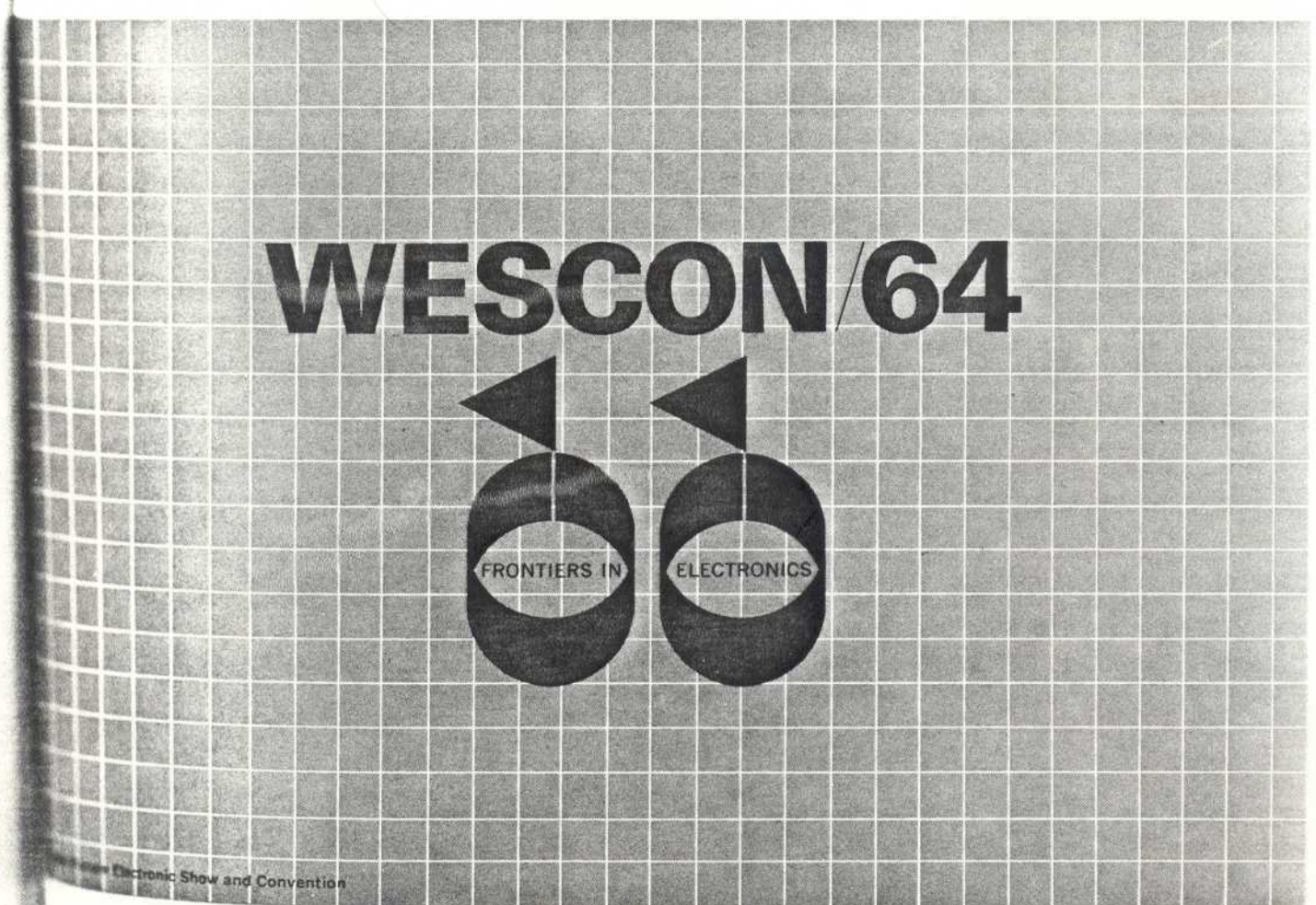
THE SURFACE CONTROLLED AVALANCHE TRANSISTOR

