MEASUREMENT OF THE DOPANT DISTRIBUTION IN THIN EPITAXIAL SI AND GaAs STRUCTURES

By

P. GOTTWALD and A. AMBRÓZY

Department of Electronic Devices and Department of Electronics Technology

Technical University, Budapest
Received October 5, 1979

I. Introduction

Manufacture of semiconductor devices relies on producing various dopant atom distributions under the surface of a semiconductor wafer.

The kind of the realized device and its electrical parameters are widely depending upon the particular structure of the realized doping profile.

There is sometimes a considerable change of the dopant atom density in a region as little as e.g. 1 μm^3.

In the future the devices will become ever smaller, at the same time the variety of the dopant distribution will be ever greater. Namely:
— Modern technology often needs very high speed electronic devices. To increase the cutoff frequency, first of all the size of the devices has to be reduced.
— On the other hand, the scale of integration of the microelectronic circuits must be increased as far as possible. Since there are always some crystal defects (e.g. dislocations) along the surface of a semiconductor wafer, the area of the devices must be decreased rather than to increase the chip.

This trend of scale-down is physically limited by the fact that in such a little space region which includes e.g. a single dopant atom, the term *doping density* becomes meaningless. (For instance, for a doping density of 10^{18} cm^{-3}, this volume would be about 10^{-6} μm^3.)

After A. Möschwitzer (TU Dresden), by the turn of this century the smallest linear size of a structure realizable by technological means will be about 0.1 μm [1].

As in the fabrication of such small devices a lot of problems arise, measurement of the dopant distribution in the devices is likely to grow increasingly difficult. Coexisting methods of doping profile measurements— each with its inherent limitation—see e.g. in [2], [3], [4], [5].

A doping profile plotter has been developed at the Technical University, Budapest.

Principles of the measurement, technical facilities, design aspects will be briefly described, supplementing earlier publications [2], [7].

II. Fundamentals of the measurement

The equipment to be described suits non-destructive measurement of the doping profile in a $p-n$ (or Schottky-) junction. The measurement principle is the so-called $C-V$ method [6].

The edges of the depletion-layer in a reverse biased $p-n$ junction are known to be shifting away with the variation of the biasing voltage.

These regions with charges of $+Q$ and $-Q$ of ionized donor and acceptor atoms, respectively, will then endure charge variations by $+\Delta Q$ and $-\Delta Q$, if the biasing voltage changes by ΔU_R as seen in Fig. 1.
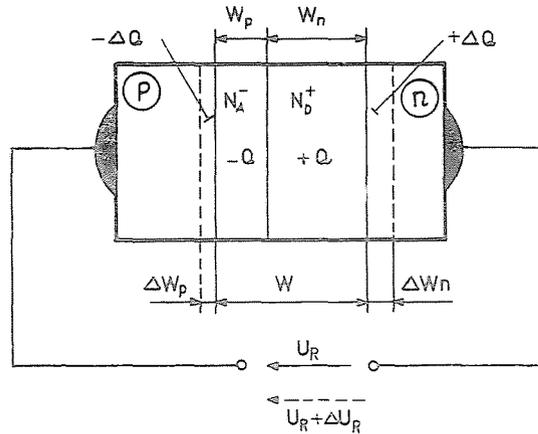

Fig. 1. Stored charges in the depletion layers of a $p-n$ junction

The well-known definition of the differential capacitance of the depletion-layer leads straightforward to the relationship:

$$C(U_R) = \varepsilon \cdot \frac{A}{w_p + w_n} = \varepsilon \cdot \frac{A}{w}. \tag{1}$$

Here $C(U_R)$ is the measurable depletion-layer capacitance, A is the area of the junction (edge effects are neglected for simplicity), ε is the permittivity of the semiconductor, and w_p, w_n, w are depletion-layer widths according to Fig. 1.

Since the charge equality must be always valid:

$$\Delta Q = A \cdot q \cdot N_D(w_n) \cdot \Delta w_n = A \cdot q \cdot N_A(w_p) \cdot \Delta w_p, \tag{2}$$

it is easy to show that:

$$\frac{dU_R}{dw} = \frac{q}{\varepsilon} \cdot N(w) \cdot w. \tag{3}$$

In this latter equation $N(w)$ is a combined doping-density of the donor and acceptor densities at the boundaries of the depletion-layers, as

$$N(w) = N_A(w_p) \| N_D(w_n).$$ (4.a)

Furthermore:

$$w = w_n + w_p.$$ (4.b)

The doping of the $p-n$ junction frequently shows a marked asymmetry, if e.g.:

$$N_A(w_p) \gg N_D(w_n).$$

then two approximations are valid:

$$w \approx w_n \quad \text{and} \quad N(w) \approx N_D(w_n).$$

Results for $N(w)$ have been published for many particular cases of faintly asymmetrical junctions [6].

