A Floating Gate and Its Application to Memory Devices

By D. KAHNG and S. M. SZE

(Manuscript received May 16, 1967)

A structure has been proposed and fabricated in which semi-permanent charge storage is possible. A floating gate is placed a small distance from an electron source. When an appropriately high field is applied through an outer gate, the floating gate charges up. The charges are stored even after the removal of the charging field due to much lower back transport probability. Stored-charge density of the order of $10^{12}/cm^2$ has been achieved and detected by a structure similar to an metal-insulator-semiconductor (MIS) field effect transistor. Such a device functions as a bistable memory with nondestructive read-out features. The memory holding time observed was longer than one hour. These preliminary results are in fair agreement with a simple analysis.

It has been recognized for some time that a field-effect device, such as that described by Shockley and Pearson,[1] can be made bistable utilizing switchable permanent displacement charges on ferroelectric material.[2] Subsequent studies of ferroelectric material have revealed,[3] however, that the inherent speed capability of a device incorporating a ferroelectric material is limited by domain motion, whose highest speed is limited by the acoustic velocity. In the absence of highly ordered, near-ideal thin film ferroelectric material, the speed capability of a bistable device, therefore, is in the microsecond range at best.[4] In addition, many ferroelectric materials suffer from irreversible mechanical disorder after many cycles of polarization switching,[2] rendering some uncertainty on the long term device reliability aspect.

An alternative to a ferroelectric gate is a floating gate chargeable by field emission, which hopefully circumvents the above mentioned difficulties. Consider a sandwich structure, metal $M(1)$, insulator $I(1)$, metal $M(2)$, insulator $I(2)$, and finally metal $M(3)$. (See Fig. 1). If the thickness of $I(1)$ is small enough so that a field-controlled electron transport mechanism such as tunneling or internal tunnel-hopping are possible, a positive bias on $M(3)$ with respect to $M(1)$ with $M(2)$ floating [$M(2)$ is called the floating gate henceforth], would cause electron accumulation in the floating gate, provided electron transport

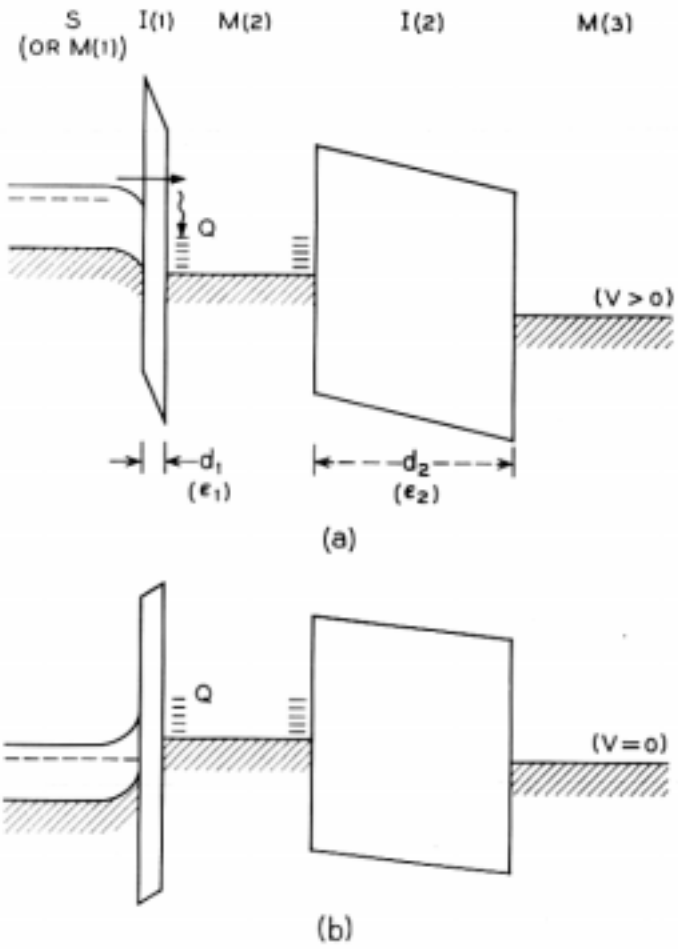


Fig. 1 — Energy band diagram of a floating gate structure with a semiconductor-insulator-metal-insulator-metal sandwich. For calculation of the stored charge, the semiconductor is replaced by a metal $M(1)$. (a) When a positive voltage step is applied to the outer gate. (b) When the voltage is removed. The stored charge Q causes an inversion of the semiconductor surface.

across $I(2)$ is small. These conditions can be met by choosing $I(1)$ and $I(2)$ such that the ratio of dielectric permittivity ϵ_1/ϵ_2 is small and/or the barrier height into $I(1)$ is smaller than that into $I(2)$. The sandwich structure is somewhat similar to the tunnel emitter metal-base transistor proposed by Mead[5] in its structure but with the following essential differences.

