UNIVERSITY OF SOUTHAMPTON

THE PROPERTIES OF GOLD IN DOPED SILICON

by

TERENCE FRANCIS UNTER

A thesis submitted for the degree of
Doctor of Philosophy

Department of Electronics
University of Southampton

July, 1977

UNIVERSITY OF SOUTHAMPTON

ABSTRACT

FACULTY OF ENGINEERING

ELECTRONICS

Doctor of Philosophy

THE PROPERTIES OF GOLD IN DOPED SILICON

by Terence Francis Unter

The properties of gold in doped silicon are discussed. A reconsideration of data concerning gold energy levels, their temperature variations and spin degeneracies leads to the development of models which describe the solubility of gold in heavily doped silicon and the effect of gold on the resistivity of silicon. A new model is proposed to explain the rapid gettering action of shallow high concentration phosphorus diffused layers. The dynamics of phosphorus gettering in silicon which has been either deliberately doped with gold or inadvertantly contaminated with gold are predicted.

Experiments on a number of structures diffused with gold in the presence of and absence of shallow phosphorus diffused layers are described. Gold concentration profiles are inferred from spreading resistance and Rutherford backscattering measurements. Important aspects of the proposed gettering model are confirmed and it is also shown that there is a strong link between the phosphorus-gold interaction and some anomalous diffusion effects associated with shallow, high concentration phosphorus diffusions. Total inhibition of the 'base push-out' effect in sequential boron-phosphorus diffusions is observed when gold is diffused simultaneously with the phosphorus.

The use of spreading resistance measurements for dopant versus depth profiling is studied. Optimization of equipment operation in conjunction with a newly developed bevelling and sample preparation technique yields measurements with high spatial resolution and excellent reproducibility.

Acknowledgements

Many people have assisted and advised me during the course of the research described herein. I would like to extend my thanks to all of them, particularly those named below.

Dr. D. R. Lamb for conscientious and patient supervision of this work. Dr. P. L. F. Hemment of the University of Surrey for carrying out the Rutherford backscattering experiments. Dr. A. F. W. Willoughby for his continued interest. Dr. S. D. Brotherton, now at Mullard Research Laboratories, for guidance in the early stages of the project. Dr. M. Brown and Dr. C. L. Jones of the Department of Mechanical Engineering for numerous stimulating discussions and much inspiration. Dr. P. C. T. Roberts for his critical comments on several aspects of the work. Mr. A. Frier for assistance with experiments on the 'push-out' effect and Mr. J. Pennock for improving my mathematics. Mrs J. Wright-Green for typing this thesis in such a short time.

I would like to acknowledge financial support from the Science Research Council.

CONTENTS

1. INTRODUCTION

The diffusion of gold in silicon is a process of great technological importance which has been in use for many years. Despite its significance in planar device manufacture - both as a deliberately added dopant and as an unwanted contaminant - there are many unanswered questions concerning its properties and diffusion behaviour.

The importance of gold arises from its introduction of two deep lying energy levels into the silicon energy band gap. These energy levels, one of which is a donor and one of which is an acceptor, cause gold to act as an efficient recombination-generation centre in both p-type and n-type silicon.

Gold diffuses into silicon very readily. This property combined with its behaviour as a recombination centre has led to wide-spread use of gold as a minority carrier lifetime "killer" in high speed bipolar switching devices. Such devices generally operate in saturation mode, that is to say while switched 'on' the emitter and collector junctions are both forward biased and there is a high density of minority carriers in the base region. The speed with which the device can switch from the 'on' state to the 'off' state is largely determined by the time taken for the excess minority carriers in the base region to recombine. This 'storage time' can be made much shorter by adding gold to the base region of the device where, by killing the minority carrier lifetime, it reduces both the amount of charge stored and the recombination time. Although recent development of Schottky clamped switching circuits has reduced the use of gold doping for low power, fast logic circuits, its use in high power switching devices (e.g. diodes, thyristors and triacs) is of continuing importance (Miller, 1976).

More recently gold has been proposed as a suitable dopant for extrinsic silicon MOS and CCD infra-red photo-sensitive devices (Parker and Forbes, 1975 ; Logan, 1976) since the gold donor level (at $\simeq$ 0.35eV above the valence band edge) provides response in the 3-5 micron wave band.

