitization and rationalization of invention narratives. He argues that "many stories romanticize scientists, inflate the drama of their discoveries, and oversimplify the process of science." In some cases, observers approach an invention process with a certain notion or ideal of how science "should work" and then try to fit the historical material into this *a priori* mold rather than the other way around i.e. letting empirical observations shape our notion of science. Allchin argues that in such cases, these authors "appropriated history for promoting a particular notion about how science works." (see also Bauer 1994; Woodcock 2014). As a countermeasure, Allchin suggests: "Suspect simplicity. Beware vignettes. Embrace complexity and controversy. Discard romanticized images. Do not inflate genius. Mix celebration with critique. Scrutinize retrospective science-made. Revive science-in-the-making. Explain error without excusing it. And above all respect historical context."

With these caveats laid out, whose responsibility is it then to formulate appropriate invention narratives – invention narratives that allow for the derivation of generalizable insights and transferable lessons? Should Bell Labs be faulted for not being more forthcoming about its engagement of a wide range of informal knowledge sources? Here, I argue one ought to consider the incentives of different key actors. In the case of transistor invention, neither Bell Labs as an organization (e.g. as represented by its PR department) nor individual researchers (e.g. Shockley) had incentives to point out the full range of "inspirations" that underlay their work. Clearly, many organizations and individuals will tend to opt for narratives, consciously or unconsciously, that emphasize, and maybe maximize, their own contributions. Related IP concerns about prior art claims have already been mentioned. But even more generally speaking. science historian Bassett points out: "Corporations typically tell the history of technology from their own perspective and have little incentive to emphasize the contributions of others." This is consistent with characterizations of Bell Labs' PR department engaging in "myth-making" (Riordan 1998) and cultivating a well-crafted self-image. Shockley in turn was described as having a "marauding ego" (Gertner 2012) and described himself that striving for personal glory was a key motivating factor for him (Shockley in Essers & Rabinow 1974). Shockley himself was prolific in providing personal accounts of transistor invention, much of which was picked up uncritically by later observers. One may speculate that an effective way for maximizing personal glory is to be the author of one's own story. Here enters the scholar and her responsibility in studying such historic developments. Rather than conveniently relying on information "from the horse's mouth," a certain critical distance and appreciation of the bigger picture is called for. One might argue that a lack of the latter resulted in the flood of heroic narratives of transistor invention which got in the way of more cooled-headed institutional analyses that tend to yield more generalizable and transferable lessons.

A takeaway point here is to remain cautious of simplified invention narratives that are overly focused on individuals instead of institutions, incentives and their wider context. A particular caveat involves primary sources with clear incentives toward a certain "spin" of their stories.

2. Emerging inventions

A central thesis of this paper is for the transistor to be viewed as an *emerging invention* – in contrast to a singular one – with a period of emergence spanning roughly from the early 1920s to the late 1950s.

Emergence is a concept frequently employed in systems theory to describe the evolution of complex systems. In this view, "emergent entities [..] 'arise' out of more fundamental entities and yet are 'novel' or 'irreducible' with respect to them" (O'Connor 2020). Paoli 2003 describes emergence as "linked to the process of transformation of the parts into a whole which, by this very process, forms, and transforms, maintains and organizes complementary tendencies, creates diversity, forges links between and organizes antagonisms, organizes antagonism within complementarities." Processes of emergence are often characterized as "phase changes" (Rennie 2018). Emergence has been described to proceed through the integration of parts at one "integrative level" toward a new whole on another, higher integrative level (Needham 1937). In this paper, the existence of a phase change has been argued to be empirically reflected in the substantial change of publication growth rates of the semiconductor literature during the 1945-50 period as well as the subsequent occurrence of commercial solid-state amplifier products. This dynamic was illustrated in Fig. 1 and associated with three phases: an exploration phase, an intermediary consolidation phase, and a maturation phase. In the case of transistor invention, the five categories of the exploration phase (Fig. 2) can be viewed as parts that enabled and drove subsequent integration during the consolidation phase. This intermediary consolidation phase then led into the maturation phase. In a more generic framework of complex system evolution, as sketched out above, the three phases can then be viewed as corresponding to integrative level 1, transition, and integrative level 2.

This paper is focused on a single case and a single technology; therefore, generalizations need to be treated with caution. Nevertheless, the general concept of emerging inventions, the three proposed phases of emerging inventions, and the simple framework for distinguishing between relevant exploration phase factors may well be applicable to other technologies as well. More research is needed to investigate such hypotheses.

