

UNCLASSIFIED

AD 282 930

*Reproduced
by the*

**ARMED SERVICES TECHNICAL INFORMATION AGENCY
ARLINGTON HALL STATION
ARLINGTON 12, VIRGINIA**



UNCLASSIFIED

NOTICE: When government or other drawings, specifications or other data are used for any purpose other than in connection with a definitely related government procurement operation, the U. S. Government thereby incurs no responsibility, nor any obligation whatsoever; and the fact that the Government may have formulated, furnished, or in any way supplied the said drawings, specifications, or other data is not to be regarded by implication or otherwise as in any manner licensing the holder or any other person or corporation, or conveying any rights or permission to manufacture, use or sell any patented invention that may in any way be related thereto.



RADIO CORPORATION OF AMERICA RCA LABORATORIES

SPECIAL SCIENTIFIC REPORT

INVESTIGATIONS OF FUNDAMENTAL LIMITATIONS
DETERMINING THE ULTIMATE SIZE OF MICROSTRUCTURES

CONTRACT NO. AF19(604)-8040

PREPARED FOR
ELECTRONICS RESEARCH DIRECTORATE
AIR FORCE CAMBRIDGE RESEARCH LABORATORIES
OFFICE OF AEROSPACE RESEARCH
UNITED STATES AIR FORCE
BEDFORD, MASSACHUSETTS

PROJECT NO. 5633
TASK NO. 56332

REPORT DATE: FEBRUARY 28, 1962



DAVID SARNOFF RESEARCH CENTER
PRINCETON, NEW JERSEY

282930

ASTIA

282 930

AFCRL-62-140

SPECIAL SCIENTIFIC REPORT
INVESTIGATIONS OF FUNDAMENTAL LIMITATIONS
DETERMINING THE ULTIMATE SIZE OF MICROSTRUCTURES

Report Date: February 28, 1962

THE VALIDITY OF SCALING PRINCIPLES
FOR FIELD-EFFECT TRANSISTORS

Project 5633

Task 56332

Contract No. AF19(604)-8040

Report prepared by: T. O. Stanley

The research reported in this document has been sponsored by the Electronics Research Directorate, Air Force Cambridge Research Laboratories, Office of Aerospace Research, United State Air Force, Bedford, Massachusetts. The publication of this report does not necessarily constitute approval by the Air Force of the findings or conclusions contained herein.

RADIO CORPORATION OF AMERICA
RCA LABORATORIES
PRINCETON, NEW JERSEY

TABLE OF CONTENTS

* * * * *

	PAGE
ABSTRACT	
1. INTRODUCTION	1
2.0 PRINCIPLES OF SCALING	1
2.1 TRANSMISSION LINE	1
2.2 FIELD-EFFECT TRANSISTOR	3
2.3 BIPOLAR TRANSISTOR	6
3.0 INSULATED GATE FIELD EFFECT TRANSISTOR ANALYSIS	7
3.1 DESIGN IMPLICATIONS	10
4. CONCLUSIONS	11

ABSTRACT

The principles of scaling are described, including the convention of scaling some material property. In particular, semiconductor doping density is scaled inversely with linear dimensions. Electrical scaling exists, by definition, when electrical properties obey some prescribed proportionality to mechanical dimensions.

The scaling of transmission lines, field-effect transistors, and bipolar transistors is examined. Transmission lines exhibit a delay time associated with series resistance and distributed capacitance which is a function of geometry but not of scale, on the order of 10^{-10} seconds for typical interconnections. Electrical scaling is valid for field-effect transistors, providing doping is scaled inversely with linear dimensions. For this convention, delay time scales linearly, and area power density is invariant with scale. Simple electrical scaling is invalid for bipolar transistors.

The range of dimensions for which electrical scaling will no longer be valid for the recently developed insulated-gate field effect transistor is not known. Continued investigation of the physical phenomena associated with a conducting channel parallel and close to a semiconductor-insulator interface, and of the dependence of these phenomena on fabrication techniques is indicated.

1. INTRODUCTION

Examination of the principles of scaling in the context of recently developed field effect devices⁽¹⁾ indicate that much is to be gained by further refinement of fabrication techniques and a more thorough understanding of the silicon-silicon dioxide interface. These principles and their limitations are discussed below, and where appropriate, referred to the premises and findings of Scientific Report No. 1, "Maximum Packing Density of Non-Redundant Semiconductor Devices", March 31, 1961.