Undesirable aspects of the efficiency of gold as a recombination-

generation centre are also of great importance. Gold is one of the most ubiquitous contaminants in high temperature furnace tubes and in chemicals commonly used in semiconductor technology. This, coupled with the ease with which it diffuses into silicon, causes great problems in the production of devices in which long minority carrier lifetimes - and hence low leakage currents - are required. It is very important, therefore, not only to be able to diffuse gold in to some devices deliberately but also to be able to remove it when it has been introduced inadvertantly.

Despite the wide application of gold diffusion for the purposes mentioned above and the necessity of preventing or removing unwanted gold contamination, the processes used by many device manufacturers are still, to a great extent, the result of an empirical approach to obtaining the desired properties in a given device.

Recent studies of the recombination-generation rates of the gold levels (e.g. Tasch and Sah, 1970) and the diffusion and solubility of gold in silicon (Huntley, 1972; Brown, 1976) have furnished much new information about the parameters of the mechanisms involved. However, one of the most problematic aspects of the diffusion process is the interaction between diffusing gold and shallow diffusions of other commonly used dopants - particularly phosphorus. There have been no systematic studies of the dynamics of this interaction although its occurrence has been known since Adamic and McNamara (1964) reported that diffused gold tended to accumulate in heavily phosphorus doped areas of silicon wafers. This phenomenon furnished two results, one desirable and one undesirable.

The most useful aspect of the preferential accumulation of gold in heavily phosphorus doped silicon is in the so-called gettering effect. Unwanted gold contamination is removed, or gettered, into a phosphorus diffused layer in parts of silicon wafers remote from active devices in which long minority carrier lifetimes are required. The undesirable aspect of gettering occurs during the deliberate gold doping of switching devices. Phosphorus diffusions are generally used to form n^+ layers (such as emitters in bipolar n-p-n transistors) and gold tends to segregate into these regions. The result can be that the region in which the lifetime killer is actually required (the base-collector junction region

for a saturation mode bipolar switch) can be starved of gold.

There is clearly an area for investigation here and in this thesis a theoretical and experimental study of the gold-phosphorus interaction is reported. The results provide new insights into the dynamics of the phosphorus gettering effect which should be valuable in the development of process models describing technologies in which deliberate gold addition or gold removal is required.

With the increasing complexity of planar processing the need to model and predict processes, rather than arrive at them empirically, is becoming paramount (Meindl, Saraswat and Plummer, 1977). In order to model the behaviour of gold in silicon, its electrical effects as well as its diffusion behaviour need to be understood and predictable. A thorough reconsideration of the data available concerning the positions of the gold energy levels, their temperature dependencies and spin degeneracies is also made in this thesis. The results of this are used in the development of a new model describing the gold-phosphorus interaction and also in a model for the prediction of the effect of gold on the resistivity of silicon already doped with moderate (up to 10^{17} atoms/cc) quantities of n-type or p-type dopants.

Diffusion profiles of gold in silicon wafers have been measured with the spreading resistance technique and by Rutherford backscattering. The interactions between gold diffusions and shallow high concentration phosphorus diffusions thus studied have yielded new information about the dynamics of the phosphorus gettering process. An interesting side effect of the phosphorus-gold interaction which has been observed, provides both confirmation of one of the important aspects of the phosphorus gettering model proposed in this work and new information on the so-called 'push-out' effect which occurs in sequential boron-phosphorus diffusions (Jones, 1976).

A substantial proportion of the period occupied by the research described here was spent establishing the method used to obtain the spreading resistance measurements. This investigation, which is considered to be an important part of this work, is described in appendix A.

A summary of the topics discussed in this thesis is given in the following section.

2. SUMMARY

The properties of gold in doped silicon which are considered in this thesis may appear, at first sight, to range over a number of almost unconnected topics. A brief summary at this stage should enable the reader to view the work in a more integrated form.

The various subjects to be considered are all connected by the close link between the diffusion behaviour of gold -- particularly its solubility -- its electrical properties and its electronic interactions with other dopants. Because the gold energy levels lie deep in the silicon energy band gap, any electrical effect due to significant amounts of gold in silicon is very sensitively dependent on the precise positions of the energy levels and any variation in them.

An attempt is made, throughout this work, to review appropriate subject matter from the literature. Emphasis is laid on obtaining new information and drawing new conclusions from the available data in the light of more recent knowledge.