(*i*) $M(2)$ is much thicker than the hot electron range, so that emitted electrons are close to the Fermi-level of $M(2)$ before reaching $I(2)$.
(*ii*) No carrier transport is allowed across $I(2)$.
(*iii*) $M(2)$ is floating.

The stored charge Q, as a function of time when a step voltage function with amplitude V is applied across the sandwich, is given by

$$Q(t) = \int_0^t j\, dt' \qquad \text{coul/cm}^2. \tag{1}$$

When the emission is of Fowler-Nordheim tunneling type, then the current density, j, has the form

$$j = C_1 E^2 \exp(-E_0/E), \tag{2a}$$

where C_1 and E_0 are constants in terms of effective mass and the barrier height. (We have neglected the effects due to the image force lowering[6] of the barrier, etc., but the essential feature is expected to be retained even after detailed corrections are made). This type of current transport occurs in SiO_2 and Al_2O_3.

When the field emission is of the internal Schottky or Frankel-Poole type, as occurs in Si_3N_4,[7] then j follows the form

$$j = C_2 E \exp[-q(\Phi_1 - \sqrt{qE/\pi\epsilon_1})/kT], \tag{2b}$$

where c_2 is a constant in terms of trapping density in the insulator, Φ_1 the barrier height in volts, ϵ_1 the dynamic permittivity.

The electric field in $I(1)$ at all times is a function of the applied voltage V and $Q(t)$, and is obtainable from the displacement continuity requirement as

$$E = \frac{V}{d_1 + d_2(\epsilon_1/\epsilon_2)} - \frac{Q}{\epsilon_1 + \epsilon_2(d_1/d_2)}, \tag{3}$$

where d_1 and d_2 are the thickness of $I(1)$ and $I(2)$, respectively.

Fig. 2(a) shows the results of a theoretical computation using (1), (2a), and (3) with the following parameters: $d_1 = 50$ Å, $\epsilon_1 = 3.8\,\epsilon_0$ (for SiO_2), $d_2 = 1000$ Å, $\epsilon_2 = 30\,\epsilon_0$ (for ZrO_2), and V = 50 volts. One notes that the stored charge initially increases linearly with time and then saturates. The current is almost constant for a short time and then decreases rapidly. The field in $I(1)$ decreases slightly as the time increases. The above results can be explained as follows: When a voltage pulse is applied at $t = 0$, the initial charge Q is zero, and the initial electric field across $I(1)$ has its maximum value, $E_{max} = V/[d_1 + (\epsilon_1/\epsilon_2)d_2]$. As t increases, Q will first increase linearly with time. This is because of the fact that for small Q such that E remains essentially the same, the current will in turn remain the same, so $Q = j(E_{max}) \cdot t$. Eventually, when Q is large enough to reduce the