So far, several authors have considered the emergence of inventions at the intersection of converging fields. This includes Hacklin 2007 who considers innovation resulting from the convergences of previously unrelated areas. Hacklin distinguishes between knowledge convergence, technological convergence, applicational convergence, and industry convergence. Whereas the latter three refer to farther downstream technical domains compared to the knowledge types considered in this present paper, Hacklin's first type of convergence appears to apply to some extent: "Knowledge convergence denotes the emergence of serendipitous coevolutionary spill-over between previously unassociated and distinct knowledge bases". However, even in this most general category, Hacklin's examples are comparatively more downstream: he considers firms that integrate knowledge areas around central processing units (CPUs) chipsets and communication technology chipsets. Maine et al. 2014, building on Sharp & Langer 2011, argue that "there is an enormous potential for innovation from the confluence of technologies" whereas confluence of technologies is defined as "a new combination of previously distinct technologies." Among their three cases, Maine et al. in fact specifically reference the transistor, although they do not consider the pre-1945 period and instead focus on what I called the maturation period i.e. the developments towards the integrated circuit. Arsov 2013 indirectly and somewhat informally considers convergence in his technical review of transistor invention: "The more we study the history of an invention, the fewer examples we find of entirely new devices conceived and perfected by one individual in isolation." With respect to Bell Labs specifically, Arsov concludes: "Looking at the existing documents, we might get the impression that Bell Labs did not invent the transistor, but that they re-invented it. Yet, what is more important is that they succeeded in its practical realization, although they were not the only ones to do so."

The concept of emerging inventions suggests that inventions can play out in more fragmented, distributed ways across longer periods of time than commonly assumed (implying a division of labor of sorts); and that such processes of emergence may not necessarily be recognized as such while still ongoing. This then raises questions as to how the potential for present and future emerging inventions can be identified and assessed; and questions on which actors contribute to the invention process in what ways.

3. The division of labor in emerging inventions

The literature on open innovation considers the division of labor between actors in innovation ecosystems as well as knowledge transfer between such actors (see Chesbrough 2006; Beck et al. 2020 for a recent review). Open innovation has been defined as "a distributed innovation process based on purposively managed knowledge flows across organisational and sectoral boundaries using pecuniary or nonpecuniary mechanisms" (Bogers et al. 2017).

When considering the activity of "generating innovation" Chesbrough 2006 distinguishes between four types of roles in this area: "innovation explorers, merchants, architects and missionaries". Innovation explorers are almost by definition associated with the conventional view of Bell Labs: Chesbrough argues that "innovation explorers specialize in performing the discovery research function that previously took place primarily within corporate R&D laboratories." In contrast, "innovation merchants will innovate but only with specific commercial goals in mind, whereas explorers tend to innovate for innovation's sake." Chesbrough 2016 and Bogers et al. 2019 describe innovation architects as follows: "in a world of widely diffuse useful knowledge, much of the real value can be gained not from developing yet another piece of knowledge, but rather from creating systems and architectures that combine these disparate pieces of knowledge together in useful ways that solve real problems." Finally, typical innovation merchants are described as nonprofit organizations driven by certain ideals.

Returning to the proposed three phases of emerging inventions, one may want to consider the significance of each of Chesbrough's roles during each phase. For instance, during the exploration phase, innovation architects can be premature whereas much activity is centered on innovation explorers. However, during the consolidation phase, the innovation architect takes center place and innovation explorers may take more of a backseat (unless they can make the jump to also become architects; or find other domains that match their exploration capabilities). Finally, innovation merchants grow in significance during the maturation stage.

In light of the body of this paper, one can argue that the early Bell Labs – at least with respect to its semiconductor research effort – appears to match the definition of an innovation architect – and to some extent that of an innovation merchant – better than that of an innovation explorer.

The division of labor in emerging inventions has implications for practitioners such as policy makers and executives: decision makers are to not expect a single blueprint of success in the context of emerging inventions. Rather, several roles can be occupied at several stages and their relative attractiveness is a function of an organization's capabilities as well as its timing.

4. Systems integration in the emerging invention ecosystem

In contrast to open innovation, Chesbrough emphasizes that in closed innovation "a company generates, develops and commercializes its own ideas" (Chesbrough 2006). Here, remarkably, Bell Labs is often presented as a prototypical case of a closed innovation research organization; for instance, in Chesbrough (2003b): "AT&T and Bell Labs are another dramatic example of a closed innovation approach. [...] It was a great scientific and technological fountain of resources." However, as argued above, de facto Bell Labs exhibited many of the major traits of an open innovation organization – even though the company or its employees did not seem to emphasize these traits in their own narratives about their work (which appears to be the cause for the later widespread perception of Bell Labs as a closed innovation organization).

