The Complete Semiconductor Transistor and Its Incomplete Forms*

Jie Binbin(揭斌斌)[1,†] and Sah Chih-Tang(薩支唐)[1,2,3,†]

(1 Peking University, Beijing 100871, China)
(2 University of Florida, Gainesville, Florida 32605, USA)
(3 Chinese Academy of Sciences, Foreign Member, Beijing 100864, China)

Abstract: This paper describes the definition of the complete transistor. For semiconductor devices, the complete transistor is always bipolar, namely, its electrical characteristics contain both electron and hole currents controlled by their spatial charge distributions. Partially complete or incomplete transistors, via coined names or/and designed physical geometries, included the 1949 Shockley p/n junction transistor (later called Bipolar Junction Transistor, BJT), the 1952 Shockley unipolar 'field-effect' transistor (FET, later called the p/n Junction Gate FET or JGFET), as well as the field-effect transistors introduced by later investigators. Similarities between the surface-channel MOS-gate FET (MOSFET) and the volume-channel BJT are illustrated. The bipolar currents, identified by us in a recent nanometer FET with 2-MOS-gates on thin and nearly pure silicon base, led us to the recognition of the physical makeup and electrical current and charge compositions of a complete transistor and its extension to other three or more terminal signal processing devices, and also the importance of the terminal contacts.

Key words: bipolar field-effect transistor; bipolar junction transistor; complete transistor; incomplete transistors; electromechanical transistors; biochemical transistors

DOI: 10.1088/1674-4926/30/6/061001 **PACC:** 7340Q **EEACC:** 2560S; 2560B

1. Introduction

The concept of the complete transistor was introduced by us[1] in the February 2009 issue of this series of invited reports on the theory of the recently discovered bipolar operation mode of the semiconductor Field-Effect Transistor (BiFET). Because of the presence of both electrons and holes in semiconductors, their electrical charge distributions and their electrical currents from the transport (or movement, by drift and diffusion) of their electrical charges, could require six terminals or contacts to the outside world, to meet the concept of a complete transistor. These are the three terminals for electrons and three terminals for holes, the input, output and reference terminals: the voltages applied to the input terminals relative to the reference terminals to control the charge and current from the electrons and from the holes, and the current or voltage at the two output terminals relative to their respective reference terminals to give the electron and hole particle current responses. It was evident that the minimum number of terminals is three, with the electron and hole charge distributions simultaneously controlled by the potential applied to one input terminal, and with the electron and hole currents present through just one output terminal, and both the input voltage and output current terminals sharing one reference terminal. The electron and hole currents can be present either simultaneously at similar amplitudes or one dominantly over the other.

The definition of the term "transistor" or "transfer resistor" was given by its originator, John R. Pierce[2], who was asked to name the newly invented solid-stated device by Bardeen and Brattain, known as the Bardeen–Brattain point-contact transistor-device[2−4]; and by Shockley, for his theoretically designed p/n junction transistor-device (later known by industry-wide consensus as the Bipolar Junction Transistor or BJT)[5,3], and for his p/n-junction gate unipolar field-effect transistor-device (later known by industry-wide consensus as the Junction-Gate Field-Effect Transistor or JGFET)[6,3]. We would like to show in this article that the Pierce definition of the "transistor" for the Bardeen-Brattain-Shockley electrical signal-amplifying and signal-processing devices can be extended and generalized to any devices; including the many, numerous solid-state devices, invented by Lilienfeld 80 years ago in the descriptions and claims of his three patents, filed and issued during 1926 to 1933 (See Ref. [3] for some figures and see his original patents for all of the many figures.); and also including the later and latest devices designed, built in research laboratories, and volume-produced in manufacturing plants, by recent engineers and scientists, that give a response, frequently electrical, to a stimulation (electrical, mechanical, optical, chemical, biological). The input stimulation is processed by the transistor, to give a response in the output of the transistor not unlike the "Central Processing Unit" or CPU of the computer.

For electrical stimulations, each of the input and output electrical signals (current or voltage) needs one signal terminal and one reference terminal. If the input and output signals share one reference terminal, a transistor could have the minimum of only three terminals, but it may not be or would not be a complete transistor because of diminishingly small magnitude or lacking of one of the two currents. For example, in a three-terminal semiconductor or semiconductor transistor,

* This investigation and Jie Binbin have been supported by the CTSAH Associates (CTSA), founded by the late Linda Su-Nan Chang Sah, in memory of her 70th anniversary. The content of this article was presented as the conference opening keynote by Jie Binbin at the Workshop on Compact Modeling on May 5, 2009 in Houston, Texas, USA.

† Corresponding authors. Email: bb_jie@msn.com and tom_sah@msn.com

Received 29 April 2009, manuscript polished > 2^4 times by the editor and two authors. © 2009 Chinese Institute of Electronics

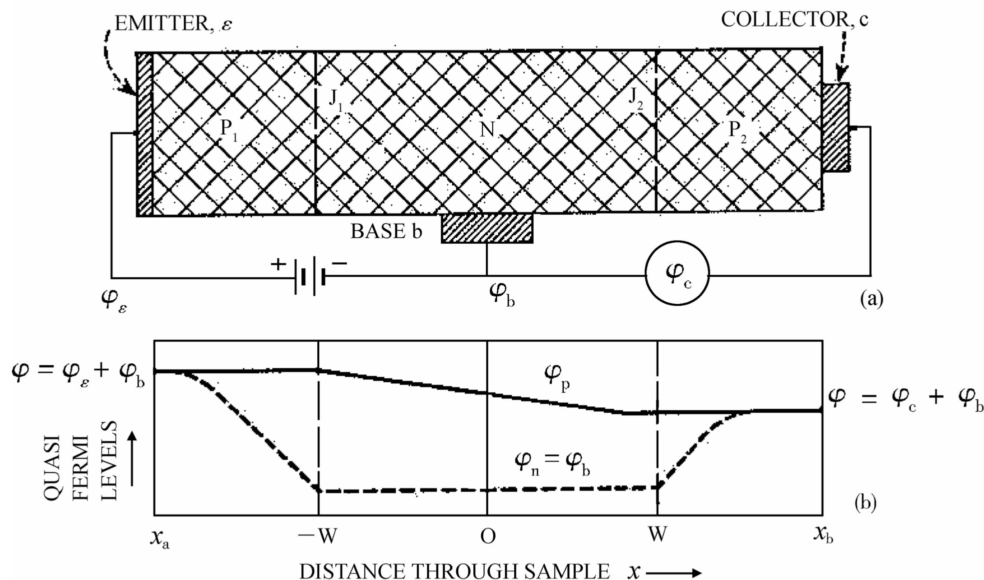

Fig. 1. The p/n junction transistor. Reprinted from Fig. 10 of the 1949 Shockley article on the invention of the Bipolar Junction Transistor (BJT)[5].