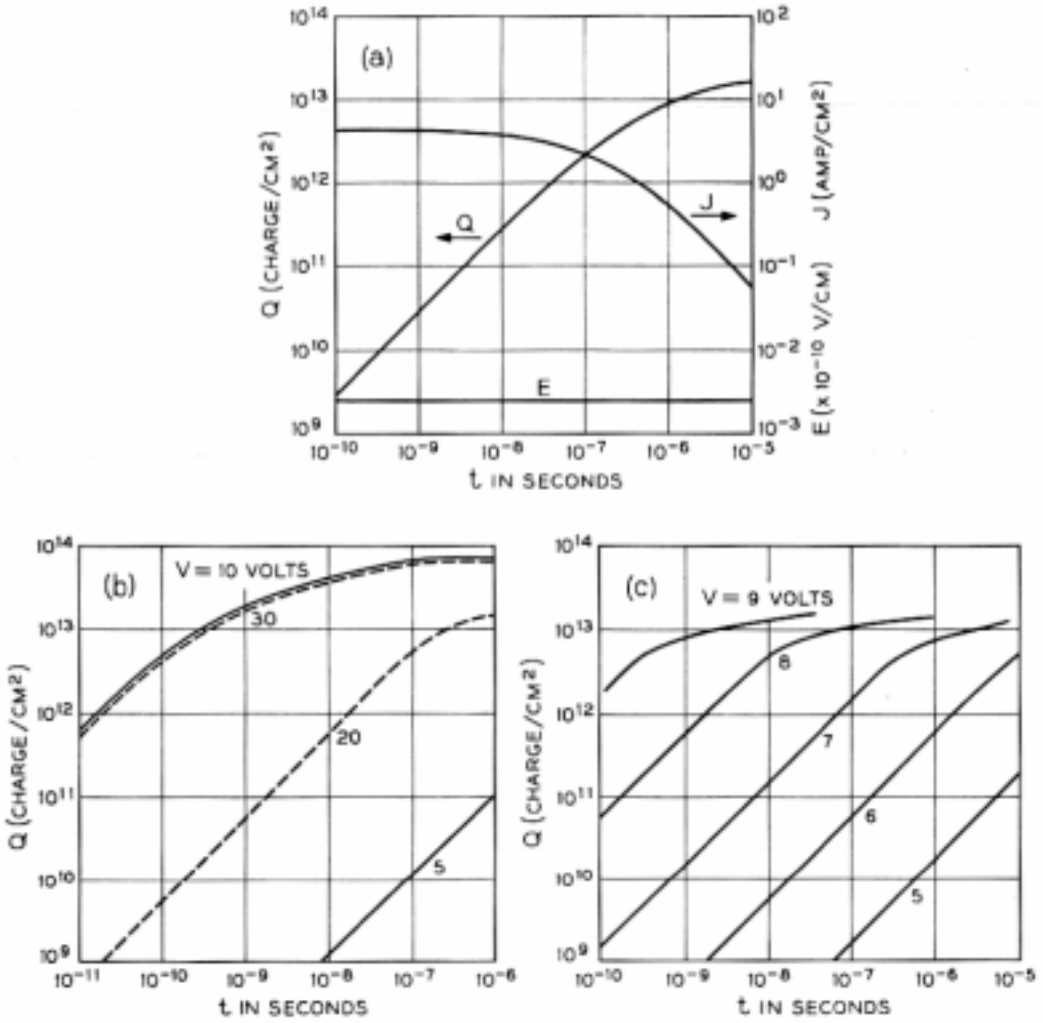


Fig. 2(a) — Theoretical results of the stored-charge density (Q), the current density (J), and the electric field across $I(1)$ as a function of time. V = 50 volts, $d_1 = 50$ Å, $\epsilon_1/\epsilon_0 = 3.8$ (for SiO_2), $d_2 = 1000$ Å, $\epsilon_2/\epsilon_0 = 30$ (for ZrO_2). (b) Theoretical results of the stored charge density as a function of time with the same ϵ_1 and ϵ_2 as in (a), and $d_1 = 10$ Å, $d_2 = 100$ Å (solid lines), $d_1 = 30$ Å, $d_2 = 300$ Å (dotted lines). (c) Theoretical results of the stored density as a function of time with $d_1 = 20$ Å, $\epsilon_1/\epsilon_0 = 60$ (for Si_3N_4), $d_2 = 200$ Å, $\epsilon_2/\epsilon_0 = 30$ (for ZrO_2) and various applied voltages.

value of E substantially, then the current will decrease rapidly with time and Q increases slowly.

Fig. 2(b) shows the stored charge as a function of time for the time ϵ_1 and ϵ_2 but different d_1, d_2, and V. It is clear that for a given structure, in order to store a given amount of charge, one can either increase the applied voltage or increase the charging time (pulse width) or both. Fig. 2(c) shows the calculated stored charge for the current transport described by (2b). Here $I(1)$ is a 20 Å thick Si_3N_4 film. There are

marked decreases in the gate voltages required for a given charge compared to Al_2O_3, SiO_2. This is largely due to the much lower barrier height (1.3 volts)[7] compared to SiO_2 ($\approx$ 4.0 volts).[8]

It is noted that the field in $I(1)$ for appreciable charge storage is in the 10^7 V/cm range. When the outer gate voltage is removed, the field in $I(1)$ due to the stored charge on the inner gate is only 10^6 V/cm or so corresponding to 5×10^{12} charges/cm^2, a large enough charge to detect easily. Since the transport across $I(1)$ is highly sensitive to the field, (2a) and (2b), no charges flow back. The charge loss is actually controlled by the dielectric relaxation time of the sandwich structure,[9] which is very long. When it is desired to discharge the floating gate quickly, it is necessary to apply to the outer gate a voltage about equal in magnitude but opposite in polarity to the voltage which was used for charging. It is evident that net positive charges (loss of electrons) can also be stored in the floating gate if the discharging gate voltages are appropriately chosen in magnitude and duration.