Combining (1) and (3):

$$\frac{dC}{dU_R} = \frac{\varepsilon^2}{q} \cdot \frac{A}{N(w) \cdot w^3}.$$ (5)

Writing (5) and (1) in another form yields the fundamental equations of the $C-V$ profiling:

$$N(w) = -\frac{1}{q \cdot \varepsilon \cdot A^2} \cdot C^3 \cdot \left(\frac{dC}{dU_R}\right)^{-1}$$ (6.a)

$$w = \varepsilon \cdot A \cdot C^{-1}.$$ (6.b)

Both the capacitance C and its slope dC/dU_R can be measured electrically [7].

Although these measurements are in principle not difficult, many problems arise if these quantities become too small. As seen from (5) this is the case far from the junction at high doping densities. The nuisances are much greater if also the area of the junction is very small.

This instrument suits measurement of doping profiles satisfying the following inequalities:

$$N(w) \leqq 10^{19} \text{ cm}^{-3}$$ (7.a)

$$w \leqq 5 \ \mu\text{m}$$ (7.b)

$$\frac{N(w) \cdot w^3}{A} = 4 \cdot 10^{21} \frac{\text{cm}^{-3} \cdot \mu\text{m}^3}{\text{cm}^2}$$ (7.c)

further:

$$1 \leqq \frac{A}{A_0} \leqq 1000 , \qquad (8)$$

where $A_0 = 10^{-4} \mathrm{cm}^2$ is the unit area. Fig. 2 shows graphically these limitations.

Relationships (7.b) and (8) admit layer capacitances approximately between 0.6 and 600 pF.

Since the equipment has a presettable input for the semiconductor's permittivity, Si and GaAs structures can be investigated, too.
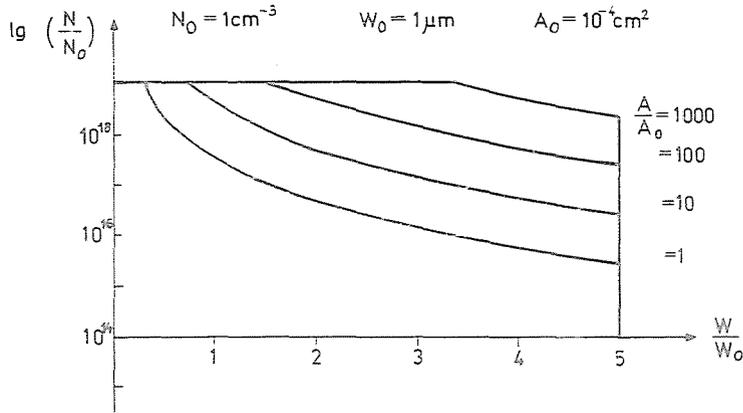

Fig. 2. Limits of the measurable structures

III. Some remarks on the measurement of capacitance and its slope

To extend the scope of measurable doping profiles in the critical direction, very small capacitance slopes i.e. very small capacitance changes have to be measured precisely.

A suitable measurement circuit for this purpose would be e.g. a symmetrical r.f. capacitance bridge, sketched in Fig. 3.

The short-circuited output current of such a bridge is proportional to the capacitance difference itself, while the open-circuited output voltage is proportional to the relative capacitance difference $\Delta C/C$.

Instead of these extreme terminations, merely approximations of them are realizable.

Taking the applied measurement frequency and the smallest junction capacitance to be measured into account (1 MHz, and 0.6 pF, respectively), it was easier to realize the condition:

$$R_{IN} \ll (2 \cdot C_0 \cdot \omega)^{-1} .$$

Since the junction investigated usually lies on a greater semiconductor wafer, it will be contacted under a microscope applying a tungsten needle.