which contains two charged particle species, the electrons and the holes, the presence of only one-pair of output terminals could or would contain the current from both electrons and holes, and one would not be able to distinguish the vanishingly small hole (or electron) current from the dominantly large electron (or hole) current. In specific applications, it may be inconsequential for separately detecting the contributions from the electron and hole particle species. But to understand the working principle of a new semiconductor device, hence how to model the device, it is mandatory to observe or measure separately both the electron and hole currents in the new semiconductor device, which have the electrical signals carried simultaneously by two charge carrier species, electrons and holes, then, each charge carrier species would need to have its own input terminal, output terminal and reference terminal. It is then apparent that a complete semiconductor transistor or device would need six terminals to accept electrical stimulations and to give electrical responses from the two charge species. The complete semiconductor transistor or device is then a bipolar transistor or device, never unipolar, because of the presence of two charge carrier species in semiconductors, the electrons and holes. The term, unipolar, used by Shockley in his 1952 invention article titled "Unipolar 'Field-Effect' Transistor"[6], now known as the JGFET, would then be inadequate since the term 'unipolar' presupposes the absence of the second charge carrier species, which in fact is important or even dominant in one of the modes under the entire operation cycle of the JGFET, as we shall demonstrate later.

Similarly, the name BJT or Bipolar Junction Transistor, coined by later investigators (from Stanford University) over the reservation and concern of Shockley[7], for Shockley's 1949 p/n junction transistor[5], was indeed a redundancy in name rather than an uniqueness in its electrical operation, because its bipolarity operation is shared by all field-effect transistors, obviously recognized by Shockely on account of his concerns[7].

In this report, we will discuss in details the similarities between the many currently application-important semiconductor transistors. These comparisons include Shockley's 1949 p/n-junction (bipolar) transistor[5], Shockley's 1952 (p/n junction-gate) "unipolar" field effect transistor[6], and the many later field-effect transistors, as well as the original Lilienfeld transistors, which were all field-effect transistors[3]. All of these transistor devices can be viewed as incomplete forms of the complete 6-terminal transistors, i.e. they are short of one to three terminals. These include the "bulk" MOS transistor, Silicon On Insulator (SiO_2) FET or SOI-FET, Thin (semiconductor) Film Transistors (TFT or TF-FET), and the 2-MOS-Gate on thin Silicon FET (the FinFET). They have one metal gate terminal on insulated thin or thick semiconductor base, or they have two gates on insulated thin semiconductor base, but none has two sets of contacts to the body or base, one set each needed for the electron and hole currents and charges.

2. The Two Shockley Transistor Inventions

Shockley invented the p/n junction transistor now called Bipolar Junction Transistor (BJT) which was described in his 1949 article published in the Bell System Technical Journal[5,2]. Figure 1 is copied from Fig. 10 of the 1949 article[5], which was given by Shockley to show the working principle of this p/n junction transistor. The bipolar nature of the operation of this p/n junction transistor will be discussed in next section. Consideration of the properties of contacts, which are either sources and sinks of electrons or holes, will be given to show their dominance, especially at high densities,

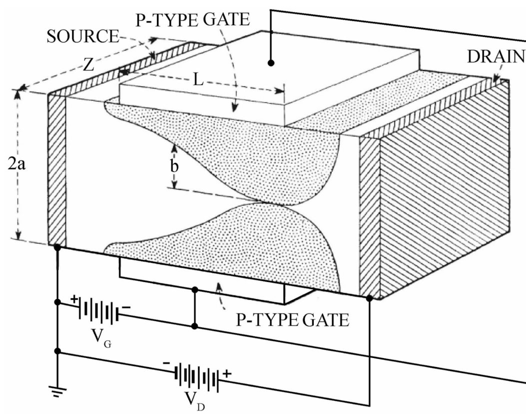

Fig. 2. The electrical pinch-off of the physical volume channel of the Junction-Gate Field Effect Transistor. Reprinted from Fig. 1 of the 1953 paper on the first reduction to practice of the JGFET by Dacey & Ross[8].

on both the spatial distributions of the spatial densities of the electron and hole currents and charges, and also the magnitudes of the electron and hole current densities.

Shockley also invented the 'unipolar' "Field-Effect" transistor or the Junction-Gate Field-Effect transistor (JGFET), which was disclosed in his 1952 paper published in Proceedings of the Institute of Radio Engineering[6]. Figure 2 is copied from Fig. 1 of the Dacey and Ross experimental paper[8], rather than the original figures in Shockley article, in order to show the spatial pinch-off of the volume channel of the JGFET. Shockley's use of the word "unipolar" in the title of his 1952 article[6] omitted one of the two currents in this JGFET, which is dominant in some of its operation modes or DC bias configurations, which will be discussed in Section 5.

3. Similarities between the Bulk MOS Transistor and the BJT

We shall illustrate the similarities between two transistors: (1) the world's most manufactured bulk MOS transistor (MOSFET or MOST) which is the one-insulated-gate FET with one metal-on-oxide or metal/oxide (Al/SiO_2) gate or metal on heavily-doped-silicon on oxide (Al/p+Si or Al/n+Si or p+Si/SiO_2) polycrystalline Si gate on a semi-infinite-thick silicon semiconductor base or body[3], and (2) the original and simple three-layer p/n junction transistor illustrated by Shockley in his 1949 invention article of the Bipolar p/n Junction Transistor (BJT)[5]. To illustrate graphically, for at one glance (A1G) acquisition by the readers[9], we shall compare the nMOST with the npn-BJT. nMOST is a MOS transistor with n-inversion surface channel on p-type Silicon base-body-or-substrate. We use the npn-BJT, rather than the original pnp-BJT used by Shockley in his 1949 article that was copied to our Fig. 1 in this article. We shall also separately illustrate their similarities at low current levels and at high current levels, in the two following sub-sections, in order to ease attaining A1G if the current level is not divided into two ranges.

