Celebrating 65th Anniversary of the Transistor

Goce L. Arsov

Abstract—The paper is dedicated to the 65th anniversary of the invention of the revolutionary electronic component that actually changed our way of life—the transistor. It recounts the key historical moments leading up to the invention of the first semiconductor active component in 1947. The meaning of the blend “transistor” is explained using the memorandum issued by Bell Telephone Laboratories. Certain problems appeared in the engineering phase of the transistor development and the new components obtained as a result of this research are reviewed. The impact of this invention on the development of power electronics is being emphasized. Finally, the possibility that the most important invention of the 20th century has been conceived not once but twice is discussed.

Index Terms—history, semiconductor, transistor.

I. INTRODUCTION

The invention of the transistor about 65 years ago was one of the most important discoveries of the 20th century. It has had enormous impact on our entire way of life. Officially it was invented by a group of scientists at Bell Telephone Laboratories (BTL), led by Shockley, on December 16, 1947, and it was demonstrated on Christmas Eve 1947. Yet, the announcement of this invention was delayed until June 1948.

For this invention Bardeen, Shockley and Brattain were awarded the Nobel Prize in Physics in 1956, "for their researches [sic] on semiconductors and their discovery of the transistor effect".

However, it should be mentioned that the pursuit of the solid-state amplifier has longer history than the transistor. It dates back to 1924–1925 and the work of Julius Edgar Lilienfeld [1]–[3], as well as to the work of Russell Ohl and his invention of the p-n junction, in 1940 [4], [5].

This invention also had great impact on the development of power electronics as a field born in 1901, this year celebrating its 112th anniversary.

Additionally, it may be interesting to note that 1947 was the centennial of A. G. Bell's birth, as well as the 50th anniversary of the discovery of the electron by J. J. Thomson, and of the publication of J. C. Bose's paper "On the Selective Conductivity Exhibited by Certain Polarizing Substances" [6], which directly led to the invention of the first semiconductor diode detector.

Although one could say that the electronics age started with Braun’s cathode ray tube (1897) and Fleming’s vacuum tube rectifier (1904), the real electronics era began with Lee de Forest’s triode, where, by placing a wire “grid” between the cathode and anode, he transformed Fleming’s rectifier into an amplifier. With the amplification, radio communications and long-distance telephony became a reality.

Despite the fact that the triode was very successful, its long-term limitations became obvious. It was a fragile device that required a lot of power. So in the mid-1920s, J. E. Lilienfeld, occupied by radio technology, began his work on finding a solid-state replacement for the vacuum triode.

II. THE INVENTION OF THE TRANSISTOR AT BTL

Aware of the limitations of vacuum tubes, several groups of scientists in different parts of the world began examining the possibility of developing semiconductor devices.

In the mid-1940s a research group for investigating the aspects of possible semiconductor applications was established at BTL. The leaders of this group were W. Shockley and S. Morgan. Two key members of the group were J. Bardeen and W. H. Brattain. Other members included G. Pearson, B. Moore, and R. Gibney. The group members had already had good theoretical and practical experience with the existing semiconductor materials and devices.

At that time, all attempts to create the semiconductor amplifying device were based on the field-effect principles described by Lilienfeld [1]–[3].

Before and after World War II, Shockley, aware of these principles, studied and analyzed different field-effect structures. He concluded that this effect should produce amplification in feasible structures (Fig. 1).

In January 1946, the group adopted two very important decisions.

The first was to focus the group’s attention only on crystals of silicon and germanium and to ignore other, more complex, materials that were frequently considered in previous investigations. At that time it was already assumed that, for most applications, silicon should be a better transistor material than germanium. This is mainly because of the higher energy gap of silicon—1.1 eV compared to 0.67 eV for germanium.