Experimental studies presented in chapter 8 shed new light on several aspects of the diffusion of gold in silicon both in the presence of and absence of shallow, high concentration phosphorus diffused layers. The dynamics of phosphorus diffusion gettering of gold at 1000°C are studied and the rate limiting step at 1000°C is identified. A result of particular interest is the observation that the base push-out effect, which commonly occurs in bipolar devices with phosphorus emitter diffusions, is inhibited by the simultaneous diffusion of gold with the phosphorus. The implications of this and other results are discussed in chapter 9.

A new model for the observed gettering effects -- based on results obtained in this work and results obtained from an appraisal and reanalysis of data available in the literature -- is developed in earlier chapters of the thesis. A number of properties of gold in silicon (its diffusion mechanism, the positions of the gold energy level positions at low and high temperatures and its solubility) are of importance in the

prediction of the gettering effect and are discussed as follows:

The diffusion mechanism which is considered to be appropriate to gold in lightly doped silicon is described in chapter 3. The discussion does not extend to the diffusion behaviour of gold in heavily doped silicon since this topic has not been well studied in the past and receives detailed attention later in this thesis. It will be seen that, although the basic diffusion mechanism is well established some of the parameters which determine the rate limiting process are still in dispute.

In chapter 4, a very detailed discussion of the gold energy levels and their interrelation with impurity level spin degeneracy is presented. The importance of knowing the gold energy level positions very accurately is stressed and consideration is given to the possible variation of the energy level positions relative to the band edges as temperature varies. It is only recently (Parrillo and Johnson, 1972) that the importance of including this effect in measurements of energy level positions has been recognised. The wealth of data on gold energy levels which is available in the literature is reconsidered in the light of possible temperature variations. It is shown that not only do apparently conflicting results agree when analysed correctly but also that the usual values employed for the gold energy level positions at room temperature (300°K) are probably quite inaccurate.

The spin degeneracies which may be associated with the gold energy levels are then considered in terms of a simple bonding model for substitutional impurities in semiconductors which was proposed by Teitler and Wallis (1960) and extended to gold in silicon by Brown (1976). Once again the relevant literature on this subject -- which is shown to be very confused -- is reviewed.

Based on recent measurements and a reappraisal of earlier ones, a model for the gold energy levels, their temperature dependences and degeneracies is chosen as the most likely. This model is used later in the thesis in discussions of the properties of gold in doped silicon and is compared with other possible models. It is shown that this model is capable of predicting the effect of heavy doping on the solubility

of gold and the effect of gold on the resistivity of silicon with good accuracy.

A very brief consideration of the behaviour of gold as a recombination-generation centre is given in chapter 5. No new information is presented on this topic but it is included for completeness since it has considerable relevance to some of the energy level measurements described in chapter 4.

The solubility of gold in silicon previously doped with shallow level impurities is discussed in chapter 6. The possible mechanisms which may contribute to the enhanced solubility of gold observed in uniformly heavily doped silicon (particularly n-type Si) are reviewed. Calculations based on the models discussed in chapter 4 are carried out and it is shown that the experimentally observed behaviour reported in the literature may be described without making any special additional assumptions about the gold energy levels or gold charge states. Previous attempts at predicting gold solubility have generally ignored the possible temperature variations and degeneracies of the gold energy levels and have had to resort to largely unjustifiable assumptions. In addition, most authors have made approximations which can cause large differences in the final predictions. The calculations here are carried out for a number of models of gold energy levels etc. in order to illustrate how dependent the final values are on the precise energy level positions.

The second part of chapter 6 is devoted to a discussion of gettering effects, particularly that due to shallow phosphorus diffusions. The solubility enhancement model developed for uniformly doped silicon is extended to the behaviour in shallow, high concentration phosphorus diffused layers. The dynamics of the phosphorus gettering process are considered and a new model of the interaction of gold and phosphorus is proposed. In this model it is suggested that the mechanism thought by many authors to explain the anomolous diffusion behaviour of shallow, high concentration phosphorus diffused layers (the phosphorus-vacancy pair of E-centre) may also be used to account for the greatly enhanced solubility of gold in phosphorus layers and the speed with which it is gettered into them. Equations are derived which should allow prediction of the effect of a phosphorus diffused layer on the distribution and amount

of gold in a silicon wafer which is either being deliberately gold-doped or from which gold is to be removed. The results which were obtained in this work are shown to be consistent with the model developed in this chapter.