3.1. Low Current Level

Figure 3 shows the similarities between the surface-conduction-channel bulk MOS transistor and the volume-conduction-channel BJT at low injection level. Figure 3(a) shows a simplified cross section of a bulk nMOS Transistor (n-inversion surface channel on p-Body). The p+ body contact or hole contact is drawn not at the bottom of the p-Body, p-Base, p-Bulk or p-substrate but at the left side of the body (base, bulk or substrate), in order to show that there is no p+ contact or hole contact at the right side of the p-Base. Therefore, although the p-Base volume channel exists, there is no hole current flowing in this volume channel due to the lack of a second contact for the hole current. We neglect the leakage current paths from the electron contacts, gate insulator and interface traps to simplify our illustration, so as not to be confused by these secondary or higher-order current pathways. The electrochemical potential representation[10, 5, 2] readily shows that because there is no steady-state hole current, the hole electrochemical potential in the base volume channel does not change or is spatially constant along the base length direction, and it can be set by the voltage applied to the p+Body (p+B) contact. When the p+ hole contact is located at the bottom edge of the base, the hole electrochemical potential in the p-Base volume channel is forced to be equal to the value in the extended p+ region of the extended source-to-drain-length body contact. It is then easy to see that the p+ body contact at the right side of the base is equivalent to the p+ body contact at the bottom of the base. This immediately gives an alternative for deriving the X-equation of the bulk MOS transistor[10]. To derive the X-equation from the Poisson Equation, the semi-infinite-thick base of the bulk MOS transistor was employed in Ref. [10] to give the boundary condition of spatially constant or zero electrical potential along the back-side surface of the base, which is far away from the silicon/silicon dioxide interface. This constant electrical potential boundary condition is now replaced by the constant electrochemical potential boundary condition without implication of semi-infinite base thickness. The current path via recombination at the interface traps is explicitly shown in this figure, Fig. 3(a). The many and spatially distributed interface traps (interfacial recombination-generation-trapping centers or GRT centers) are represented by just one open triangle, with the understanding that they are distributed spatially over the entire length and width of the interface area of the SiO_2/Si interface. Interface GRT centers are important on transistor's reliability or operation life, and noise figure-of-merits in RF applications.

The transition one-electron energy band versus distance diagram, the E–y diagram[9], is shown in Fig. 3(b). It does not show the E–y diagram in the bulk of the metal and not at the M/Semiconductor interface. Both of these can be easily included[9] to show the complete picture. Figure 3(b) is not the two-dimensional cross-section view; it is a 'schematic' **projection view** of all the electronic events from all the Y–Z layers or X-planes. It is schematic for the inheritent 2-Dimensional MOSFET, because the edges of the energy gap and the localized energy levels are shifted in X due to the gate–voltage produced normal electric field or the X-dependence of the electric potential, which renders the 2-D projection view untenable or at best highly cluttered and confused. This is not the case in the 1-D BJT where uniformity in the x-direction is

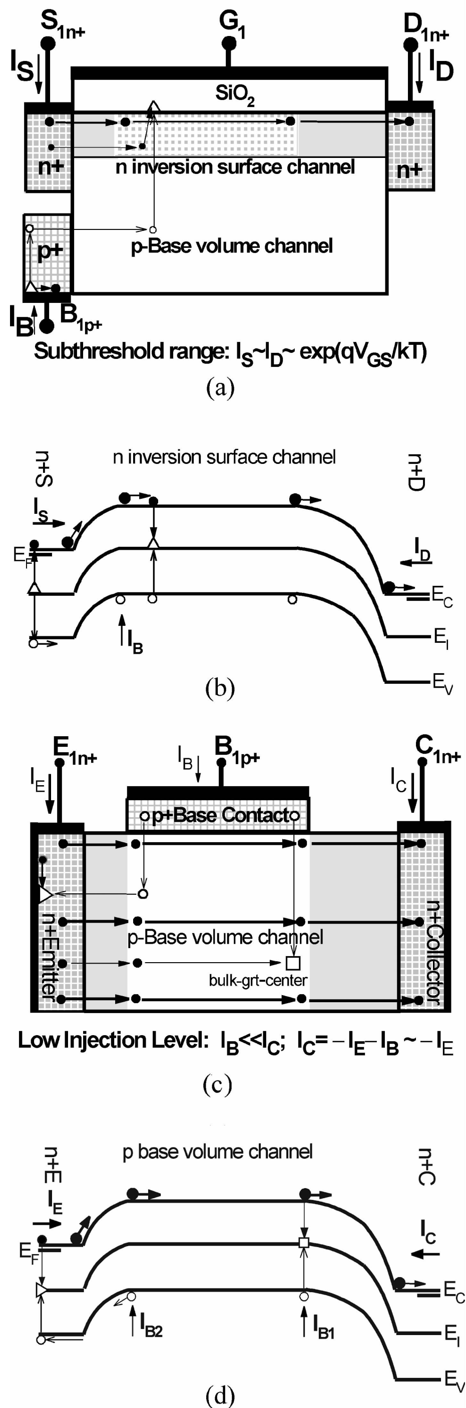

Fig. 3. Similarity at low current levels between the n-inversion surface-conduction-channel semi-infinite-thick bulk MOS transistor, the nMOST, and the volume-conduction-channel npnBJT. (a) Cross-sectional view of the nMOST operating in the subthreshold current range. (b) Schematic projection view of the transition energy band diagram, *E*–*y*, in the direction of the surface inversion channel, *y*-axis, of the nMOST. (c) Cross-sectional view of the npnBJT. (d) Projection view with no variation in the *x*-direction, of the *E*–*y* diagram along the p-type volume-conduction-channel, *y*-axis, of the npnBJT.

assumed. Thus, our projection views for the MOSTs are understood to be schematic with *X* location of the *Y*–*Z* layers at some given *x*-plane, such as the planar SiO_2/Si or a body plane. In the physical cross sectional view shown in Fig. 3(a), the *Y*-axis is in the horizontal direction and the *X*-axis is in the downward vertical direction. In the *E*–*y* one-electron transition energy-band diagram of Fig. 3(b), we show the trajectory or pathway of the electrons from the n+Source which are moved or injected over the build-in potential barrier into the n-inversion surface channel by diffusion due to the concentration gradient (from a very high concentration in the n+S to a very low concentration in the n-inversion-surface-channel). A small fraction of the electrons, during and after injection, will recombine with holes, at the SiO_2/Si interfacial GRT centers (the hollow triangle on the SiO_2 side of the SiO_2/Si interface). The holes come from the p+Base through the p-Base volume channel. These p+Base holes are replenished by the thermally emitted holes from the GRT interfacial centers at the p+Base/Metal contact, shown also by just one triangle, while the thermally emitted electrons move into the metal contact and metal wire to complete the electrical circuit loop. The current along this recombination pathway via the Si/SiO_2 interfacial GRT centers is a part of the base or substrate current I_B, shown both in Fig. 3(a) and *E*–*y* energy band in Fig. 3(b). The other part, not shown, is the similar electron–hole recombination current at the bulk GRT centers in the p-Base volume channel, its n-inversion surface channel and its (nearly depleted of electrons and holes) surface space-charge layer. However, the density of these bulk GRT centers are very low in present and recent manufacturing technologies, therefore, they can be neglected.