2. PRINCIPLES OF SCALING

To scale, as defined here, is to alter the dimensions of a device in prescribed proportions. One may, in principle, also alter material properties: for example, materials of varying dielectric constant might be chosen in accordance with some linear dimension. In semiconductor devices, the scaling of doping density constitutes alteration of material properties.

Once the rules of proportion are prescribed, the dependence on scale of the device electrical parameters may be predicted, and the physical phenomena most likely to limit the range of dimensions for which these parameters follow a prescribed proportionality to device dimensions may be ascertained. Within this range, electrical scaling is said to be valid.

2.1 TRANSMISSION LINE

The mechanical dimensions of the transmission line in question are scaled in equal proportion, linearly with an arbitrary

(1) Quarterly reports, Contracts AF19(604)8040, AF19(604)-8836.

reference dimension. This is indicated by the degree of proportionality in parentheses: (1). Conductivity and dielectric constant are assumed to be constant.

Thus:

length	(1)
conductor spacing, etc.	(1)
conductor area	(2)
conductivity	(0)
dielectric constant	(0)
surge impedance	(0)
electrical length	(1)
total shunt capacitance	(1)
total series resistance	(-1)
total series inductance	(1)
lumped impedance time constant, L/R	(2)
lumped impedance time constant, RC	(0).

The last item indicates that for a given geometry of interconnections when scaled to very small dimensions, a lower-limit time delay is encountered. This is on the order of 10^{-10} seconds for geometries typical of integrated logic nets, and for optimum device impedance level. Devices exhibiting awkward impedance levels, e.g. tunnel diodes,⁽²⁾ suffer greater delays.

(2) Tunnel Diode Balanced Pair Switching Characteristics, J. J. Gibson, G. B. Herzog, A. S. Miller and R. A. Powlus; 1962 International Solid-State Circuits Conference.

Two physical phenomena which limit the validity of electrical scaling for transmission lines are skin effect, which leads to larger-than-scale resistance for large dimensions at high frequencies and electron scattering by defects and surfaces, which leads to larger-than-scale resistance for very small dimensions (order of $1000 \text{ \AA}$ conductor thickness).

2.2 FIELD EFFECT TRANSISTOR

Field-effect transistors, as we shall see, lend themselves to analysis based on scaling, while bipolar transistors, in their present degree of refinement, do not. For this reason they are considered separately. The mechanical dimensions of the field-effect transistor in question are scaled in equal proportion, and doping density is scaled inversely with an arbitrary reference dimension. As will be seen, this convention holds the operating electric fields within the transistor invariant with scale. Again, the mechanical dimensions scale:

length	(1)
area	(2)
volume	(3)

But now we scale doping density, and thereby fixed charge, (doping x volume):

doping density	(-1)
fixed charge	(2)

The fixed charge in the transistor is accompanied by equal and opposite mobile charge which may be extracted by applying a suitable electric field, as in the depleted channel of a field-effect

transistor (or the collector depletion region of a bipolar transistor). On the premise that the depletion depth is to be a fixed fraction of some transistor dimension, the electric fields in the transistor scale as (doping) x (Length), and the applied voltages as (field) x (length):

electric field (0)

voltage (1)

A lower limit to the thickness of the depletion region may be set by the Debye length, on the order of a few hundred angstroms for the devices contemplated.

The capacitance associated with the depletion region is proportional to the ratio of fixed charge to voltage. In reality it is a geometrical capacitance: it scales linearly:

capacitance (1)

Mobile charges in the presence of an electric field gives rise to conduction when they are suitably replenished. The ratio of the velocity with which they move to the magnitude of the applied field is the carrier mobility of the material. It follows that the conductivity of a region is the product of the mobile charge density in that region and its mobility. In a region where mobility is fixed and the density of mobile charges is equal to the doping density, the scaling, then, is:

conductivity (-1)

resistance (0)

For the voltage drops in these conducting regions to remain proportional to voltage applied to depletion regions, current scales

linearly. Mobility is not fixed, however, but depends both on the electric field and on scattering by dopants, surfaces, and defects in the material. By holding fields constant regardless of scale, only the latter effect applies, and constitutes a limit, for small dimensions, to the validity of electrical scaling. A quantitative measure of this limit, particularly for conduction parallel and close to a silicon-silicon dioxide interface, is a primary objective of our research.