W. Shockley and W. W. Hooper

Clevite Semiconductor, Palo Alto Plant
Palo Alto, California

12.1



THE SURFACE CONTROLLED AVALANCHE TRANSISTOR

W. Shockley and W. W. Hooper
Clevite Semiconductor, Palo Alto Plant
Palo Alto, California

Summary

A new principle for obtaining amplification in transistors consists of utilizing the control of avalanche breakdown voltage by external fields applied through surface insulating layers. The resulting control of reverse current across the p-n junction indicates the possibility of devices operating at frequencies comparable to the reciprocal transit time across the space charge layer for carriers moving with close to their peak velocities (estimated as 2×10^7 cm/sec) in silicon. This suggests that devices with frequencies in the 10 Gc range may be possible with structures having dimensions of the order of 10^{-4} cm.

Introduction

Transit time effects, especially diffusion through base layers of junction transistors and drift at relatively low fields in channels of field-effect transistors, have adverse effects on high frequency transistor performance. In the "avalanche emitter" mode of operation[1] the carriers that constitute the output current are generated in high electric fields and thus can have relatively short transit times. However, the input circuit draws relatively high currents, and these currents require input power and result in reduced power gain.

A new transistor amplifying principle has been discovered that has the advantageous short transit-time features of the avalanche emitter without the disadvantage of drawing large input currents. It promises significant improvement of high frequency performance over junction or field-effect transistors of comparable dimensions. Power amplification results from appropriately utilizing control of avalanche breakdown of a p-n junction by applying voltage to an external electrode so that an electric field penetrates the semiconductor surface and extends into the space-charge layer. This field modifies that produced by reverse bias across the p-n junction, thus changing the voltage across the junction at which avalanche breakdown occurs, consequently controlling the avalanche breakdown currents, and therefore the power delivered to a load in series with the junction.

Control of reverse breakdown voltage of p-n junctions has been reported for germanium by Forster and Veloric[2] and for silicon by Hofstein and Heiman.[3] Forster and Veloric used an insulating layer consisting of a "thin mica spacer" lying over a specimen in the form of a grown p^-n^+ junction. Hofstein and Heiman used the SiO_2 layer of a "Silicon Insulated Gate Field-Effect Transistor" to transmit an electric field to the junction. The use of the resulting valve action of control of breakdown voltage to give power gain is, however, not discussed in either reference.

General Features

In one form the surface controlled avalanche transistor consists of a silicon $(n^+)p$ junction covered with a thin insulating layer of SiO_2. The electrodes are the "source" or (n^+) region, the "drain" or p region, and the "gate" which consists of a metal layer lying on the oxide and over the intersection of the junction with the surface. The application of an electric field between gate and source controls avalanche breakdown voltage between source and drain. For SiO_2 as the dielectric, the electric displacement $K(SiO_2)E$ can substantially exceed that required to produce avalanche breakdown in silicon,[4] namely $KF_B \doteq (10^{-12}$ farad/cm) x $(5 \times 10^5$ volt/cm) = 5×10^{-7} coulomb/cm^2. For the abrupt junction between a heavily doped and a weakly doped region, the avalanche field will occur at the junction where the potential is nearly equal to that of the heavily doped region.

Secondary carriers of one sign generated by the avalanche multiplication flow to the heavily

doped side. Carriers of the other sign flow across the wider portion of the space-charge layer to the other side; during this flow these latter carriers will, for the initial portion of their paths, be drawn towards the silicon surface if the gate voltage is applied so as to decrease the breakdown voltage. The field tangential to the surface produced by reverse bias across the junction will cause these carriers to be collected by the weakly doped region.

High frequency response results from generation of carriers at the avalanche source point and subsequent drift in high fields where drift velocities are near their maximum value[4] of about $v_m = 2 \times 10^7$ cm/sec. To ensure high drift velocities all along the path the p region may consist of a (p^-) or intrinsic layer with a (p^+) electrode region.