The notion of an innovation architect is closely related to the notion of a systems integrator. In Prencipe et al. 2003, the following definition of systems integration as an organizational capability is presented: "the meta-process of systems integration is above all the integration of knowledge." However, in order to be capable of integrating knowledge sourced externally, firms also "must retain and dominate, in-house, a whole host of generative contexts of knowledge in order to control systems integration." Prencipe 2003 also emphasizes the "relevance of external sources of component and knowledge for a firm's competitive advantage." As a consequence, "managing external relationships (through the development and maintenance of an extensive flow of information across the boundaries of the firm) becomes critical".

Closely related to the concept of systems integration, is the concept of "architectural or integrative capabilities" as put forth by Henderson & Cockburn (1994). These capabilities are defined as "the ability to access new knowledge from outside the boundaries of the organization and the ability to integrate knowledge flexibly across disciplinary and therapeutic class boundaries within the organization."

A related perspective on knowledge integration that shifts attention from the organizational to the operational level is provided by Lester & Piore 2009. They suggest that activities of good managers involve "initiating and guiding conversations among individuals and groups." More specifically, they compare the manager's role in animating these conversations to "the role of the hostess at a cocktail party, identifying the 'guests,' bringing them to the party, suggesting who should talk to whom and what they might talk about, intervening as necessary to keep the conversations flowing." This image clearly evokes such interactions as described in section C2, when manager Kelly deliberately connected Ohl – the tinkerer familiar with semiconductor anomalies of the radio community – with Shockley – the MIT PhD and expert in theoretical solid-state physics.

The role Bell Labs played under Mervin Kelly during the 1945-1950 transformation of the semiconductor field matches in many ways above-mentioned definitions related to knowledge integration and specifically the integration of diverse, externally sourced knowledge.

Considering the early Bell Labs in this light raises a number of questions for practitioners: Which fields today would benefit from such kinds of integration activities? And which organizations today are capable of playing that role? By extension: Can organizations be purposefully designed to take on such roles and if so, how?

5. Formal and informal knowledge

A key insight from this paper concerns the role of informal knowledge among externally sourced knowledge.

Here, formal knowledge refers to knowledge transferred through formal relationships with external partners. This includes such mechanisms as formal research collaborations, licensing agreements, and firm acquisitions. Informal knowledge then refers to knowledge absorbed through non-formal mechanisms. This includes the study of patents and publications – including nonacademic ones, insights from research trips, and technical anecdotes. Informal knowledge defined as such is different from tacit knowledge (Polanyi & Sen 1967) since informal knowledge can still be codified – although not necessarily in a systematic and consistent way. What comes perhaps closest to this definition of informal knowledge is what some scholars have called "practical knowledge" or "clinical knowledge" (Piore 2018). The notion of clinical knowledge originates in the kind of experiential, practical knowledge employed by clinicians which often complements and sometimes inspires academic or formal knowledge.

The study of knowledge flows is at the heart of the open innovation literature. However, much research in this literature focuses on flows of formal knowledge, rather than informal knowledge. This is perhaps understandable, given that formal knowledge is often codified and more readily accessible and measurable. However, this research preference creates a bias against understanding the role of informal knowledge – which this present paper suggests to be problematic as informal knowledge can indeed play significant roles in emerging inventions such as in the case of the transistor.

The diversity and complexity in the nature of knowledge has been pointed out by other authors. Paoli (2003) suggests: "We argue that knowledge tends to be increasingly represented in unique [...] and transdisciplinary (by transcending the classic boundaries between disciplines) ways". Piore 2018 specifically draws attention to clinical knowledge: "The relationship between formal and clinical knowledge is unclear in large part because clinical knowledge is seldom explicitly recognized, and because it goes unrecognized it is understudied." Piore adds: "Recognition is complicated" in part because it "draws on anecdotal evidence which is easily dismissed as atypical or anachronistic."

Returning to the transistor case, I argue that in the case of transistor invention, the sourcing, interpretation, evaluation, and integration of various forms of informal knowledge played a critical role in the invention process. In the concrete case, such informal knowledge included information from patent specifications, reports of anomalies in amateur and popular media reports, as well as oral anecdotes regarding certain physical behaviors and technical configurations. Stressing the significance of informal knowledge does not downplay the significance of formal knowledge. Rather, I argue that both types of knowledge were critical in transistor invention – and that Bell Labs excelled, like few, others in working with both types of knowledge.

The above suggests that informal knowledge ought to be further investigated, including its various types and its role in invention processes. There are also practical implications: This study suggests that practitioners may want to pay more explicit attention to informal knowledge, including its creation, propagation, and evaluation.