The actual power dissipation of a transistor depends on its mode of operation. Once this has been determined, the scaling of dissipation follows the scaling of (voltage) x (current):

$$\text{power} \quad (2)$$

In an integrated net one is interested for thermal reasons, in power density. For integration in a two-dimensional array, as on the surface of a semiconductor wafer, it is noteworthy that surface loading remains constant:

$$\text{area power density} \quad (0).$$

Summarizing the scaling laws for field-effect devices, with limitations cited above:

length	(1)
area	(2)
doping	(-1)
fixed charge	(2)
electric field	(0)
voltage	(1)
capacitance	(1)

current	(1)
resistance	(0)
time constant, RC	(1)
power	(2)
area power density	(0).

These relations and limitations will be discussed quantitatively following a review of bipolar transistor scaling relationships.

2.3 BIPOLAR TRANSISTOR

In bipolar transistors, or in general where minority carrier injection is employed, regions exist into which carriers are injected over a barrier and from which carriers flow by diffusion. The injected charge in such a region is proportional to the injected current and to the square of the thickness of the region. Efficient injection requires* a high ratio of emitter doping to the sum of base doping and injected charge density. Thus, any simple scaling of current and doping is unrealistic. On the other hand, the barrier voltage over which injection must occur is relatively fixed, which in turn fixes a lower limit for operating collector voltage in switching applications. This limit, a few volts, has been reached in present operation of silicon switching transistors, and it too invalidates further electrical scaling.

* High injection efficiency is one of the benefits that would derive from an emitter region of wider band-gap than the base region. The fabrication of a transistor incorporating such an emitter remains a formidable task.

The switching delay for bipolar transistors comprises several terms: transit time delays, diffusion-limited delay, "storage" time, as well as delays associated with simple capacitance charging. Refinements in fabrication techniques have ameliorated various terms in turn, and focussed attention on others. As a consequence, one may estimate that major improvements in switching time will not be gained by scaling. Further area reduction at constant power density appears quite feasible.

3.0 INSULATED-GATE FIELD-EFFECT TRANSISTOR ANALYSIS

A detailed description of the recently developed insulated-gate field-effect transistor will be issued as a scientific report under the associated contract AF19(604)-8836 "Integrated Logic Networks." To provide a quantitative estimate of the benefits to be derived from scaling, a simplified analysis is made here of a somewhat idealized structure based on the same principles, as illustrated in Fig. 1. The following symbols are employed:

B	channel doping, coulombs-cm ⁻²
d	channel length, cm (direction of current flow)
w	channel width, cm
s	insulator thickness, cm
k	insulator dielectric constant, farads-cm ⁻¹
μ	channel mobility cm ² volt ⁻¹ sec ⁻¹

The voltage required to pinch off the channel is sum of the voltages appearing across the silicon depletion region, the voltage across