The second decision was to pursue the field-effect principle as the most promising for leading to realization of a useful device.
Between the mid-1920s and the early 1950s numerous attempts to achieve the field-effect in semiconductors were made, but all failed. Bardeen and Brattain tried confirming this theory by experimenting with metal probes placed on the germanium surface. They showed that the theory was correct. So, for the first time, they came to some understanding of the reasons they had failed to observe the field-effect, but also to some knowledge of how to intervene. After several attempts Bardeen replaced the two contacts with an ingenious arrangement of two strips of gold foil separated by just a few millimeters that were pressed onto the germanium surface. In the case when one gold contact was forward-biased and the other reverse-biased, he observed power gain. That was the discovery of the transistor action [8] (Fig. 2). This happened on December 16, 1947, nearly two and a half years after the establishment of the Shockley group.

The transistor effect was demonstrated for the top management of Bell Labs by Brattain and Moore on Christmas Eve, 1947. (The original structure and the inventors are shown in Fig. 3.) However, the announcement of the invention was postponed until June 1948. This six-month period was used to gain better understanding of the device and its possible applications, but also to obtain adequate patent position (Fig. 4) [9].

Bardeen and Brattain directed their attention to the surface effect [10] and continued experiments on that basis. Shockley, on the other hand, recognizing the role of minority carriers, worked on formulating a p-n junction theory that would include the role played by the injection of minority carriers in the case of a forward-biased junction and their collection when it was reverse-biased [11]. This research and analysis resulted with the invention of a junction transistor.

In an attempt to increase the gain, Shockley proposed a structure with a fourth region obtained by the addition of an extra junction at the collector. This was sometimes called a “hook collector” [12]. He noted that this would lead to a current gain which is greater than unity. It was later discovered that silicon p-n-p-n diode had a bistable characteristic, behaving like a reverse-biased junction in one state and a forward-biased junction in the second state, which led to the invention of the “thyristor” in 1957.

III. NAMING THE TRANSISTOR

In all dictionaries and textbooks, the word transistor is explained as a blended form of the words “transfer” and “resistor”. But, the internal “Memorandum for File” of Bell Telephone Laboratories concerning the name of the newly-invented device, “semiconductor triode”, issued on 28 May 1948 [13], [14], clearly shows that the word transistor is a blended form of the words “transconductance” or “transfer” and “varistor”. This memorandum, together with a ballot, was sent to 26 individuals or groups, including
Fig. 5. The copy of the "Memorandum for File" concerning the name for the transistor ("semiconductor triode") [13].

Shockley, Brattain and Bardeen. The propositions: "Semiconductor Triode", "Solid Triode", "Surface States Triode", "Crystal Triode", "Iotatron" and "Transistor", were all considered (Fig. 5). Although "Semiconductor triode" was recommended by the creators of this memorandum, "Transistor" was the clear winner of the internal poll. The rationale for the name was: "Transistor is an abbreviated combination of the words 'transconductance' or 'transfer', and 'varistor'. The device logically belongs in the varistor family, and has the transconductance of transfer impedance of a device having gain, so that this combination is descriptive" [13].

IV. THE DEVELOPMENT AND PRODUCTION PHASE

The invention of the transistor itself did not resolve the problems of its production. The challenge was to find ways to design a product that could be manufactured and that could succeed on the market. Eight additional years were needed to fulfill this engineering task.

In early 1950s there were two transistor structures that were proven to work—the point-contact transistor and the junction transistor—but neither of them was suitable for large-scale manufacturing. The former was difficult to make and its electrical characteristics were far from ideal. It was very variable, hard to control, and inherently unstable, but nevertheless it was manufactured for ten years, starting from 1951. Point-contact transistors were used in telephone oscillators, hearing aids, an automatic telephone routing device, and the first airborne digital computer (TRADIC).

The latter, on the other hand, had predictable and more desirable electrical characteristics. But, it was wasteful in its use of precious semiconductor material and required very special contacting techniques. Crystals were grown using a precise doping procedure to create a single thin layer of base material embedded in material of the opposite type from which the "emitter" and "collector" were made. It was possible to "grow" only one "slice" of base material in one
crystal. The process was very complex; it was conducted in special laboratories and was not convenient to automatization. The manufacturing of the “grown” junction transistor began in 1952. It was the same year that the development of the alloy junction transistor was announced at GE by J. E. Saby [15].