The lack of recognition of informal knowledge and its significance is also reflected in certain institutional behavior. For instance, many scientific publications are strongly biased towards publishing only research results where little or no ambiguity remains. However, this present paper suggests that even anomalous or imperfect results can represent important data points relevant to the decision-making of research managers. Such data points – as long as they are understood for what they are i.e. to exhibit larger uncertainty and associated error bars than other data points – can influence the assessment of opportunities and risks as well as the allocation of resources and research timing.

6. The significance of timing

Section F3 above included a short discussion of timing in view of different phases of emerging inventions and different roles of actors in the innovation ecosystem. Specifically, I argued that during the exploration phase, the activity of the innovation architect is premature. Rather, the would-be innovation architect needs to understand what conditions need to be met for integration/consolidation and subsequent maturation to become feasible. In the case of the transistor, I suggested that five conditions needed to be sufficiently advanced and aligned for that to be the case: concepts, observed anomalies, materials, theory, and economic demand.

Once sufficient advance and alignment occurs, a phase change (in the sense of the emergence literature) becomes feasible. Some authors have likened such aligned preconditions and subsequent transitions to situations where breakthroughs are “in the air” (Gladwell 2008). The frequent occurrence of multiple simultaneous but independent discoveries within short periods of time provides some empirical evidence for this assertion (Epstein 1926; Merton 1961; Lemley 2011; Griswold 2012).

This discussion of timing is somewhat reminiscent of the entry timing literature which traditionally concerns itself with the timing of market entry. Metzler 2019 discusses entry timing in view of product development at the intersection of converging technologies. In that research, too, the findings suggest that converging knowledge domains ought to be assessed along key metrics, each of which will exhibit certain thresholds whose crossing marks the feasibility for integration activities – and thus for subsequent market entry.

Returning to the transistor case, I suggest that it may be viewed in a similar light: particular thresholds needed to be crossed in each of the five domains proposed in Fig. 2 for integration to become feasible. For concepts, this threshold may be viewed as the first publication of a solid-state amplifier design with accompanying data and sufficient detail for reliable reproduction (such as in Hilsch & Pohl 1938); for anomalies, the reliable production of anomalies due to refined materials (also partially owed to Pohl); for theory, the emergence of quantitative first-principle models that matched experiments (such as exemplified by Slater’s and Wigner’s work); for materials, the reliable production of single-crystal high-purity germanium and silicon; and for economic demand, the solid expectation for long-term and growing amplifier demand. It was when these thresholds were crossed, that integration across these increasingly related knowledge domains became feasible.

It is such dynamics that strong managers ought to be aware of and act accordingly. As was the implication both in Metzler 2019 and in this paper, managers ought to 1) identify relevant knowledge domains; 2) assess in each domain the present state of the art and estimate thresholds that enable integration; 3) estimate in each domain the rate of improvement – and thus determine the expected point of convergence; 4) enter into development efforts at the intersection of these knowledge domains at the earliest time feasible – but not earlier.

In the case of solid-state amplifiers, threshold conditions were not met before the late 1930s, which would have made any earlier attempts for integration and commercialization hopeless unless the integrating organization would have single-handedly filled the remaining knowledge gaps. This would explain the lack of success of early patents, as discussed in section C1 and C5. Note that this implies that Bell Labs' timing was excellent. Bell Labs got involved in the emerging semiconductor field early on – but not too early. By the late 1930s, when Kelly had just formed his first solid-state physics teams, relevant factors just about aligned to make integration feasible. At the same time, this was not too early. Had Bell Labs started a semiconductor group already during the 1920s, it might have well run out of steam before any tangible progress had been made. At that time, too much work was still needed – for instance with regard to theory and materials – and such amounts of basic research could not have been done inhouse at Bell Labs. This would have been the kind of long-term "basic research" often associated with Bell Labs – but in reality, it would have simply note been economically justifiable in view of remaining risks and projected returns.

Again, these arguments are reminiscent of discussions in Metzler 2019, specifically with respect to Apple's entry into the mobile phone industry: Apple recognized early on the impending convergence between computation and telecommunication technologies as well as corresponding thresholds. Apple then got involved in product development of a convergence device (the iPhone) at the earliest possible time when performance thresholds were crossed and when a commercial device had thus become feasible.

A lack of appreciation of this timing aspect is reflected in some observers' comments about Bell Labs. For instance, Chesbrough 2006 argues: "Lucent can no longer sustain the investigation of basic scientific phenomena in Bell Laboratories because it can't make use of it fast enough to warrant the high level of investment." However, I suggest that Bell Labs – and particularly the early Bell Labs which this paper focuses on – was perhaps never the kind of patient, long-term "basic science" research organization as which it is often presented today (or as which it has presented itself). Rather, in the case of solid-state amplifiers, it started out with a small, lean effort in 1936 which initially mostly studied the lay of the land; Bell Labs then determined in 1945 that the time was right to make a substantial push in this area. It obtained preliminary results less than two years later, and immediately began to exploit them through patents and PR which served as a justification for a scale-up of research activities. This approach has more of the feel of a lean, well-timed commando operation rather than a long-term siege.