-7a-

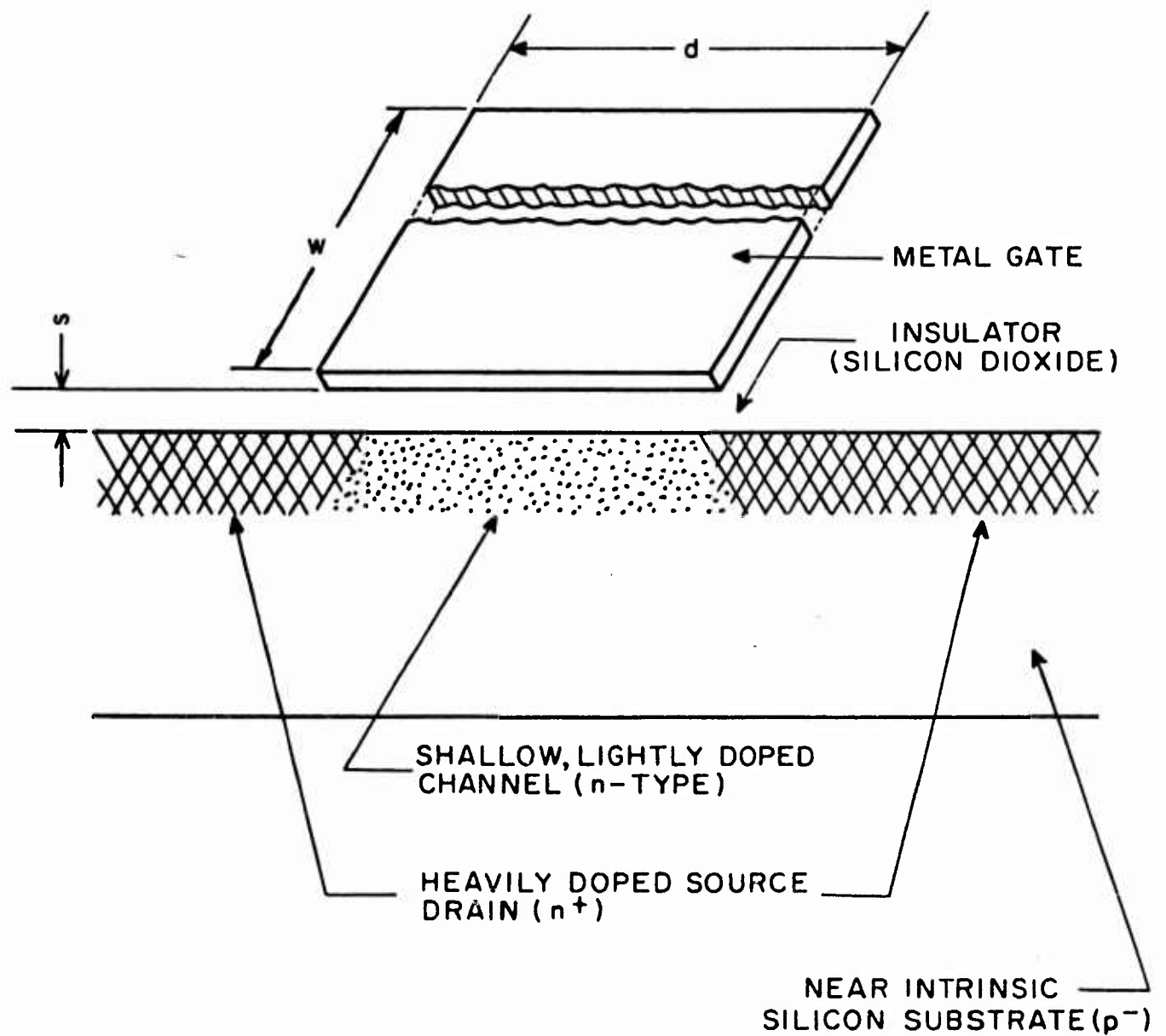


FIG. 1 INSULATED GATE FIELD EFFECT TRANSISTOR (SCHEMATIC)

the insulator ascribable to immobile surface states at the silicon-insulator interface and the voltage across the insulator ascribable to β . In the ideal case, only the last term exists, and for

$$V \quad \text{pinch-off voltage, volts}$$

$$V = \beta s/k.$$

If the gate is biased so as to create no depletion region and the voltage, V , is applied between source and drain, a lateral field E , will exist, giving rise to a mobile charge velocity, v , and current I . Any contact resistance in series with the channel will reduce both E and I . In the ideal case,

$$E \quad \text{lateral field, volts cm}^{-1}$$

$$E = V/d = \beta s/kd$$

$$v \quad \text{mobile charge velocity, cm sec}^{-1}$$

$$v = \mu E$$

$$I \quad \text{channel current, amperes}$$

$$I = \beta v w$$

$$\text{also, } G \quad \text{channel conductance, mhos}$$

$$G = I/V = \mu \beta w/d.$$

If the current, I , is employed to deplete a channel of charge $\beta w d$, the associated delay, T , is

$$T \quad \text{delay, seconds}$$

$$T = \beta w d / I = d/v$$

T may be identified as the channel transit time.

A measure of the power consumed in the device is the product of V and I :

$$P \quad \text{characteristic power, watts}$$

$$P = VI$$

By substituting appropriately, this may be expressed:

$$P = (E^2 k_v d/s) (dw)$$

The first factor is the area power density, dependent only on the design electric field, E and the consequent velocity, v, and on the ratio, d/s; the second factor is the channel area.