Experimental units having low breakdown regions near the surface have shown controlled variations in breakdown voltage of a factor of two and current and power gains greater than 10^4 at low frequencies. The drain-source differential resistance at constant gate-source voltage is consistent with predictions[4] of plasma-space-charge resistance based on a "unit-cube" conductance of $Kv_m \doteq 2 \times 10^{-5}$ mhos.

Design Considerations

The high frequency potential of the surface controlled avalanche transistor can be appreciated in terms of its design theory which we illustrate for the structure of Fig. 1. The representation may be considered to be one unit of a repeating structure, such as may be produced by interdigitation; the periphery or the total length perpendicular to the plane of the figure of the working region being P cm. Figure 2 illustrates the carrier paths under operating conditions, and introduces the equivalent thickness, a, of the avalanche control layer and the sign conventions for voltage and current symbols.

Structures of the type illustrated may be fabricated by the usual photo-lithographic techniques and the well known methods used for producing oxides. It should be noted that a much thinner oxide layer lies under the working region of the gate electrode than where the electrode overlaps the heavily doped source region. This reduces unwanted input capacitance between the gate and source.

For steady state conditions, the carrier current consists of holes generated by avalanche in the immediate vicinity of the avalanche source point (asp), the secondary electrons being drawn to the source. Provided the voltage is large enough to sustain avalanche, this current is represented by the following expression

$$I = -I_d = I_s = -g\,[\,V_g + (V_d/\mu) + V_{go}\,]$$

If the square bracket term is positive, avalanche is not sustained and $I = 0$. The other quantities are as follows:

$$g \doteq K(Si)\, v_m)(P/a)$$

$$a = [\,K(Si)/K(SiO_2)\,]\ (\text{thickness } SiO_2)$$

$$\mu = wb/\pi a^2$$

$$V_{go} \doteq \pi a F_B/2$$

The value of μ is based on an approximate potential theory treatment in which a region of uniform dielectric constant K(Si) = 1.04 pF/cm is bounded by one equipotential at ground along the y-axis of Fig. 2, another equipotential at voltage V_g on the circle of radius b - a and a third equipotential at V_d at x = w. The capacitance per unit area of a layer of thickness "a" having the dielectric constant of silicon is the same as the actual layer of SiO_2 which is thinner in the ratio of the dielectric constants. The field at the asp due to V_g is approximately $V_g/(\pi a/2)$, and due to V_d is approximately $V_d 2a/bw$. These approximations lead to $\mu = bw/\pi a^2$ and to the avalanche field F_B occurring at the asp when $[\,V_g + (V_d/\mu)/(\pi a/2) = F_B$; this leads to the V_{go} term. The value of the transconductance g can be estimated by noting that the current (I/P) amp/cm flowing uniformly at velocity v_m on the barrier cylinder of radius b produces a charge density (I/Pv_m) C/cm and a voltage drop $(I/Pv_m)(a/K) = \Delta V$ across the avalanche control layer so that in order to cancel its effect requires a change of equal magnitude in V_g. The resulting expression for g can also be interpreted as a row containing (P/a) cubes, each of edge "a" and each providing the conductance $v_m K$ associated with a "unit cube" of avalanche breakdown plasma.[4]

A figure of merit of SCAT can be obtained by comparing capacitance and transconductance. For an iterated structure of the form shown in Fig. 1, the capacitances will be approximately C_{sg} = (2 to 3)KP, C_{sd} = (s/w)KP, C_{gd} = (b/w)KP. The corresponding figure of merit obtained by dividing g by the sum of capacitances leads to a circular frequency of $v_m/(4 \text{ or } 5)a$. This is probably overoptimistic because of the fact that capacitative and loss terms will result from the sheath of carriers flowing along the boundary surface. Their effects will be important at frequencies corresponding to the transit time w/v_m which will be somewhat larger than $(4 \text{ or } 5)a/v_m$. For stripe dimensions 2s and 2b of the order of 10^{-4} cm, (which have been achieved in junction transistors) and aw values of 3×10^{-4} cm, $v_m/2\pi w \doteq 10^{10}$ cps. Detailed design considerations involving problems of excess heating, control of parasitic impedance and the like will be necessary to evaluate the practicality of actual operation in keeping with these values.