This discussion bears a number of implications: Certain types of inventions may be almost inevitable and essentially simple a matter of time once critical thresholds in relevant knowledge domains are crossed and boundary conditions for continued scientific progress maintained. At the same time, it may be very difficult if not impossible to prematurely force an invention if its time has not yet come i.e. if the knowledge domains it relies on have not yet sufficiently evolved. Organizations involved in emerging inventions will want to be aware of the progression of relevant knowledge domains, remaining knowledge gaps, and estimated thresholds. Such insights can then inform the choices as to the amount of resource to be allocated and the type of activities to be engaged in. Organizations would do well to carefully consider the state of the emerging invention; the roles they can and want to play; and when the right time for such a desired role has come.

G. Conclusions

The broader goal of this research was to contribute to our understanding of the occurrence of major inventions. To this end, the article investigated – from technical and managerial perspectives – the emergence of transistor technology, starting with early indications and concepts of solid-state amplification during the 1910s and 1920s to the development of the transistor's dominant design in the form of the MOS-FET during the 1950s.

As part of this process, several widely held beliefs about transistor invention were challenged. This includes the notion that transistor invention ought to be viewed as a singular event, centered around two months in late 1947/early 1948. Instead, I proposed for transistor invention to be seen as a process – a process that spanned several decades. Also challenged is the interpretation of Bell Labs' role in transistor invention. In the transistor context, Bell Labs is typically described as a monolithic inward-looking R&D organization whose long-term basic science orientation produced revolutionary inventions, almost as side effects. I argued that – rather than conforming to this image of a prototypical closed innovation organization – Bell Labs in fact played the role of a systems integrator. As such, Bell Labs managed to source externally a wide range of relevant knowledge and integrated such knowledge skillfully – an activity that was conducted with mission- and device-orientation as well as good timing.

More specifically, I proposed for the transistor to be viewed as an *emerging invention* which implies an invention process comprising multiple phases. I proposed three phases which I called exploration phase (1910-1945), consolidation phase (1945-1950), and maturation phase (1950-). Across these phases, the activities of different actors can be associated with different roles such as Chesbrough's innovation explorers, innovation architects, and innovation merchants. In this picture, Bell Labs appears to correspond much more to an innovation architect rather than an innovation explorer.

In bringing about the maturation phase which led to the emergence of the transistor's dominant design and ultimately to commercialization, the earlier exploration and consolidation phases were critical. I argued that the exploration phase was characterized by the simultaneous evolution of several knowledge domains whose respective advancements were preconditions for subsequent integration and consolidation. The relevant domains include 1) proposed design concepts; 2) an empirical body of observed anomalies; 3) theoretical understanding of underlying physics; 4) advances in materials design and production. As a fifth relevant factor, which in this case is not itself a knowledge domain, I identified economic demand. I argued that once thresholds were crossed in view of each of these factors, consolidation became feasible and did indeed take place in short order.

The effectiveness and pace of such consolidation did, however, depend on the nature of the dominant systems integrator i.e. in this case Bell Labs. In this regard, Bell Labs exhibited many relevant capabilities. The organization, under the leadership of Mervin Kelly, managed to field various sources of external knowledge, assess, and digest such knowledge. The organization was further able to effectively integrate the various types of relevant knowledge which accelerated the pace of invention. Particular emphasis was placed in this paper on the role of informal as opposed to formal knowledge. Here, informal knowledge specifically refers to such knowledge that originated from patent specifications, technical articles and technical reports in nonacademic magazines, as well as oral reports of anomalous behavior of materials. Bell Labs recognized the significance of such types of informal knowledge and made productive use of it by combining and integrating it with complementary formal knowledge.

Although emphasis here is placed on a single case – the transistor – I suggest that some implications of these findings may be generalizable to other contexts. This may particularly pertain to 1) the presented considerations about the concept of *emerging inventions*; 2) the significance of informal knowledge; and 3) the role of systems integrators in the context of emerging inventions. More research is needed to determine the extent to which such concepts and conjectures apply to other circumstances.

References

Allchin, D. (2003). Scientific myth-conceptions. *Science Education*, *87*, 329–351.

Arns, R. G. (1998). The other transistor: Early history of the metal-oxide semiconductor field-effect transistor. *Engineering Science & Education Journal*, *7*, 233–240.