Figure 3(c) is a simplified physical cross-sectional view of the p/n junction transistor. The bulk or base-layer generation-recombination-trapping or GRT centers are represented by one open square in this figure. Electrons move over the build-in potential barrier of the emitter/base junction, shown in the corresponding one-electron transition energy band diagram Fig. 3(d), reach the base quasi-neutral region and recombine at these bulk GRT centers with the holes supplied from the p+ base contact. The current along this pathway is a component of the base current I_{B1}. The other base-current component, I_{B2}, comes from the holes moving over the build-in potential barrier of the n+emitter/p-base junction into the n+ emitter region. It is crucial to understand that there are many interfacial GRT centers located at the Metal/n+Emitter interface (also M/n+Collector interface) which are represented by just one open triangle in this figure to simply the illustration. These centers are expected from Sah's interface trap model due to the random distribution of the interfacial bond angles and lengths, which was mathematically analyzed by Sah based on Slater's many-body quantum mechanical perturbation theory of the semiconductor periodic potential in the presence of randomly located atomic imperfections[11]. At the M/Semiconductor interface, there are also GRT centers from the metal impurities. Due to the heavily doped n-type emitter, the space charge layer of the Metal/n+silicon junction is so thin that the holes from the p-base region can also reach these interfacial GRT centers by tunneling through the build-in potential barrier of the Metal/n+silicon junction. Then, these holes recombine with the electrons from the Metal conduction band, as shown in Fig. 3(d). See Ref. [11] for examples that illustrate the one-electron energy-bands in the metal/material contacts. At the Metal/non-Metal contacts, the surface recom-

bination velocity or rate is almost infinite or has reached the maximum value allowed, due to the high concentration of the interfacial GRT centers and high electron concentration in the metal conduction bands. Large recombination rates lead to the ohmic contact properties of the Metal/n+ junction[9].

Let us further study the energy band pictures. Figure 3(b) shows the transition energy band diagram along the surface-conduction-channel (E–y or EY diagram) of the bulk MOS transistor operating in the subthreshold range. The resistance of the channel is relatively small compared with that of the space-charge region of the base/drain p/n or n-/n+ junction. The voltage drop along the channel, from the space charge boundary of the base/source junction to the space charge boundary of the base/drain junction, is negligible. As shown in this figure, Fig. 3(b), electrons move through the space charge region by first being injected via diffusion due to concentration gradient from the n+Source to the n-inversion-surface-channel over the build-in potential barrier of the n+Source/n-inversion-surface-channel junction, diffuse along the n-inversion surface channel, accelerate through the space charge region of the n-inversion surface-channel/n+drain junction, and reach the n+drain contact. Therefore, in this low current range, the current is limited by the diffusion of the electrons over the potential barrier at the n+S/n-inversion-surface-channel, and hence controlled by the diffusion barrier height at the n+Source. And, the diffusion barrier height at the n+Si/n-surface-inversion-channel is controlled by the amount of n-surface-inversion on the p-Base, by the voltage applied between the Gate and Source, or applied to the gate SiO_2 film, and the resulting electric field at the SiO_2/n-surface-channel interface at the Source. So, source junction barrier lowering is the fundamental mechanism that determines the electron current or the drain terminal current in the low level or subthreshold range of the current–voltage characteristics of the MOSFET.

Continuing the description of the similarities of the energy band diagrams, EY, of MOSFET and BJT at low current levels, Figure 3(d) shows the transition energy band diagram along the volume-conduction-channel of the p/n junction transistor or BJT. It is apparent that Fig. 3(d) is similar to Fig. 3(b), except the relabeling of the n+Source/p-Base junction of the nMOST by the n+Emitter/p-Base junction of the npnBJT; and the p-Base/n+Drain junction of the nMOST by the p-Base/n+Collector junction of the npnBJT. Thus, this analysis proves no physical difference on the basic parameter that controls the low level electron current in the two transistors, nMOST in the subthreshold current range and npnBJT in the low-level current range, because they are both controlled by the potential barrier height of the electron injection junction, the source n+/n-inversion-surface-channel junction of the nMOST and the emitter n+/p-base-volume-channel junction of the npnBJT. The only difference is the way the potential barrier height is controlled or modulated. In the nMOST or any MOST, the source/base barrier height at the Gate-oxide/Si-base interface is controlled or modulated by the gate oxide transverse electric field or applied voltage to the gate oxide, hence 2-Dimensional. In the npnBJT, the corresponding electron injection barrier height is controlled or modulated by the forward voltage applied to the volume-channel n+Emitter/p-Base junction of the npnBJT. An in-depth description of the GRT kinetics at the interfacial and bulk centers is omitted here at low current levels and is given in the next section at high current levels which also applies to the present low current levels.

3.2. High Current Level

High current level is the result of vanishing or vanished electric potential barrier height of the injection n/p junction in both the nMOST and the npnBJT. Consequently, the current is no longer determined by the vanished or zero potential barrier height through the transition or injection layer (the E/B and S/B transition layers.) Therefore, the current must be determined by what is left, the drift-diffusion through the base region, either in the surface inversion layer of the nMOST or the base volume layer of the npnBJT, hence it is a resistance-like current, which can have a spatially varying resistance modulated or controlled by an input signal, applied to the gate/source junction in the MOST or the base/emitter junction in the BJT.

The four parts in Fig. 4 show similarity at high injection or high current levels between the surface-conduction-channel nMOS transistor and the volume-conduction-channel npnBJT. Figure 4(a) shows a simplified cross section view of the nMOS transistor operating in the strong inversion range, which is very similar to Fig. 3(a) except the large current indicating by the thicker arrows. Similarly, the E–Y energy band diagram of Fig. 4(b) is very similar to that of the low-level injection given in Fig. 3(b), except that here the n+S/p-B or n+S/n-surface-inversion junction is wiped out by the higher applied gate/source voltage.

Similarly, Figure 4(c) shows a simplified cross section view of the npnBJT operating at high injection level. The hole current from the p+ base to the n+ emitter has significantly increased to maintain the essentially electrical neutrality, known as quasi-neutrality[5, 2, 9] in the p-Base, i. e., defined as the injected electron concentration exceeding the existing hole concentration in the p-Base from the acceptor impurities, $N > P >> P_{\rm IM}$, so $N - P = P_{\rm IM} + (\varepsilon/q)\nabla\cdot E \simeq 0$ or $P \simeq N$. So, the electron current from the n+ emitter to the n+ collector through the originally p-type Base, is almost equal to the hole current from the p+Base contact through the p-Base to the n+emitter. The current gain of the npnBJT at such a high injection level would have dropped toward 1/2, since half of the emitter current is now from holes in the p+Base recombining with the electrons in the n+Emitter, at the Metal/n+Emitter interfacial GRT centers. The trajectory arrows of the three electrons in the n+Emitter pointing to the three M/n+E interfacial GRT centers are shown in Fig. 4(c) but in the transition energy band diagram in Fig. 4(d) only one interfacial GRT center is shown with the two vertical transition arrows, the solid downward arrow for electrons and the upward open arrow for holes to avoid cluttering, although generally the interfacial GRT centers are distributed in energy in the silicon energy gap[11]. There are usually few GRT