Chapter 7 describes calculations of the effect of gold on the room temperature resistivity of silicon which is lightly doped with either p-type or n-type dopants ($<10^{17}$ atoms/cc). Once again the models discussed in chapter 4 are all considered and the effects of small variations in the gold energy level positions and degeneracies on the predicted resistivities are shown to be considerable. Although a good fit to the available experimental data is not possible because of uncertainty in what proportion of the total gold concentration is electrically active, the closest correlation is shown to occur for the same model of energy levels and degeneracies that gives the best fit to the gold solubility predictions and was chosen as the most likely in chapter 4.

As already mentioned, experimental work is described in chapters 8 and 9. Gold diffusion profiles in bulk silicon are inferred from resistivity profiles obtained with the spreading resistance technique by angle bevelling the samples to expose the regions of interest. A substantial period was spent setting up and optimising the operation of the spreading resistance apparatus and in developing a new surface preparation and bevelling technique for the samples. This work, which is described in the first appendix, has resulted in spreading resistance measurements which, in terms of spatial resolution, accuracy and reproducibility, are considerably better than state-of-the-art measurements reported in the literature. Such performance was necessary for the measurements described in chapter 8 and the description has been placed in an appendix so as not to interrupt the logical presentation of the properties of gold which are considered rather than to de-emphasize its importance.

It is hoped that the interrelation between the topics discussed in this thesis is now apparent to the reader and that, as a result, it may be more easily read.

3. THE DIFFUSION MECHANISM OF GOLD IN SILICON

Despite a number of studies since the first report of an experiment designed to measure the diffusion coefficient of gold in silicon (Dunlap et al. 1954) there are still a number of aspects of the diffusion process which are ill understood and in dispute. Of particular interest in this thesis, is the interaction of gold diffusion and phosphorus diffusion in thin silicon wafers. The gold diffusion mechanism is reviewed briefly in this chapter in order to provide the necessary background to the theory and experiments which follow.

The early studies of gold diffusion (Dunlap, Bohm and Mahon, 1954 , Struthers, 1956 and 1957; Trumbore, 1960; Boltaks, Kulikov and Malkovich, 1961) resulted in differing diffusion coefficients and , in the case of Boltaks et al., scatter in the results from sample to sample. This scatter was attributed to "structural defects" in the silicon samples.

Dash (1960) observed non-conservative motion of dislocations in silicon wafers during the diffusion of gold. The dislocation climb, which was in such a direction as to cause emission of vacancies, was taken to indicate that the diffusing gold had caused an undersaturation of vacancies in the silicon samples. This result led Dash to suggest that gold diffused by the "dissociative mechanism" which had been proposed by Frank and Turnbull (1956) for the diffusion of copper in germanium. Since this proposal, a number of works have confirmed that the dissociative mechanism is almost certainly applicable to the diffusion of gold in silicon. This mechanism is outlined briefly below and the indications of various results available in the literature are then discussed.

3.1 The dissociative diffusion mechanism

The dissociative mechanism is, in essence, very simple. The diffusing atoms move in two fashions within the silicon lattice: interstitially and substitutionally. Conversion from an interstitial to a substitutional site is possible when an interstitial atom and a lattice vacancy combine. In practice the diffusion rates of interstitial and substitutional atoms are very different and so are their solubilities. Due to the disparate

diffusion rates, the local ratios of interstitial atoms A_i to substitutional atoms A_s will be upset so that a kinetic process to reestablish equilibrium will occur according to the reaction:

$$A_s \underset{k_2}{\overset{k_1}{\rightleftharpoons}} A_i + V \qquad 3.1$$

where V is a vacancy in the silicon lattice.