Arsov, G. L. (2013). Celebrating 65th anniversary of the transistor. *Electronics*, *17*, 63–70.

Bardeen, J. (1936). *Quantum Theory of the Work Function* (PhD Thesis). Princeton University.

Bardeen, J. (1947). Surface states and rectification at a metal semi-conductor contact. *Physical Review*, *71*, 717.

Bardeen, J., & Brattain, W. H. (1949). Physical principles involved in transistor action. *Physical Review*, 75, 1208.

Bardeen, J., & Brattain, W. H. (1950). Three-Electrode Circuit Element Utilizing Semiconductive Materials. *United States Patent No. US2524035A*.

Bassett, R. K. (2002). *To the digital age: Research labs, start-up companies, and the rise of MOS technology*. JHU Press.

Bauer, H. H. (1994). *Scientific literacy and the myth of the scientific method*. University of Illinois Press.

Beck, S., Bergenholtz, C., Bogers, M., Brasseur, T.-M., Conradsen, M. L., Marco, D. D., … Xu, S. M. (2020). The Open Innovation in Science research field: A collaborative conceptualisation approach. *Industry and Innovation*, *0*, 1–50.

Bloch, F. (1929). Über die Quantenmechanik der Elektronen in Kristallgittern. *Zeitschrift Für Physik*, *52*, 555–600.

Bogers, M., Chesbrough, H., Heaton, S., & Teece, D. J. (2019). Strategic Management of Open Innovation: A Dynamic Capabilities Perspective. *California Management Review*, *62*, 77–94.

Bogers, M., Zobel, A.-K., Afuah, A., Almirall, E., Brunswicker, S., Dahlander, L., … Wal, A. L. J. T. (2017). The open innovation research landscape: Established perspectives and emerging themes across different levels of analysis. *Industry and Innovation*, *24*, 8–40.

Bray, R. (1982, May 14). *Interview of Ralph Bray by Paul Henriksen*. Niels Bohr Library & Archives, American Institute of Physics. Retrieved from https://www.aip.org/history-programs/niels-bohr-library/oral-histories/4533

Bridgers, H. E., & Biondi, F. J. (1958). *Transistor technology* (Vol. 1). Van Nostrand.

Chesbrough, Henry W. (2006). The era of open innovation. *Managing Innovation and Change*, *127*, 34–41.

Chesbrough, Henry William. (2003). *Open innovation: The new imperative for creating and profiting from technology*. Harvard Business Press.

Crawford, B. E. (1991). *The invention of the transistor* (MS Thesis). University of Vermont.

Cuff, T. (1993). *The Copper Oxide Rectifier*. Retrieved from https://www.researchgate.net/publication/294259463_The_Copper_Oxide_Rectifier

Darrow, K. (1964, April 2). *Interview of Karl Darrow by Henry Barton and W. J. King*. Niels Bohr Library & Archives, American Institute of Physics. Retrieved from https://www.aip.org/history-programs/niels-bohr-library/oral-histories/4567-1

Davydov, B. (1937). On the theory of the motion of electrons in gases and semi-conductors. *Phys. Zeit. Der Sow*, *12*, 269–300.

Davydov, B. (1938). On the rectification of current at the boundary between two semiconductors. *Compt. Rend. Doklady. Acad. Sci. USSR*, *20*, 279.

Davydov, B. (1939). On the contact resistance of semiconductors. *Zh. Eksp. Teor. Fiz*, *9*, 451.

Drude, P. (1900). Zur Elektronentheorie der Metalle. *Annalen Der Physik*, *306*, 566–613.

Epstein, R. C. (1926). Industrial Invention: Heroic, or Systematic? *The Quarterly Journal of Economics*, *40*, 232–272.

Essers, F., & Rabinow, J. (1974). *The Public Need and the Role of the Inventor: Proceedings* (Vol. 388). National Bureau of Standards.

Frenkel, J. (1931). Some remarks on the theory of the photoelectric effect. *Physical Review*, *38*, 309.

Gertner, J. (2012). *The idea factory: Bell Labs and the great age of American innovation*. Penguin.

Gladwell, M. (2008). In the air. *The New Yorker*, *84*, 50–60.

Griswold, M. (2012, November 25). Are Inventions Inevitable? Simultaneous Invention and The Incremental Nature of Discovery. Retrieved from The Long Nose website: https://mgriz.wordpress.com/2012/11/25/are-inventions-inevitable-simultaneous-invention-and-the-incremental-nature-of-discovery/

Hacklin, F. (2007). *Management of convergence in innovation: Strategies and capabilities for value creation beyond blurring industry boundaries*. Springer Science & Business Media.