The alloy device was the first junction transistor to be easily manufactured and it was the leading semiconductor product of the industry for a number of years (Fig. 6). On the other hand, the research on the alloy junction process had an interesting byproduct—the junction field-effect transistor (FET). In fact, in 1951 Shockley “reinvented” the field-effect transistor, but this time as a junction device. So, in 1952 he asked G. C. Dacey and I. M. Ross to try and build a unipolar field-effect transistor. The structure built (Fig. 7) conducted exactly as Shockley’s theory had propounded [16]. Unfortunately, at that time they could not find any significant advantage of the new device over the bipolar transistor. Moreover, it was much more difficult to produce unipolar than bipolar transistors. So, although the field-effect theory was validated, its production was abandoned for a while.

In 1952 it was recognized that dopants could be introduced to very precise depths by diffusion from the semiconductor surface. The diffusion process could achieve precise control of surface concentration over a range of 10,000 to 1, and control of the number of atoms introduced over a range of 100,000 to 1. But, most importantly, the depths of diffused layers were easily controllable in the range from a fraction of a micrometer to 20μm [17].

Another problem to solve was the reliability of the transistor. Different methods were investigated. In an attempt to solve this problem, M. M. Atalla and D. Kahng studied the surface properties of silicon in the presence of a silicon dioxide layer. A byproduct of their investigations was the fabrication of another design of a field-effect device—the metal-oxide-semiconductor field-effect transistor (MOSFET) [18].

Since the beginning of the semiconductor era in electronics, the number of transistors in an integrated circuit has been increasing exponentially with time [19]. The bipolar technology has been largely replaced with CMOS and it has been improved many times, but above all it has been miniaturized to an inconceivable level. Also, the number of transistors produced per year has been growing incessantly [20].

**Fig. 6. Schematic of the alloy transistor.**

**Fig. 7.** (a) Schematic of the junction field-effect transistor. (b) G. C. Dacey and I. M. Ross (seated) testing a field-effect transistor.

### V. THE IMPACT ON POWER ELECTRONICS [21]

The invention of the transistor had a major impact on power electronics and its development.

Officially, power electronics was born in 1901 with the invention of the glass-bulb mercury-arc rectifier by Peter Cooper Hewitt of USA [22]. Then, it went through the eras of gas tube electronics in the 1930s and saturable core magnetic amplifiers in the 1940s. Hot cathode thyatron was introduced in 1926 and the ignitron rectifier in 1933. The applications of power electronics began to expand and in 1930 the New York Subway installed a grid-controlled mercury-arc rectifier (3 MW) for DC drive. Then, in 1931 the German railways introduced a mercury-arc cycloconverter for universal motor traction drive. In 1934 the thyatron cycloconverter (synchronous motor 400 hp) was installed in the Logan Power Station as a first variable frequency drive.

The present era of solid-state power electronics started with the introduction of the thyristor, or silicon controlled rectifier (SCR). Actually, the Silicon Controlled Rectifier or Thyristor was proposed by William Shockley in 1950. It was theoretically described in several papers by J.J. Ebers [23], and especially by J. L. Moll [24] and others at Bell Telephone Laboratories (BTL). In 1956 the SCR was developed by power engineers at General Electric (G.E.) led by Gordon Hall. The commercial version was developed 55 years ago, in 1958, by G.E.’s Frank W. “Bill” Gutzwiller (Fig. 8).
Fig. 8. Frank “Bill” Gutzwiller.

The idea of the p-n-p-n switch was first simulated by a circuit model, the Ebers’ model [23] (Fig. 9). The principle of operation of a proposed p-n-p-n switch was that a p-n-p transistor (bottom) is driving an n-p-n (top) and, in turn, the n-p-n is driving the p-n-p. The collector of the one, either one, drives the base of the other. That produces a positive feedback that is guaranteed to yield instability. When the voltage from A to K (A+ to B-) reaches avalanche breakdown of the “n-p” diode (between $R_1$ and $R_2$ in Fig. 9) and sufficient current flows in emitter shunt resistors $R_1$ and $R_2$ to bias “on” the emitters, the sum of $\alpha_{PnP}$ and $\alpha_{NPN}$ approaches unity, and to maintain current continuity, switching to low voltage occurs. The two collectors switch from reverse to forward voltage, and to the “on” state of the A-K switch, which, of course, is still not a p-n-p-n switch in a single “slab” of Si.