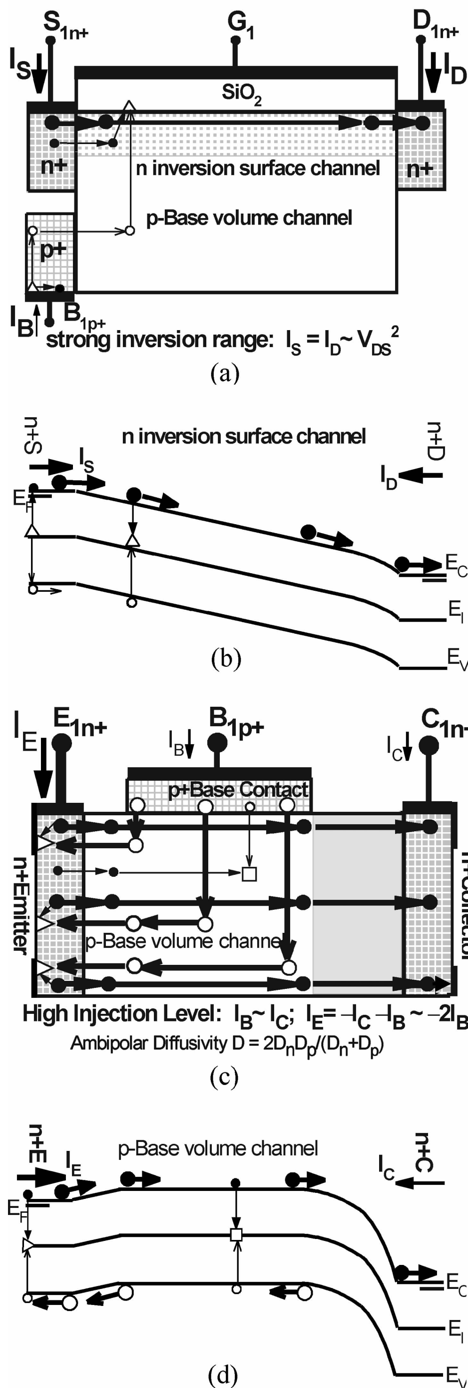

Fig. 4. Similarity at high current level between the n-inversion surface-channel MOS transistor, nMOST, and the volume-conduction-channel npnBJT. (a) Cross-sectional view of the nMOST operating in the strong inversion range. (b) The schematic projection view of the transition energy band diagram along the surface-conduction-channel, y-axis, of the nMOST. (c) Cross-section view of the npnBJT operating at the high injection level. (d) Schematic projection view of the transition energy band diagram along the volume-conduction-channel, y-axis, of the npnBJT with variation along the x-axis ignored.

centers in the n+Emitter layer and in the p-Base, the latter is shown as a single open square in Fig. 4(d) with the one vertical transition arrow each for the electrons and holes. These descriptions of the electronic transitions at the interfacial

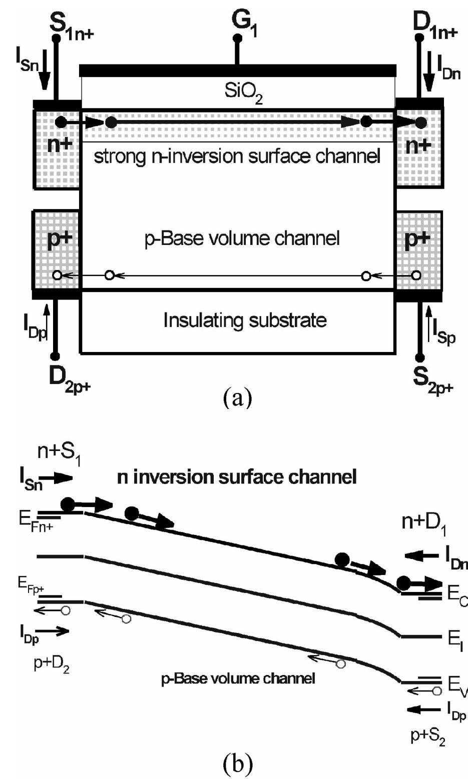

Fig. 5. (a) The cross-sectional view of the device structure and (b) the schematic projection view of the transition energy band diagram of a Thin-Film Transistor, TFT, with one electron contact and one hole contact at each of the two ends of the p-type thin-film. The transistor operated in the strong n-inversion surface channel bias mode, linking the two electron contacts, and a conducting p-Base volume channel, linking the two hole contacts.

and bulk GRT centers are also applicable to the low current levels shown in Figs. 3(a) to 3(d), but they were omitted for A1G brevity of first-time description of the importance of the GRT kinetics at the interfacial traps.

4. Bipolar Currents and Charges in TFT and FinFET

Figures 5(a) and 5(b) respectively show the simplified device-structure cross-section and transition-transport energy-band projection views of a p-base n-inversion surface-channel one-Gate (1G) Silicon or Semiconductor on Insulator (SOI) Thin-Film Transistor (nTFT), 1G-SOI-nTFT, operating in strong n-inversion surface-channel range. To illustrate the presence of both electron and hole currents, Bipolar Currents, we have included two contacts for both electrons and holes in order to provide the individual pathways or channels for surface-channel electrons and volume-channel holes. Comparing Fig. 5(b) here of the nTFT with Fig. 4(b) of the nMOST in strong inversion, we see that they are very similar. However, there could be a hole current flowing through the base volume hole-channel in this nTFT because of the two p+ hole contacts, S_{2p+} and D_{2p+}, with an applied voltage or a contact potential difference in a temperature gradient between the two p+ hole contacts. For the nMOST in Fig. 4(b), there is no hole current or negligible (secondary) hole current flowing into the p+Base

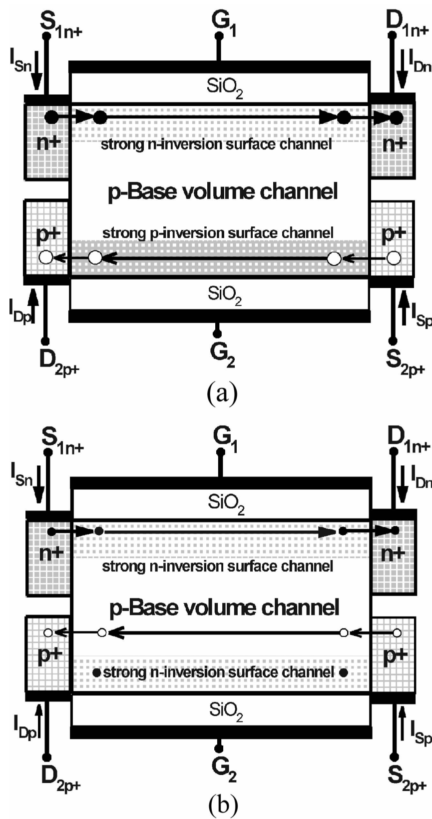

Fig. 6. Physical cross-sectional views of a FinFET with one electron contact and one hole contact at each of two ends of the p-type pure silicon base. (a) The two gate voltages has different polarity, for example, $V_{G1/n+S1} > 0$ and $V_{G2/p+S2} < 0$. (b) The two gate voltages has the same polarity, for example, $V_{G1/n+S1} > 0$ and $V_{G2/p+S2} > 0$.

contact because there is not the second p+Base contact.