Handel, K. C. (1999). Anfänge der Halbleiterforschung und-entwicklung. *The Beginnings of Semiconductor Research and Development], Doctoral Thesis, Cologne Univ*, 29.

Heil, O. (1935). Improvements in or Relating to Electrical Amplifiers and Other Control Arrangements and Devices. *Great Britain Patent No. GB439457A*.

Henderson, R., & Cockburn, I. (1994). Measuring Competence? Exploring Firm Effects in Pharmaceutical Research. *Strategic Management Journal*, *15*, 63–84.

Hilsch, R., & Pohl, R. W. (1938). Steuerung von Elektronenströmen mit einem Dreielektrodenkristall und ein Modell einer Sperrschicht. *Zeitschrift für Physik*, *111*, 399–408.

Hoddeson, L. E., & Daitch, V. (2002). *True genius: The life and science of John Bardeen*. Joseph Henry Press.

Holst, G., & Geel, W. C. V. (1939). Electrode System of Unsymmetrical Conductivity. *United States Patent No. US2173904A*.

Katz, R., & Allen, T. J. (1982). Investigating the Not Invented Here (NIH) syndrome: A look at the performance, tenure, and communication patterns of 50 R & D Project Groups. *R&D Management*, *12*, 7–20.

Kelly, M. J. (1950). The Bell Telephone Laboratories—An example of an institute of creative technology. *Proceedings of the Royal Society of London. Series A. Mathematical and Physical Sciences*, *203*, 287–301.

Lashkaryov, V. E. (1941). Investigations of a barrier layer by the thermoprobe method. *Izv. Akad. Nauk SSSR, Ser. Fiz*, *5*, 442–446.

Lee, T. H. (2004). *The Design of CMOS Radio-Frequency Integrated Circuits*. Cambridge University Press.

Lemley, M. A. (2011). *The Myth of the Sole Inventor* (SSRN Scholarly Paper No. ID 1856610). Rochester, NY: Social Science Research Network.

Lester, R. K., & Piore, M. J. (2009). *Innovation–the missing dimension*. Harvard University Press.

Lilienfeld, J. E. (1930). Method and Apparatus for Controlling Electric Currents. *United States Patent No. US1745175*.

Lojek, B. (2007). *History of semiconductor engineering*. Springer.

Lossev, O. V. (1924). Oscillating crystals. *Wireless World Radio Rev*, *271*, 93–96.

Lossev, O. V. (1928). CII. Luminous carborundum detector and detection effect and oscillations with crystals. *The London, Edinburgh, and Dublin Philosophical Magazine and Journal of Science*, *6*, 1024–1044.

Maine, E., Thomas, V. J., & Utterback, J. (2014). Radical innovation from the confluence of technologies: Innovation management strategies for the emerging nanobiotechnology industry. *Journal of Engineering and Technology Management*, *32*, 1–25.

Matare, H. F., & Welker, H. (1954). Crystal Device for Controlling Electric Currents by Means of a Solid Semiconductor. *United States Patent No. US2673948A*.

Merton, R. K. (1961). Singletons and multiples in scientific discovery: A chapter in the sociology of science. *Proceedings of the American Philosophical Society*, *105*, 470–486.

Metzler, F. (2019). *Firms, Industries, and Technological Change: A Patent-Based Approach to Studying Disruption and Disruptors* (PhD Thesis). Massachusetts Institute of Technology.

Mott, N. F. (1938). Note on the contact between a metal and an insulator or semi-conductor. *Mathematical Proceedings of the Cambridge Philosophical Society*, *34*, 568–572.

Mott, N. F. (1939). The theory of crystal rectifiers. *Proceedings of the Royal Society of London. Series A. Mathematical and Physical Sciences*, *171*, 27–38.

Mott, N. F. (1976). Atoms in contact. *New Scientist (London)*, *70*, 172–174.

Mott, N. F. (2020). *Bristol Physics in the 1930s*. Retrieved from http://www.bristol.ac.uk/physics/media/histories/11-mott.pdf

Murty, K. K. (2008). *50 Timeless Scientists*. Pustak Mahal.

Needham, J. (1937). Integrative Levels: A Revaluation of the Idea of Progress. *Journal of Philosophy*, *34*, 612.

Nelson, R. R. (1962). The link between science and invention: The case of the transistor. In *The rate and direction of inventive activity: Economic and social factors* (pp. 549–584). Princeton University Press.

Nelson, R. R. (1982). *Government and Technical Progress: A cross-industry analysis*. Pergamon.

Nix, F. C. (1975, June 27). *Interview of Foster Cary Nix by Lillian Hoddeson*. Niels Bohr Library & Archives, American Institute of Physics. Retrieved from https://www.aip.org/history-programs/niels-bohr-library/oral-histories/4798

North, H. Q. (1945). *Final Report on K-band Germanium Crystals* (No. NDRC, 14-427, GK Co.).