Present experimental devices have the following approximate dimensions:

β	10^{-7} coulombs cm^{-2} (estimated)
d	0.0017 cm
w	0.13 cm
s	3×10^{-5} cm
k	4×10^{-13} farads cm^{-1}

The electrical characteristics for such a device, assuming $\mu = 500 \text{ cm}^2 \text{ volt}^{-1} \text{ sec}^{-1}$, calculated on the idealized basis discussed above, are gratifyingly within reach of experimental results:

Parameter	Calculated	Experimental
V	7.5v	8v
I	29ma	14ma
T	$8 \times 10^{-10} \text{ sec}$	$4 \times 10^{-9} \text{ sec}$
P	230mw	112mw
Channel Power Density	$1000 \text{ watts} \cdot \text{cm}^{-2}$	$500 \text{ watts} \text{ cm}^{-2}$

The biggest discrepancy, that in delay, T, is accounted for in part by parasitic capacitance contributed by the gate electrode terminal, and in part by the other idealizations.

3.1 DESIGN IMPLICATIONS

The operating electric fields in this device are near optimum for maximizing speed, since velocity is approaching terminal velocity, but the geometrical factor, d/s , could profitably be reduced an order of magnitude (for d/s much smaller than 10 the calculations of device parameters when reformulated indicate further reduction of d/s to be unprofitable). Maintaining constant field E at reduced d/s requires a corresponding reduction in channel doping, β . (β remains invariant with scale, having the dimensions of charge per unit area). Here, two factors enter. It is desirable that β be appreciably greater than the density of immobile surface-state charge at the silicon-insulator interface, and for this reason the measurement and control of surface states is an objective of our research. Second, the effects of statistical variations in the doping distribution must be examined, as predicted in Scientific Report No. 1. The channel charge calculated for the parameters of the previous section, 3.0, is 2.2×10^{-11} coulombs, on the order of 10^8 electrons. If, for example, β , d , and w are decreased by factors of 10, 10, and 100, the electron count is only 10^4 . A factor that may prove to be mitigating is that the mobile electrons in the channel appear to arise not from donor impurities, but from distortion of the energy band structure at the silicon-insulator interface. This too is of interest in our research effort.

The lateral dimensions d and w of present experimental devices were chosen for convenience and are not indicative of fabrication limits. Reduction of d by more than a factor of two will

require refinements in techniques or introduction of new techniques. The transition between the heavily doped source-drain regions and the channel region must also be scrutinized. A twenty-fold reduction in w should present no problems. The limitation to scaling set by edge uncertainty, as noted in Scientific Report No. 1, depends directly on the the fabrication technique employed.

The insulator thickness, s , can be controlled to great accuracy and uniformity. The dependence on the insulator of channel characteristics is more likely to set a lower limit to insulator thickness than edge uncertainty.

Area power density, as noted in Section 2.1, is invariant with scale. For planar integration, the ratio of the area devoted to electrodes, interconnections, terminals and encapsulation to the channel area may be on the order of 100 to 1000. The corresponding area power density (again assuming reduction of β and d/s by an order of magnitude) is on the order of 0.1 to 1 watt cm^{-2} . The feasibility of such power densities will depend on the application. It is worth noting, however, that for the rather extravagant assumptions which led to a total channel charge of 10^4 electrons, the dissipation per element is only on the order of 10^{-5} watts.

4. CONCLUSIONS

The principles of electrical scaling are applicable to field-effect devices provided volume doping varies inversely with scale (doping per unit area remains constant). Under these conditions, power density per unit area remains constant, and delay varies

linearly with scale . This is in contrast to bipolar transistors, for which simple rules of electrical scaling do not apply.

The physical phenomena most likely to limit the range of dimensions for which electrical scaling is no longer valid remain largely unexplored. These include:

- 1) The mechanisms of conduction in a channel parallel and close to a semiconductor-insulator interface, in particular the silicon-silicon dioxide interface.
- 2) The modulation of conduction in such a channel by the application of a transverse electric field.
- 3) The phenomena associated with the conductivity transition (laterally) from a heavily doped source-drain region to a relatively less doped channel region.
- 4) The dependence of the phenomena in the three areas cited above on the fabrication techniques employed.

The electrical characteristics of recent experimental insulated-gate field-effect transistors are comparable to the characteristics predicted theoretically. Furthermore, such parameters as delay time and power, when compared to the corresponding parameters observed in silicon bipolar switching transistors of similar geometry are sufficiently promising to encourage further scaling to smaller dimensions.