The simplified cross-section views of a FinFET (2-MOS-Gate on thin p-Base with a pair of electron and a pair of hole contacts) are shown with different polarities of the two gate voltages in Fig. 6(a), and the same polarity gate voltages in Fig. 6(b). This is the complete and also more general transistor, rather than the nanometer laboratory and advanced engineering FinFETs recently reported which contain just one continuous or wrap-around MOS gate on a nanometer thin and presumable nearly pure base and only one drain and one source contact that presumably are either electron or hole contacts. When these two gates are applied different polarity voltages, as shown in Fig. 6(a), a strong n-inversion surface channel induced by the positive voltage applied to the top gate, G1, is formed which links two electron contacts, and a strong p-inversion surface channel is also formed between two hole contacts, by the negative voltage applied to the bottom gate, G2. Since the surface p-channel is much conductive than the p-base volume channel, most hole current flows through the surface p-channel.

When these two gates are applied the same polarity voltages, as shown in Fig. 6(b), two strong n-inversion surface channels are formed when the applied gate voltages are positive (relative to a reference, such as the n+S), and in addition, the hole current also flows through the p-base volume channel contacted by the two p+ hole contacts. However, electron current flows into only one n-inversion channel (under the top gate, G1, in this illustration) through the two electron contacts, and no electron current flows into the lower n-inversion surface channel due to lacking the two n+ contacts to the bottom n-inversion channel under the bottom gate, G2, in this illustration. Physical realizations of bipolar contacts as source and sink of both electrons and holes were discussed by us[1, 10, 13].

5. Similarity Between the Bulk and SOI MOS Transistors

The similarities between the bulk and SOI nMOSTs are illustrated in their cross-sectional structure views in Figs. 7(a) and 7(c) and in their projection transition energy band views in Figs. 7(b) and 7(d). Note in Fig. 7(c) of the SOI that the p+ body contact is an electrical contact to the thin p-Base **S**ilicon **o**n the buried oxide **I**nsulator. As discussed in Section 3, the bulk nMOS transistor is an incomplete transistor, even though there is a very small hole current flowing into the semi-infinitely thick base volume channel. The SOI nFET has a much thinner base volume channel than the bulk MOS transistor but still an incomplete transistor because it does not have a second p+ hole contact to complete the hole channel current loop. The SOI FET manufactured by IBM does not have the feature drawn in Fig. 7(c), instead, the buried oxide completely isolates the p-base volume channel from the p+ body. Thus, the IBM SOI nFET is even a less complete or more incomplete transistor. In fact, it does not operate about a steady-state condition due to its lack of any hole contact, rendering the depletion or enhancement of hole charges in the p-type SOI base film incapable of reaching a true DC steady state in the desired circuit operation time scale, further compromising its reproducibility and reliability. With the p+ region in contact to the p-Base of the SOI shown in Fig. 7(c), it is obvious that the bulk and SOI nMOSTs work the same way, as also illustrated by the projection view of their transition energy band diagrams shown in Figs. 7 (b) and 7(d).

6. Bipolar Currents and Charges in JGFET

In this section, we shall demonstrate that the p/n junction-gate field-effect transistor, with the volume channel conducting width modulated, which was invented and named by Shockley in the title of his 1952 article, A Unipolar 'Field-Effect' Transistor[6], is really a Bipolar field-effect transistor. Indeed it has both electron and hole currents in its four terminals. The four terminals put it in the incomplete transistor category. Although its present form does not conform to the complete transistor definition, it can. For example, if one of the p+Gate is replaced by a n+Gate/p-Base double layer, with the addition of two p+Contacts to the p-Base, then, it becomes a complete p/n junction-gate field-effect transistor with six terminals. It has two gates, and it has electron contacts to the electron volume channel and hole contacts to the hole volume channel. And, its two volume channels are adjacent to each other, each serving as the second p/n junction gate of the other. Such a two-channel complete transistor structure was actually realized (i. e. fabricated by the silicon planar transistor techno-

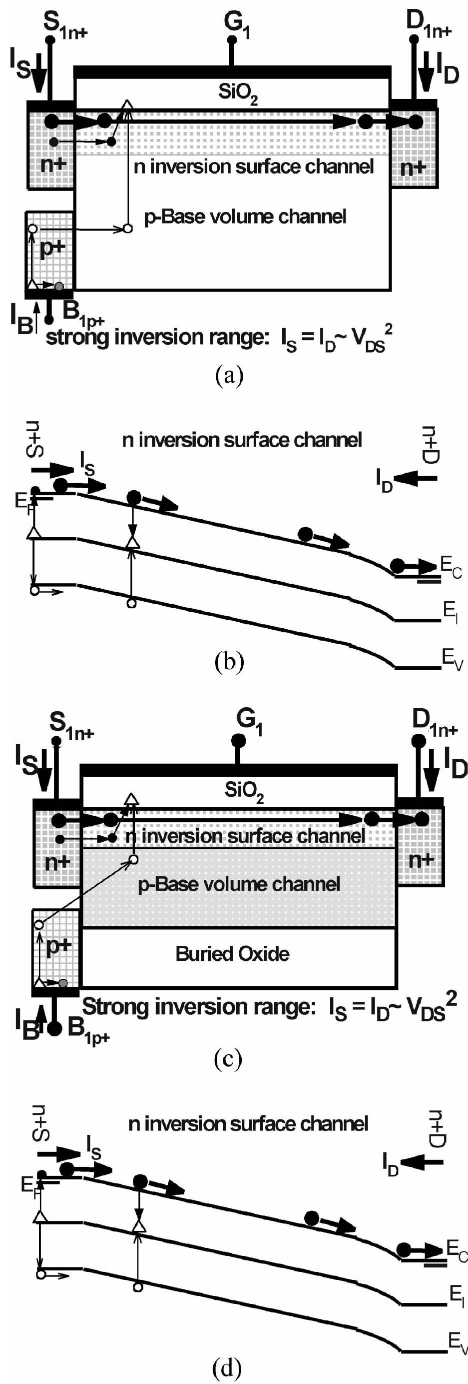

Fig. 7. Similarity between the bulk MOS transistor, nMOST, and the thin-film nSOI-FET, operating in the strong n-inversion surface channel range. The cross-sectional views of: (a) the nMOST and (c) the nSOI-FET. The schematic projected transition energy band diagrams of: (b) the nMOST and (d) the nSOI-FET.

logy invented by Hoerni[3]) in a hybrid form by Fu and Sah in 1971[12], in which, the p+Gate was replaced by a M/SiO_2/p-Si to give the hybrid-gate six-terminal M/SiO_2/p-Si/n-Si MOS-JG FET. In a more recent project to demonstrate a step function jump by a factor of 2 of the Moore's Law with existing technology, via the realization of the CMOS voltage inverter circuit in one physical FinFET transistor, both gates were replaced by the MOS gate and the body is a p/n junction-less pure silicon thin base[13].