O'Connor, T. (2020). Emergent Properties. In E. N. Zalta (Ed.), *The Stanford Encyclopedia of Philosophy* (Fall 2020). Metaphysics Research Lab, Stanford University.

Ohl, R. (1976, August 19). *Interview of Russell Ohl by Lillian Hoddeson*. Niels Bohr Library & Archives, American Institute of Physics. Retrieved from https://www.aip.org/history-programs/niels-bohr-library/oral-histories/4804-1

Ohl, R. S. (1949). Point Contact Negative Resistance Devices. *United States Patent No. US2469569A*.

Ohl, R. S. (1950). Electric Wave Transmission. *United States Patent No. US2522395A*.

Paoli, M. (2003). The cognitive basis of systems integration: Redundancy of context-generating knowledge. *The Business of Systems Integration*, 152.

Peierls, R. (1932). Elektronentheorie der Metalle. In *Ergebnisse der exakten Naturwissenschaften* (pp. 264–322). Berlin, Heidelberg: Springer.

Pickard, G. W. (1908). Oscillation-Receiver. *United States Patent No. US886154A*.

Piore, M. (2018). Economic Policy in the Time of Reactionary Populism. *Issues in Science and Technology*, *34*, 24–29.

Polanyi, M., & Sen, A. (1967). *The tacit dimension*. Doubleday New York.

Prencipe, A. (2003). Corporate strategy and systems integration capabilities: Managing networks in complex systems industries. *The Business of Systems Integration*, 114–132.

Prencipe, A., Davies, A., & Hobday, M. (2003). *The Business of Systems Integration*. Oxford University Press.

Ringer, W., & Welker, H. (1948). Leitfähigkeit und Hall-Effekt von Germanium. *Zeitschrift Für Naturforschung A*, 3, 20–29.

Riordan, M. (2005). How Europe missed the transistor. *IEEE Spectrum*, *42*, 52–57.

Riordan, M. (1998). *Crystal Fire: The Invention of the Transistor and the Birth of the Information Age*. New York: W. W. Norton & Company.

Schottky, W. (1938). Halbleitertheorie der Sperrschicht. *Naturwissenschaften*, *26*, 843–843.

Schottky, W. (1939). Zur Halbleitertheorie der Sperrschicht- und Spitzengleichrichter. *Zeitschrift Für Physik*, *113*, 367–414.

Shockley, W. (1936). Electronic energy bands in sodium chloride. *Physical Review*, *50*, 754.

Seitz, F., & Johnson, R. P. (1937). Modern theory of solids. II. *Journal of Applied Physics*, 8, 186–199.

Shive, J. N. (1949). The Double-Surface Transistor. *Physical Review*, 75, 689–690.

Silverman, M. (1947, July). Ma Bell's House of Magic. *Saturday Evening Post*.

Slater, J. C. (1934). The electronic structure of metals. *Reviews of Modern Physics*, 6, 209.

Sommerfeld, A. (1927). Zur Elektronentheorie der Metalle. *Naturwissenschaften*, *15*, 825–832.

Srinivasan, R., Lilien, G. L., & Rangaswamt, A. (2006). The emergence of dominant designs. *Journal of Marketing*, *70*, 1–17.

Suarez, F. F., & Utterback, J. M. (1995). Dominant designs and the survival of firms. *Strategic Management Journal*, *16*, 415–430.

Torrey, H. C., & Whitmer, C. A. (1948). *Crystal rectifiers* (No. Vol. 15). McGraw-Hill.

Tufekci, Z. (2019). The real reason fans hate the last season of Game of Thrones. *Scientific American*, *17*.

Van Dormael, A. (2004). The "French" Transistor. *Proceedings of the 2004 IEEE Conference on the History of Electronics*.

Van Dormael, A. (2009). Biographies: Herbert F. Mataré. *IEEE Annals of the History of Computing*, *31*, 68–73.

Welker, H. (1973). Halbleiteranordnung zur kapazitiven Steuerung von Stroemen in einem Halbleiterkristall. *Germany Patent No. DE980084C*.

Wilson, A. H. (1931). The theory of electronic semi-conductors. *Proceedings of the Royal Society of London. Series A, Containing Papers of a Mathematical and Physical Character*, *133*, 458–491.

Wolff, M. F. (1983). Mervin J. Kelly: Manager and motivator. *IEEE Spectrum*, *20*, 71–75.

Woodcock, B. A. (2014). "The scientific method" as myth and ideal. *Science & Education*, *23*, 2069–2093.