Fig. 9. Circuit model of a p-n-p-n switch [23].

The first prototype of an actual “Silicon Controlled Rectifier” (SCR) was demonstrated in 1957. The SCR became a huge success for General Electric. It was reported in the press by Business Week in their December 28th issue, under the headline “New Way to Change AC to DC”. Commercial SCRs were on the market in early 1958 [25] and Bill Gutzwiller was responsible for their technical and promotional support.

Nowadays, the SCR family consists of: SCR – Thyristor; ASCR – Asymmetrical SCR; RCT – Reverse Conduction Thyristor; LASC R – Light Activated SCR (or LTT – Light Triggered Thyristor); DIAC & SIDAC – both forms of trigger devices; BOD – Breakover Diode or Diode Thyristor (a gateless thyristor triggered by avalanche current); TRIAC – Triode for Alternating Current (a bidirectional switching device); BRT – Base Resistance Controlled Thyristor; SITh – Static Induction Thyristor (or FCTh – Field Controlled Thyristor) containing a gate structure that can shut down anode current flow; LASS – Light Activated Semiconducting Switch; GTO – Gate Turn-Off Thyristor; MCT – MOS Controlled Thyristor (has two additional FET structures for on/off control); IGCT – Integrated Gate Commutated Thyristor; ETO – Emitter Turn-Off Thyristor. The last four are thyristors with forced turn-off capabilities.

VI. DISCUSSION AND COMMENTS

An interesting question to be discussed is: who really invented the transistor? This question can be answered in different ways. There were many preceding discoveries (starting with the isolation of “silicium” by Jakob Berzelius in 1824) and inventions that were used in the development of the transistor by the BTL group. The work of J.C Bose [6], Pickard [26], Adams [27] etc. should be seriously considered.

The work of Russell Ohl and his discovery of the p-n junction and the photovoltaic effect about 73 years ago [4], [5] during the investigation of silicon rectifiers used in radar detectors, should also be given credit. Although accidental, the discovery of the p-n junction was one of the critical inventions for the further development of semiconductor electronics.

Another experimenter of this era who deserves far greater credit is Dr Julius Lilienfeld from Germany, who, in the early 1930s, patented the concept of a field-effect transistor (FET). He believed that applying voltage to a poorly conducting material would change its conductivity and thus amplification can be achieved. Lilienfeld is justly renowned for his work on the electrolytic capacitor, but according to his patents [1]-[3] he should also be recognized for his pioneering work on semiconductors. He created his non-tube device around 1923, with one foot in Canada and the other in the USA. The date of his Canadian patent applications was October 1925. These patents were followed by American ones, which should have been familiar to the Bell Labs patent office. Lilienfeld demonstrated his remarkable tubeless radio receiver on many occasions. He followed his 1925 (Canadian) and 1926 (American) patent applications for a “Method and Apparatus for controlling Electric Currents” with another granted in 1933. US patent 1,900,018 [2] clearly describes the field-effect transistor, constructing it using thin film deposition techniques and using dimensions that became normal when the metal-oxide FET was indeed mass-produced well over 30 years later. The patent (and subsequent ones) describes the advantages of the device over “cumbersome vacuum tubes”.

The next question here is how should be evaluated the “Transistron” invented by Herbert Mataré and Heinrich Welker (Fig. 10) in early 1948, while working at...
Westinghouse Subsidiary in Paris [28]. The first germanium devices with a negative Hall coefficient capable of amplification were developed in July of 1948 (Fig. 11). The diode and transistor production was based on the ceramic holder design. Manufacturing started immediately. The complete production was sold to the French telecommunications laboratories and the military.

On April 21, 1948, Mataré applied for a patent in the United States, which was granted on May 8, 1951. On August 13, 1948, Westinghouse applied for a French patent that was granted on March 26, 1952.