It was mentioned that the stored-charge density of 5×10^{12}/cm^2 was sufficient for easy detection. One of the detection or read-out schemes is to use the surface field effect transistor (MOSFET or IGFET) first fabricated and described by Kahng and Atalla[10] in 1960. For inversion at a silicon surface, the charge required is only about 2×10^{11}/cm^2 for 1 ohm-cm n-type silicon. However, surface-state charges at the silicon-silicon dioxide interface may be as high as 10^{12}/cm^2, depending on the fabrication techniques used. For this reason we have chosen 5×10^{12}/cm^2 as the stored charge required for easy detection. When the Insulated Gate Field Effect Transistor (IGFET) principle is used for read-out, $M(1)$ is now replaced by silicon. This requires a slight correction in the calculation of charge flow through the insulator, but the major features of the results are not expected to be altered significantly. It is to be noted that about one half of the stored charge can be active in creation of the inversion layer since the other half resides near the $M(2)$-$I(2)$ interface due to Colombic repulsion.

To check the feasibility, a floating gate device has been fabricated using an IGFET as shown in Fig. 3(a). The substrate is an n-type silicon, 1 ohm-cm, and ⟨111⟩ oriented. $I(1)$ is a 50 Å SiO_2 thermally grown in a dry oxygen furnace. $M(2)$ and $I(2)$ are Zr (1000 Å) and ZrO_2 (1000 Å), respectively. $M(3)$ and the ohmic contact metals are aluminum deposited in a vacuum system. Fig. 3(b) is another version of the floating gate device using a thin film transistor (TFT) structure.[11]