A number of additional considerations are relevant to the operation of the device. In brief, some of the principle features are these: Thermal instability leading to hot spots in junction transistors arises from the positive temperature coefficient of emitter, and hence, collector current. The temperature coefficient of avalanche current in the SCAT, however, is negative, and it is this characteristic which leads to thermal stabilization of this device.

In the SCAT, the drain voltage should be sufficiently large to make space charge effects of the current I unimportant, but not so large as to produce F_B at the point x = b, y = 0 in Fig. 2. These conditions impose limits on the power efficiency of operation. It should be noted, however, that the peak drift velocity v_m probably extends over about one decade of electric field from roughly 50,000 to 500,000 volts/cm. This range will permit relatively wide fractional swings of V_d.

Consideration of the upper frequency limits associated with reducing the dimensions of the device must take into account the fact that avalanche breakdown does not occur at a fixed field, but at an appropriate integral of secondary ionization coefficients for holes and electrons. Furthermore, these coefficients apply only to uniform field conditions, and refinements are necessary when distances are so short that an appreciable fraction of the total voltage drop occurs over the distance required to give an avalanche generated carrier sufficient energy to produce secondaries. As an example of the effect of such considerations, recent data on ionization rates in silicon[5] suggests that to reduce the extent of high field required for secondary breakdown from 10^{-4} to 10^{-5} cm requires increasing F_B from about 0.35×10^6 to about 1.0×10^6 volts/cm.

It is quite possible that experiments with transistors of this type will lead to better knowledge of the physical phenomena of secondary ionization.

We are indebted to all the members of the Shockley Research Laboratory who helped in preparation and evaluation of the surface controlled avalanche transistor.