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.

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 fact that they completely failed to acknowledge the pioneer work done by others could be explained in different ways. It is perfectly true that the world was not ready for the previous incarnations of the transistor, but that was no reason to deny the fact that Lilienfeld patented the original solid-state triode oscillator/amplifier well before others claimed all the credit.

Moreover, the most important “re-invention” of the 20th century remarkably occurred twice—and independently. On account of the secrecy surrounding the Bell Labs device until late June 1948, it is highly unlikely that Mataré and Welker knew anything about it before July 1948. And it seems clear from the still-sketchy historical record that they had really had a working, reliable device by that time.

This dual, nearly simultaneous breakthrough may be attributed in part to the tremendous wartime advances in purifying silicon and, in particular, germanium. In both cases, germanium played the key role, as in the immediate postwar years it could be refined much more easily and with considerably higher purity than silicon. Such high-purity semiconductor material was absolutely necessary for fabricating the first transistors. But the Bell Labs team had a clear advantage with their better physical understanding of how the electrons and holes were flowing inside germanium. This proved crucial to the following achievements. The first among them was Shockley’s junction transistor, which was much easier to manufacture with much higher reliability and homogeneity.

However, regardless of who invented the transistor, the BTL group initiated the semiconductor revolution, which has completely changed our way of life. Many new semiconductor components followed the BJT: FETs, MOSFETs, thyristors and the whole family of power electronics devices, integrated circuits etc. They have enabled the design and production of many different types of electric and electronic equipment that have become indispensable in contemporary life.

VII. **In Lieu of a Conclusion**

Since the beginning of semiconductor electronics the number of transistors in an integrated circuit has been doubling approximately every 18 months. This exponential trend was first described by Gordon Moore [19] and has become known as Moore’s Law. In Fig. 12 the number of transistors in successive Intel processors is plotted as a function of time [29].

![Fig. 12. Number of transistors in successive Intel processors as a function of time [29].](image-url)
The capabilities of many digital electronic devices are strongly linked to Moore's Law: processing speed, memory capacity, sensors and even the number and size of pixels in digital cameras. All of these have been improving at (roughly) exponential rates as well. This exponential improvement has dramatically enhanced the impact of digital electronics on nearly every segment of the world economy.

Although nowadays over 90 percent of integrated circuits are manufactured in CMOS technology, Moore’s Law is still true in many aspects of the development trends of semiconductor microelectronics (with the appropriate time constant). The MOS transistor has been improved countless times but it has also been miniaturized exceeding all expectations.

The reduction in feature size, as shown in Fig. 13, has been more or less exponential. As a matter of fact, the current technology allows solid matter to be manipulated at the molecular and atomic level. So, the production and use of nano-scaled systems has become feasible.

It should also be noted that the number of transistors produced per year and the average cost, shown as a function of time in Fig. 14, have been changing exponentially, as well. In fact, in 2010, more than a billion transistors were produced for every person living on the Earth.

But, it does not end here. Nowadays, extensive research is being carried out in different areas (materials and devices): graphene, organic electronics, quantum devices, microsystems, integration of silicon with other materials and many other issues. Thus, many improvements and new discoveries are to be expected in the very near future.

Therefore, there is no need to discuss the importance of the invention of the transistor for the development of electronics in the last 65 years. It is quite obvious from the fact that electronic equipment nowadays permeates all aspects of our lives.

So, in lieu of a conclusion, the words of Robert Wallace, said at a BTL meeting on the problems in emulating the vacuum tube, are given here [32]: “Gentlemen, you’ve got it all wrong! The advantage of the transistor is that it is inherently a small size and low power device. This means that you can pack a large number of them in a small space without excessive heat generation and achieve low propagation delays. And that’s what we need for logic applications. The significance of the transistor is not that it can replace the tube but that it can do things that the vacuum tube could never do!”

ACKNOWLEDGMENT

The author wishes to express his sincere gratitude to Prof. Vladimir Katić from the University of Novi Sad, Serbia, for his very helpful comments and suggestions.

REFERENCES

[31] The singularity is near, http://singularity.com/charts