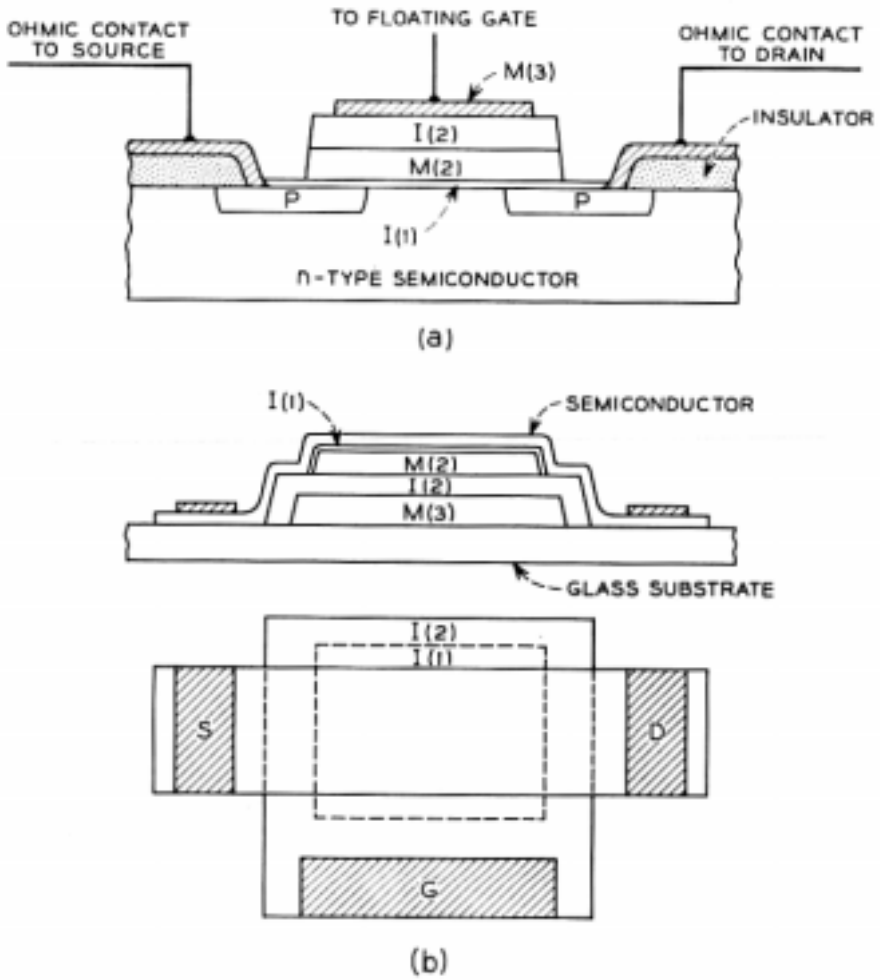
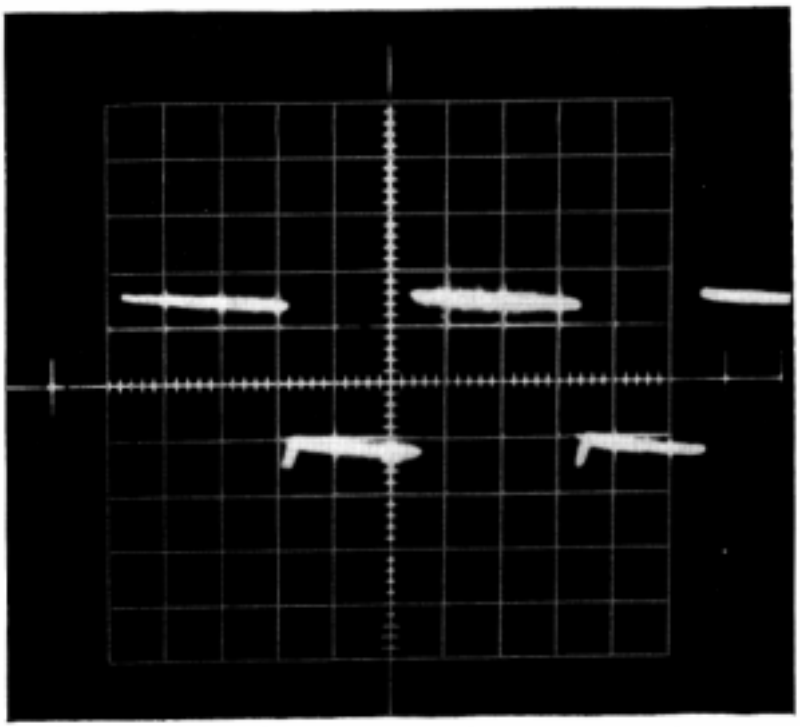
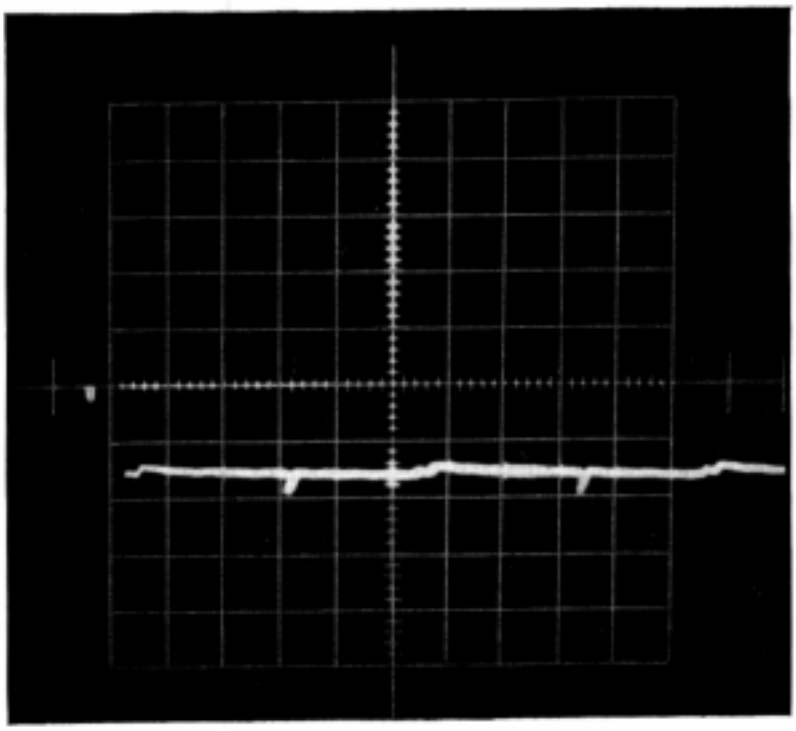


Fig. 3 — (a) Schematic diagram of a floating gate device using an IGFET. The numbers indicated correspond to those shown in Fig. 1. (b) Schematic diagram of a floating gate device using thin film transistor structure.