To demonstrate the presence of both electron and hole

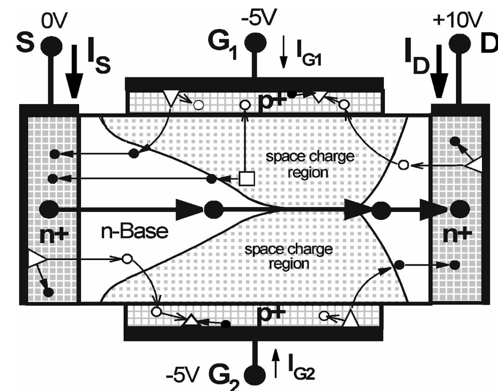

Fig. 8. Cross section view of the Junction-Gate Field-Effect Transistor (JGFET) when its n-base volume channel starts to pinch off. Electron and hole trajectories are shown to indicate the bipolar nature of the JGFET.

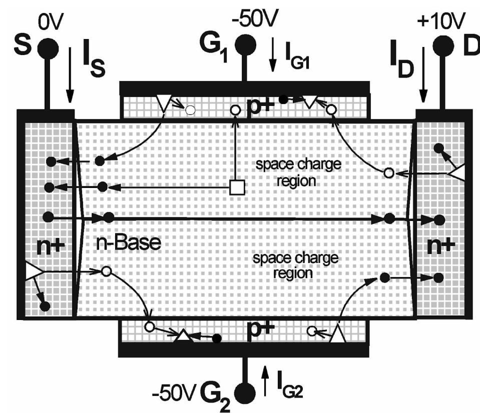

Fig. 9. Cross section view of the Junction-Gate Field-Effect Transistor (JGFET) when its n-base volume channel is completely depleted between the two gates due to very large reversed biases applied to both the source/gates and drain/gates junctions. Electron and hole trajectories are shown to indicate the bipolar nature of the JGFET.

currents or the bipolar current property of Shockley's 1952 "unipolar" Junction-Gate Field-Effect Transistor (JGFET), we employ the simplified 2-D cross-section view of the JGFET with n-type volume channel and two p+Gate, shown in Fig. 8, in place of the 3-D view of Dacey and Ross's original p-type volume channel JGFET shown in Fig. 2 in Section 2. The p+Gate/n+Source junctions and p+Gate/n+Drain junctions in Fig. 8 are both reversely biased to make the two space-charge regions contacting each other. Figure 8 (also Fig. 2) shows that the width of the conductive n-type volume channel is now reduced to zero over a fraction of its length near the n+Drain by the two carrier-depleted space-charge regions in touch near the n+Drain . This was called the pinch-off of the volume channel by Shockley[6]. At the pinch off, there is still an electron current, as indicated by the thick electron-current arrow in the mid-line or mid-plane (in the *Y*–*Z* plane) in the n-base of Fig. 8. But this volume channel current, due to electrons flowing from the n+Source through the electrically pinched-off n-Base volume channel to the n+Drain, is no longer dependent on the voltage applied between the n+Drain and n+Source, V_{DS}, which no longer changes much the electrical shape or the varying width of the n-Base volume channel, other than the slight reduction in its

length. Thus the drain terminal current, I_D, saturates to a constant value with only slight further increases of the magnitude with increasing V_{DS} (negative to reversely bias the two p+D/n+G junctions) due to the just mentioned electrical length reduction. The (electrical) channel length shortening effects is neglected in the long channel (physical channel) theory, and is treated as a first order correction to the physical channel length. Nevertheless, Figure 8 shows that there are indeed hole currents, indicated by hole-arrows, which could increase the drain and source terminal currents, which make this Shockley's "unipolar" Field-Effect transistor indeed bipolar. Figure 8 shows that there is a hole current path in each of the four reverse-biased junction regions (two n+Gate/p+Source junctions and two n+Gate/p+Drain junctions). Only one of each of the two hole current paths at the n+S and n+D is shown in Fig. 8. Similarly, the corresponding electron current paths are shown in this figure. The important point here is the unavoidable presence of the interfacial GRT centers (shown as triangles in Fig. 8) at all four contacts, the M/n+S and M/n+D interfaces to serve as the sources or the sites for the generation of the holes, and at the M/p+G_1 and M/p+G_2 interfaces to serve as the sources or sites for the generation of electrons. As an illustration, we describe one complete transport pathway of holes from n+S to p+G_2 in the reverse biased n+Source/n-Base/p+$Gate_2$ which gives the hole current in the S and G_2 terminals. Holes are generated at the GRT centers at the M/n+S interface, diffuse through the n+S and n-B regions, drift through the n-B/p+G_2 space charge region, drift through the p+G_2 to the GRT centers at the p+G_2/M interface, and then recombine with the electrons from the Metal in contact with the p+G_2. We again neglect the bulk GRT centers in the n+S region, the n-B region, the n-B/p+$G_{1,2}$ space charge regions in the entire n-Base region, and the p+$G_{1,2}$ region. Nevertheless, a partial pathways (interface generation centers not shown) of the generation current from holes and electrons generated at a bulk GRT center (open square) in the reverse-biased p+G_1/n-B space-charge region is shown in Fig. 8. As indicated earlier, bulk GRT centers are nearly eliminated in production silicon transistors, while residual interfacial GRT centers due to lattice mismatch at the SiO_2/Si, M/Si and M/SiO_2 interfaces remain.

Figure 9 shows the simplified cross section view of the above JGFET operated in the well-known current cut-off mode when the reverse biases on the G/S and G/D junctions are sufficiently large to deplete the electrons in the entire n-Base volume channel. The electron and hole pathways are self explanatory following the example just given for Fig. 8 in the preceding paragraph. This shows that both electron and hole currents are present in comparable magnitudes simultaneously in the same space regions of the transistor, therefore, it is a bipolar transistor, not the unipolar field effect transistor, which was named in the title of Shockley's 1952 article.