Further electrical connection to the r.f. bridge is made through a coaxial cable. The impedance of this cable appears essentially in form of a relative high capacitive load.
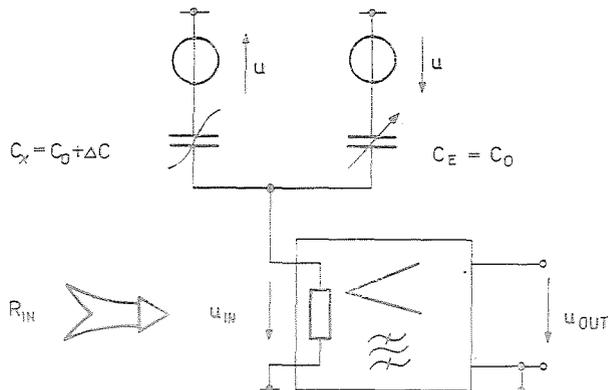

Fig. 3. Circuit diagram of a symmetrical capacitance bridge

Although this can be tuned out, the resulting resonance impedance is not high and stable enough to detect the open-circuited bridge voltage. The input impedance of the indicator r.f. amplifier—terminating essentially the bridge—was therefore chosen to be a low value, $R_{IN} = 180 \ \Omega$.

The remaining small phase shift caused by the cable was compensated for in the other arm of the bridge, with an appropriate coaxial cable.

Under these conditions the proportionality factor between the output voltage of the bridge and the capacitance difference ΔC will soon markedly depend on the instant value of the capacitance C_0, if this increases beyond about 100 pF.

This deviation from the exact linear response can, however, easily be taken into account.

Use of a lower R_{IN} impedance might seem to be better. But then, the decrease of the bridge output voltage could not be avoided, since the nonlinearities of the investigated junction do not allow to increase arbitrarily the r.f. supply voltage of the bridge. (Therefore, for r.f. supply voltage a practical value of 100 mV was chosen.)

Regarding the amplifier input noise voltage, which was about $2.1 \ nV \cdot s^{0.5}$ in our case, after optimizing the circuit, one can deduce a value of 300 nV as a noise-limited sensitivity. (The bandwidth is about 11 kHz.)

It is easy to show that the least value of the measurable capacitance-difference is about $\Delta C = 3$ mpF.

Since the highest peak to peak value of the biasing DC voltage variation (see later) is limited to 1 V, the smallest measurable capacitance slope is $3 \ mpF \cdot V^{-1}$. This value corresponds to (5) and (7.c).

IV. Functional description of the measurement

One can easily measure the slope of a voltage-dependent capacitance [7]

Regarding Fig. 3 it is obvious that the output r.f. voltage of the amplifier will be proportional to the capacitance slope if the capacitance difference ΔC is due to a small constant deviation of the biasing D.C. voltage from its value, which just equalizes the bridge.

To distinguish whether ΔC has a positive or negative value, the output r.f. voltage must be detected with a phase sensitive detector (e.g. with a sampling detector).

In our practical realization a low frequency (50 Hz) trapezoidal voltage U_T was superimposed upon the biasing D.C. voltage U_{R0} of the investigated junction.

As a consequence a trapezoidal voltage appears at the output of the phase sensitive detector; this voltage is symmetrical about zero, if the equilibrium biasing voltage U_{R0} just equalizes the bridge.

In other cases the D.C. component of the output voltage will indicate whether the controllable capacitance in the other arm of the bridge is greater or smaller than necessary.

This D.C. component can also be used to equalize the bridge automatically by means of a negative feedback loop. Thereafter the amplitude of the trapezoidal output voltage will be proportional to the capacitance slope.

Since the value of the capacitance slope itself varies in a very wide range, to avoid nonlinearities of the indicator amplifier, it is better to keep the *output* trapezoidal voltage on a constant reference value, controlling automatically the input trapezoidal voltage.

Regarding (5), it gives:

$$U_T = \text{const} \cdot \frac{1}{A \cdot \varepsilon^2} \cdot N(w) \cdot w^3 . \tag{9}$$

To eliminate the quantities A and ε, the gain of the r.f. amplifier must be set inversely proportional to the area A, and the reference voltage for the amplitude of the output trapezoidal voltage must be chosen proportional to the magnitude of ε^2.

Then:
$$U_T = \text{const} \cdot N(w) \cdot w^3 . \tag{10}$$

The other voltage, which is proportional to w, is obtained from a two-port network, having a transfer response proportional to $C_0(U_R)$ characteristic of the equalizing diode.

If the input voltage of this two-port is the automatic biasing voltage equalizing the bridge, and the output voltage is attenuated proportional to

$(\varepsilon \cdot A)^{-1}$, one obtains a voltage:

$$U_w = \text{const} \cdot w. \tag{11}$$

From the instant values of U_T and U_w an analog computing network produces the output voltages corresponding to $\lg N(w)$ and w. They control an X—Y recorder displaying graphically the dopant profile itself.