The IGFET-type floating gate devices have been tested in a pulsing circuit. Because of the relatively thick insulator layers, large voltages ($\approx$ 50 V) and long pulse width ($\approx$ 0.5 μs) have to be applied in order to store $\approx 5 \times 10^{12}$ charges/cm^2. Fig. 4 shows the experimental results. A positive pulse of 50 volts is first applied to the gate electrode, and 60 ms later a negative pulse of 50 volts is applied. Then the pulsing cycle repeats. In Fig. 4(a) the pulse widths are 0.5 μs. One notes that when the positive pulse is applied, a sufficient amount of charge is stored in the floating gate so that the silicon surface is inverted; a conducting channel is thus formed, and the channel current is "on." It can be seen that the channel current decreases only slightly at the end of 60 ms. When the negative pulse is applied, the stored charge is eliminated, and also the channel. The channel current reduces to its



(a)



(b)

Fig. 4 — Experimental results of the channel current of a IGFET-type floating gate device. A positive voltage pulse, V_1, with pulse width W_1, is first applied to the gate, and 60 ms later a negative pulse V_2 with pulse width W_2 is applied. Then the pulsing cycle repeats. Horizontal scale: 20 ms/div. Vertical scale: 0.1 ma/div. (a) $V_1 = V_2 = 50$ volts, $W_1 = W_2 = 0.5$ μs. (b) $V_1 = V_2 = 40$ volts, $W_1 = W_2 = 0.5$ μs.

"off" state. Fig. 4(b) shows results for pulses with the same widths but smaller amplitude (40 V). Since the stored charge is a strong function of the pulse amplitude, only a very small amount of charge is stored, too small to cause inversion. For non-leaky units, the memory holding time of longer than one hour has been observed.

It is clear that a modified IGFET such as a TFT can be used for read-out, as shown in Fig. 3(b). For an academic study of device operation, the floating gate can be partially exposed and a potential probe can be placed nearby.

In conclusion, it has been demonstrated that the controlled field emission to the buried "floating" gate may be capacitively induced by pulsing the outer gate electrode. This combination can therefore, be used as a memory device, with holding time as long as the dielectric relaxation time of the gate structure and with continuous nondestructive read-out capability. There seems to be no inherent reason why read-in read-out cannot be performed in a very short time, say in the nanosecond range or even shorter.

We wish to acknowledge helpful discussions and technical assistance rendered by M. P. Lepselter and P. A. Byrnes in connection with ZrO_2, and skillful technical assistance given to us by G. P. Carey and A. Loya, and J. F. Grandner and his group.

REFERENCES

1. Shockley, W., and Pearson, G. L., Phys. Rev., *74*, 1948, p. 232.
2. Ross, I. M., U. S. Patent 2,791,760, issued May 7, 1957.
3. Jona, F. and Shirane, G., *Ferroelectric Crystals*, MacMillan Co., 1962.
4. Heyman, P. M. and Heilmeier, G. H., Proc. IEEE, *54*, 1966, p. 842.
5. Mead, C. A., Proc. IRE, *48*, 1960, p. 359.
6. Simmons, J. G., J. Appl. Phys., *34*, 1963, p. 1793.
7. Sze, S. M., Current Transport and Maximum Dielectric Strength of Silicon Nitride Films, J. Appl. Phys., June, 1967.
8. Williams, R., Phys. Rev., *140*, 1965, A569.
9. Kahng, D., Semipermanent Memory Using Capacitor Charge Storage and IGFET Read-out, B.S.T.J., this issue, pp. 1296–1300.
10. Kahng, D. and Atalla, M. M., Silicon-Silicon Dioxide Field Induced Surface Devices, presented at the IRE-AIEE Device Res. Conf., Carnegie Inst. Tech., Pittsburgh. Pa., June 1960.
11. Weimer, P. K., An Evaporated Thin Film Triode, presented at the IRE-AIEE Device Res. Conf., Stanford University, Stanford, Calif., June 1961.