References

1. W. Shockley, U. S. Patent 2,936,425 issued May 10, 1960.

2. J. H. Forster and H. S. Veloric, J. Appl. Phys. 30, 906 (1959).

3. S. R. Hofstein and F. B. Heiman, Proc. IEEE 51, 1190 (1963).

4. W. Shockley, Solid State Electronics 2, 35-67 (1961).

5. C. A. Lee, R. A. Logan, R. L. Batdorf, J. J. Kleimack and W. Weigmann, Phys. Rev. 134, A761 (1964).

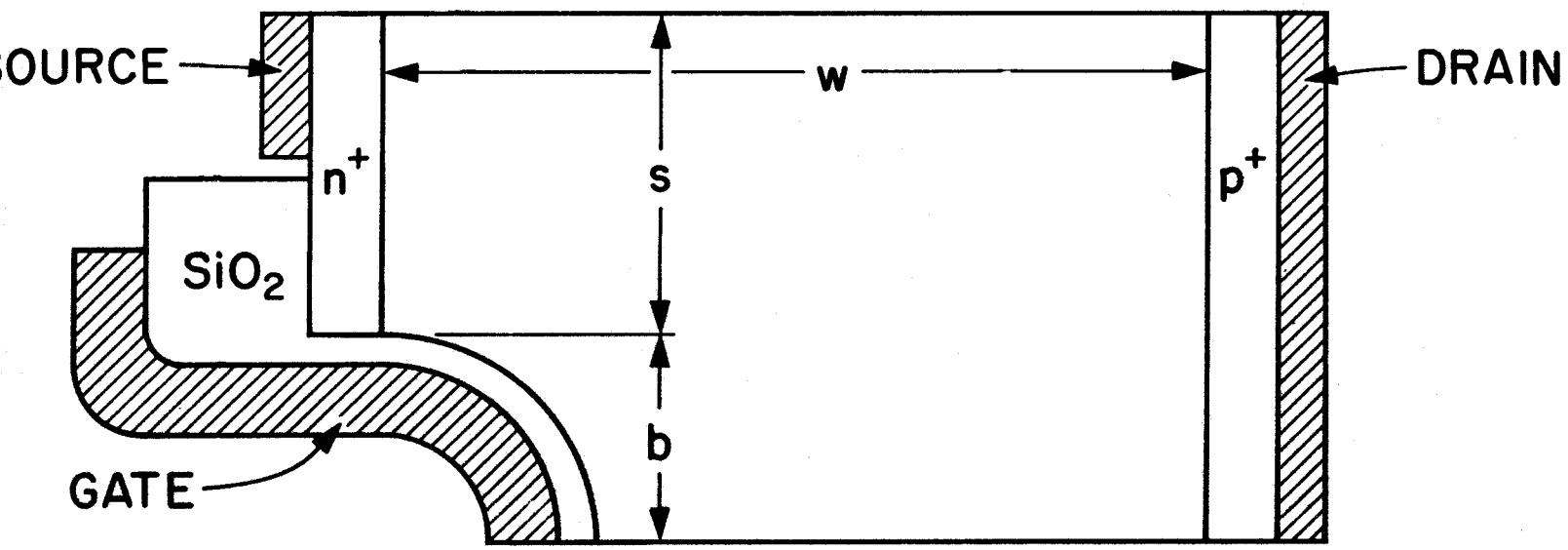


Fig. 1. Schematic representation of the surface controlled avalanche transistor.