Some remarks must be made about losses in the bridge.

During the measurement the biasing voltage of the investigated junction is sweeping throughout between its presettable lowest and highest value.

The automatic biasing voltage of the equalizing capacitance always maintains its necessary value belonging to the instant dynamic equalization.

The losses varying differently in the two arms of the bridge should cause many problems. Consequently the losses must be continuously, automatically compensated for.
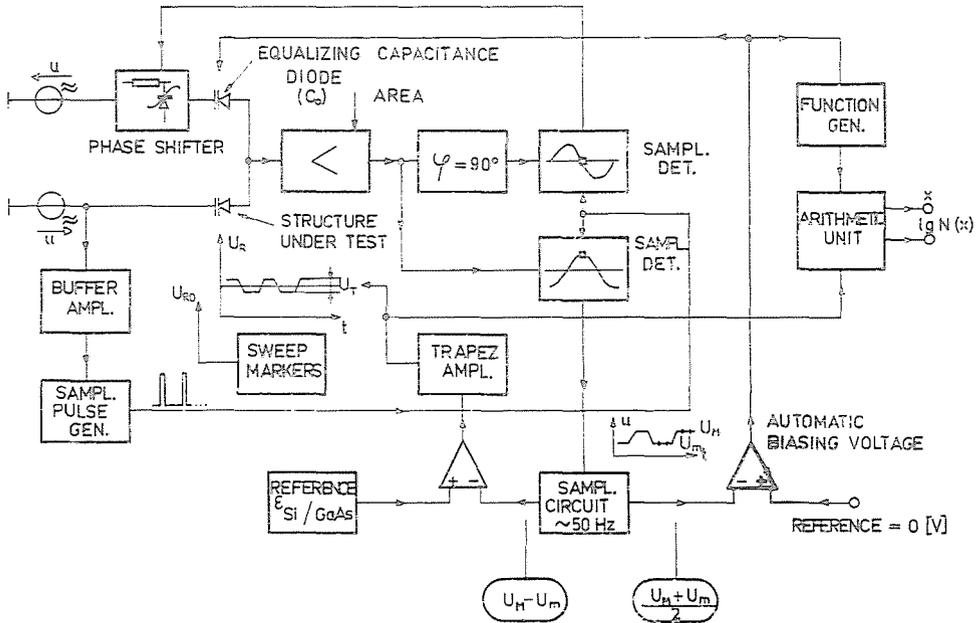

Fig. 4. Schematic diagram of the equipment measuring the doping distribution in a structure

For this reason, not only the direct output r.f. voltage has been detected but this has been done also after a phase shift by 90°. The reference sampling pulses of the phase-sensitive sampling detectors were the same.

This latter detection gives a voltage proportional to the sinus function of the phase mismatch, and so it can be used in a negative feedback, to control an electrically variable phase shifter placed in the reference arm of the bridge.

The control loops are designed to work dynamically; each one is compensated for with a PID network, and the loop gains are nonlinear to avoid latch-up effects.

A schematic diagram of the equipment is sketched in Fig. 4.

Summary

Doping profile is a very important technological parameter of almost all semiconductor devices. Its non-destructive measurement is generally based on the so-called $C - V$ method. Instead of the direct $C - V$ relationship, dC/dV vs. C gives, however, a more straightforward and accurate means of calculating the doping profile. The paper describes an instrument measuring directly $dC/dV = f(C)$ and discusses the practical limits of C, dC/dV and the area of the junction to be measured considering the sensitivity limit of the detector. The instrument is designed around three feedback loops for automatic compensations; one of them compensates for the losses varying during the measurement.

References

1. MÖSCHWITZER, A.: Parameter und Grenzwerte bipolarer Transistoren für hochintegrierte Schaltkreise. (Rep. I.-78.) Private communication
2. AMBRÓZY, A.: Periodica Polytechnica, Electrical Engineering Vol. 20. (1976.) No. 2. 141—155.
3. SZENTPÁLI, B.: Hiradástechnika, 29, 335—341., 1978.
4. SIMEK M. et al.: TÁKI Évkönyv, 1975. 227—244., 1975.
5. NAGASAWA, E. et. al.: Sol.-State El., 20. 207—513., 1977.
6. OPDORP, C.: Sol.-State El., 11. 397—406., 1968.
7. AMBRÓZY, A.: Sol.-State El., 13. 347—353., 1970.

Péter GOTTWALD
Prof. Dr. András AMBRÓZY   } H-1521 Budapest