7. Transistors with Different Input Stimulations

In the above sections, we have discussed the traditional silicon transistors which are presently manufactured in volume for electrical signal processing, with a minimum number of electrical input and output terminals, three. For the BJT, its input signal could be the base current or voltage, and its output signal could be the collector current or voltage. It has only one base terminal which does not control the electrons and the holes separately, nor provide the second channel for the current of the second charge carrier species. Both species, electrons and holes, flow through only one channel, a volume channel. So, the BJT is an incomplete semiconductor transistor. For the MOS transistor, with the n-inversion surface channel on semi-infinite thick p-silicon body, known as the "bulk" nMOS transistor, there is only one gate, and the voltage applied to the gate simultaneously induces both the n-inversion surface channel and modulates the p-Base volume channel. The source and drain electron contacts connect the n-inversion channel and electron channel current to outside load, while the single hole contact to the p-Base provides the unlimited electrical equilibrium sources and sinks of holes to the base volume channel although it does not provide any significant hole current to outside load, unless there are two hole contacts and an applied or temperature gradient voltage appeared between them. So, the four-terminal, one-gate, one-source, one-drain and one-body contact (G, S, D, B), bulk MOS transistor is an incomplete semiconductor transistor because it carries only the current of one carrier species, although its one input gate terminal controls the charge distributions of both carrier species (electrons and holes) but not the current of the second charge carrier species, which will require a second contact for the second carrier species.

The semi-infinite-thick bulk MOS transistor structure with the minimum of four terminals (G, S, D, B) can be extended to a broad range of signal processing devices, simply by sensitizing the gate to other signals which are not electrical, such as light, temperature, pressure, atom (hydrogen) and molecule. An silicon MOS transistor with a gate which is sensitive to DNA molecules was reported[14,15] which was named the BioFET[15]. It is a bulk-silicon MOS transistor with a two-layer gate over the gate oxide, consisting of electrolyte (for electrical contact) over a DNA layer. DNA molecules contain negative charges on their sugar backbones which modify the silicon surface charge distribution under the gate oxide. Hybridization of the DNA molecules is the input stimulation which modulates surface channel charge density and the electrical current flowing between the drain and source terminals of this BioFET.

It is readily recognized by the readers that the solid-state electron device structures described in the three 80+year old Lilienfeld patents[3] are all incomplete transistors in our proposed definition of the complete transistor in this exposition.

8. Summary

We have described in this paper the model of the complete semiconductor transistor. Similarity between the supposedly unipolar semi-infinite-thick bulk MOS field-effect transistor and the supposedly bipolar p/n junction transistor, BJT, are extensively described to show that they are both bipolar transistors, with incomplete physical structures, as gauged by our proposed model of the complete semiconductor transistor.

The bulk MOS transistor is extended to many recently manufactured transistors, all with incomplete physical structures, including also a DNA sensitive gate MOS transistor known as the BioFET.

Acknowledgments

We thank Professor Wang Yang-yuan of Peking University, Academician of the Chinese Academy of Sciences, for his supports which have enabled us to undertake and complete this and previous studies, all presented in the issues of this journal. We also thank especially Professor Xing Zhou of Nanyang Technological University of Singapore for convincing us five years ago to rejoin the Compact Model Community after 40+ years of absence of the co-author (CTS), without which, the results of this and the previous reports would not have been obtained, and our future efforts on device and system modeling, would not have been planned.

References

[1] Sah Chih-Tang and Jie Binbin, "The Bipolar Field-Effect Transistor: XIII. Physical Realizations of the Transistor and Circuits (One-Two-MOS-Gates on Thin-Thick Pure-Impure Base)," Journal of Semiconductors, 30(2), 021001-1–12, February 2009.

[2] William Shockley, Electrons and Holes in Semiconductors with Applications to Transistor Electronics, 558 pp, D. Van Nostrand Company, Inc. New York, November 1950.

[3] Chih-Tang Sah, "Evolution of the MOS Transistor – From Conception to VLSI," Invited paper, IEEE Proceedings, 76(10), 1280–1326, October 1988.

[4] J. Bardeen and W. H. Brattain, "Three-electrode Circuit Element Utilizing Semiconductive Materials," U.S. Patent 2524035. Application filed June 17, 1948, granted Oct. 3, 1950.

[5] W. Shockley, "The Theory of p-n Junctions in Semiconductors and p-n Junction Transistors," Bell System Technical Journal, 28(7), 435–489, July 1949.

[6] W. Shockley, "A Unipolar 'Field Effect' Transistor," Proceedings of the IRE, 40(11), 1365–1376, Nov. 1952.

[7] Personal communication from Shockley to Tang Sah, in one of the weekly one-on-one tutorial sessions during 1956–1959.

[8] G. C. Dacey and I. M. Ross, "Unipolar 'Field Effect' Transistor," Proceedings of the IRE, 41(8), 970–979, August 1953.

[9] Chih-Tang Sah, Fundamentals of Solid State Electronics, 1001 pp, World Scientific Publishing Company, 2001. For A1G, see Preface-Pedagogy on page xii. For contact properties see Section 583 Contact Resistance on page 506 and related sections.

[10] Sah Chih-Tang and Jie Binbin, "The Theory of Field-Effect Transistors: XI. The Bipolar Electrochemical Currents (1-2-Gates on Thin-Thick Pure-Impure Base)," Journal of Semiconductors, 29(3), 397–409, March 2008.

[11] Zuhui Chen, Bin B. Jie, and Chih-Tang Sah, "Effects of Energy Distribution of Interface Traps on Recombination DC Current-Voltage Line Shape," Journal of Applied Physics, 100, 114511-1–11, 2006.

[12] H. S. Fu and C. T. Sah, "A Distributed-Gate Bistable MOS Transistor," Solid-State Electronics, 14(9), 799–804, September 1971.

[13] Jie Binbin and Sah Chih-Tang, "The Bipolar Field-Effect Transistor: VI. The CMOS Voltage Inverter Circuit (Two-MOS-Gates on Pure-Base)," Journal of Semiconductors, 29(11), 2079–2087, November 2008.

[14] E. Souteyrand, J. P. Cloarec, J. R. Martin, C. Wilson, I. Lawrence, S. Mikkelsen, and M. F. Lawrence, "Direct Detection of the Hybridization of Synthetic Homo-Oligomer DNA Sequences by Field Effect," Journal of Physical Chemistry B, 101(15), 2980–2985, 1997.

[15] M. J. Deen and M. W. Shinwari, "Modeling the Electrical Characteristics of FET-type Sensors for Biomedical Applications," Technical Proceedings of NSTI-Nanotech2007, vol. 3, Workshop on Compact Modeling, 599–602, May 2007. M. J. Deen, et. al. Monograph in the ASSET Series, World Scientific Publishing Company, 2010.