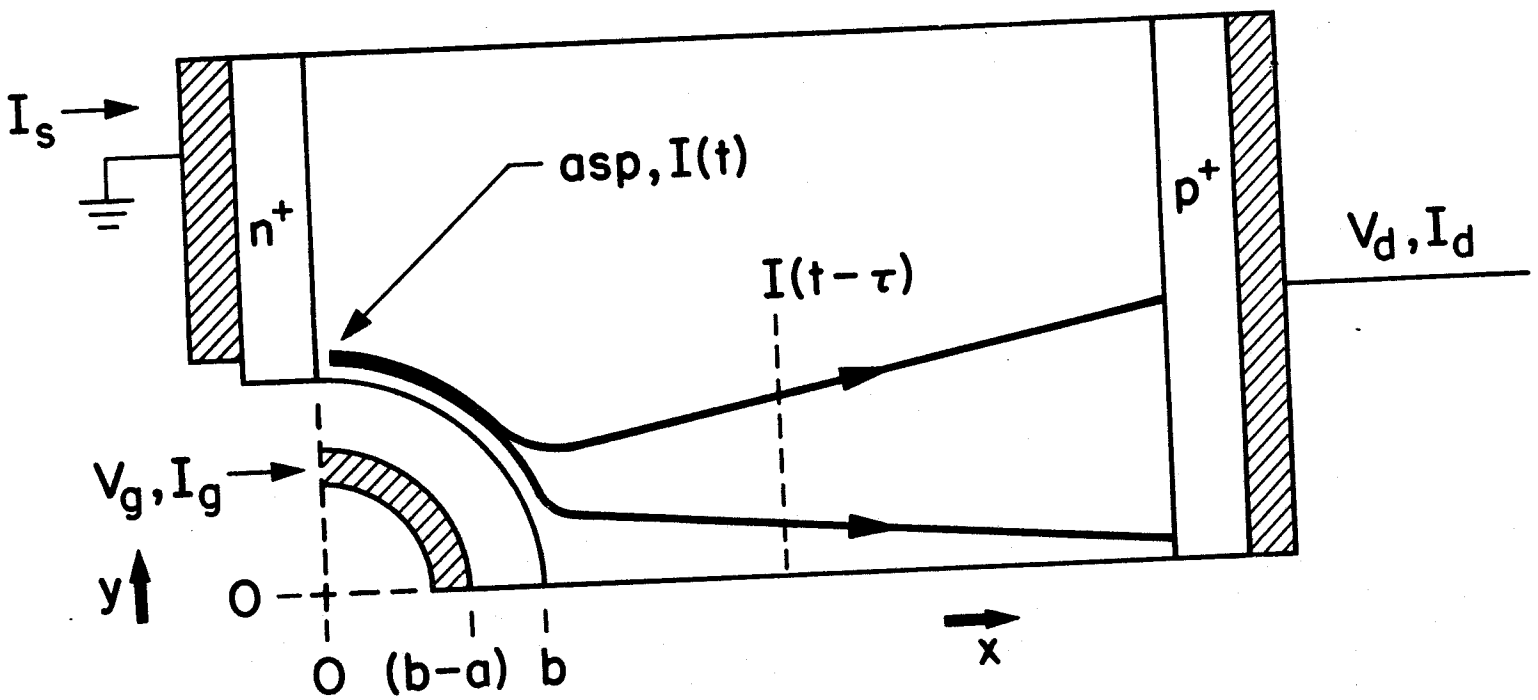


Fig. 2. Carrier paths illustrated for operating conditions of the surface controlled avalanche transistor.

next cascade equals NQ_1 where $Q_1 = C_1 \cdot V_{DI}$
This shows that the device has a charge gain (equal to the fan-out N) which is necessary for logic operation.

10 April 4/10/84, pp. 11-17

Jack Lewis

Read & Understood

Apr 10, 1984 pp 11-17

4/27/84 Tunnel Triode or "Breakdown Injection Transistor" (BIT? BRIT

The following is a generalization of the ideas described on pp. 11-17. — to include other than avalanche mechanisms of Breakdown, notably the tunneling breakdown.

It has been suggested to me (by D. Kahng) that an important limitation on the speed of AVIT operation results from the high field required for avalanching, which in turn requires high gate voltages. Mathematically, this limitation can be described as follows:

In a given geometry and fixed doping profile the avalanche current I_a is a unique function

of the applied electric field, which is determined by the amount of charge Q placed on the gate, viz.

$$I_a = f(Q) \qquad (1)$$

The gate delay of the AVIT is therefore limited by the time τ defined by

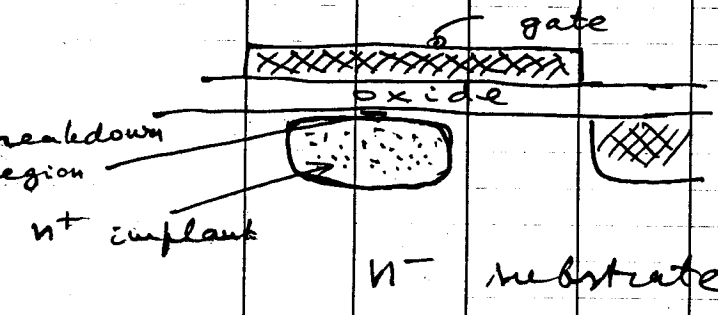
$$1/\tau = \partial I_a / \partial Q \qquad (2)$$

To optimize the performance of AVIT one has to maximize $\partial I_a/\partial Q$ by varying the doping profile and the geometry of the device & the avalanche region. One way of enhancing $\partial I_a/\partial Q$ is by using asperity of the surface — which would enhance local fields at a given overall Q (this is similar to the ion-microscope idea, where sharpness of a needle enhances the local field near its tip). The other approach is to make a narrower field region — by bringing the implanted region closer to the surface and using thinner oxides. This, of course, cannot be ~~done~~ improved indefinitely, since the oxide thickness is limited to about, say, 100 Å and the avalanche region cannot be made too short either — because for very thin

films the breakdown field begins to increase with decreasing film thickness.

∴ For sufficiently thin fields regions (very high implant density, very close to the surface) the avalanche mechanism will not be dominant in breakdown - yielding the dominant role to tunneling mechanism. This is not at all bad!

The entire reasoning of pp 11-17 equally applies to the tunneling mechanism. The prototype structure is then shown in the figure below:



The entire proposal thus reduces to a new type of source, viz. n^+ type for a p-channel transistor and p^+ type for an n-channel one.

4/27/84

Red & Underlined — Serge Luryi. pp 17-19

pp 17-19 — April 27, 1984