If the interstitial diffusivity is much higher than the substitutional diffusivity, in a region of high concentration

$$\frac{A_i}{A_s} < \frac{(A_i)_e}{(A_s)_e} \quad , \qquad 3.2$$

where the subscript 'e' denotes the equilibrium concentration. In a region of low concentration :

$$\frac{A_i}{A_s} > \frac{(A_i)_e}{(A_s)_e} \qquad 3.3$$

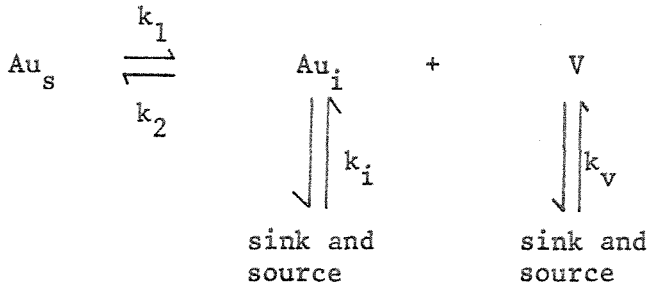
Clearly in the high concentration region, the reaction described by 3.1 takes place in the k_1 direction and the diffusion is truly "dissociative" in that it proceeds by the dissociation of a substitutional atom into an interstitial atom and a vacancy. In the region of low concentration the reaction takes place in the k_2 direction.

It is generally assumed that in the case of gold in silicon, a high interstitial diffusion rate is combined with a low interstitial solubility and that the low substitutional diffusion rate is combined with a high substitutional solubility. The result is that the diffusive flux is carried almost entirely by interstitial atoms, Au_i, and the measured concentrations of gold generally comprise substitutional atoms, Au_s. The actual disparity in solubility between the substitutional and interstitial species of gold has not really been assessed satisfactorily and is discussed again later in this chapter.

The diffusion profiles which may result from this process will be complex and will depend on a number of factors. Sturge (1958) expanded the Frank-Turnbull theory and put it on a sounder mathematical basis. This

theory has been further extended by Huntley and Willoughby (1970 and 1973). The detailed mathematical analyses which result from these theoretical studies will not be entered into here, since a qualitative understanding of the various processes which may occur is sufficient in this work.

The reaction described in 3.1 may be viewed in greater detail :



3.4

where k_i and k_v are the reaction constants linking interstitial atoms and vacancies to their respective sources. In the case of interstitial gold atoms the 'source' is the diffusion source of gold. This is usually a layer of elemental gold plated or evaporated onto one surface of the sample and as such generally constitutes an infinite source. For vacancies the situation is a little more complex and gives rise to one of the main areas of contention in the modelling of gold diffusion processes. Vacancy sources may be the sample surface (which is usually considered to be an infinite vacancy source) or may be some reaction within the wafer. An example of the latter source is the generation of vacancies by climbing dislocations observed by Dash (1960) and modelled by Huntley and Willoughby (1973).

Differential equations describing the continuity of the processes involved in 3.4 may be obtained. The general solutions of such equations are difficult to obtain and give little conceptual insight to the problem; however, if one of the reactions in 3.4 is considered to be the rate limiting step of the whole process, useful solutions may be obtained:

The possible rate limits are :

(i) the reaction rate between Au_i and V described by k_1 and k_2

(ii) the generation rate of vacancies in the sample bulk or the diffusion rate of vacancies from the sample surface described by k_v.

(iii) the diffusion or formation rate of interstitial atoms described by k_i.

A combination of these rate limits may apply.

The diffusion rate of interstitial gold is very high (shown by the rapidity with which gold diffuses through thin silicon wafers - Sprokel et al. 1965 - see below) and will only be a limiting process if k_2 is very fast and there is an adequate supply of vacancies or if the sample is very thick. The nature of the other possibilities has been the subject of studies by several authors (Wilcox and La Chapelle, 1964 ; Sprokel, 1965 ; Malkovich, 1968 ; Yoshida, 1969; Yoshida and Saito, 1970; Yoshida, 1973; Huntley and Willoughby, 1970, 1973 i and ii; Brotherton and Rogers, 1972).

3.2 Studies of gold diffusion in thick silicon samples

Wilcox and La Chapelle (1964) carried out the first comprehensive set of experiments on the diffusion mechanism of gold and interpreted their results in terms of the Frank-Turnbull dissociative mechanism. Radiotracer gold was diffused into thick silicon specimens over a range of temperatures and for a range of times. The samples had "low" and "moderate" dislocation densities.

The diffusion profiles showed a complex dependence on temperature and dislocation content. The most common occurrence was a kinked profile of the type shown in figure 3.1. The data was analysed by assuming that the kinked profile was in fact made up of two ideal complementary error function profiles (see section 6.5) which normally result from a Fickian diffusion process from an infinite source (see, for example, Grove, 1967 p 42 ff). Error function complement curve-fits were made to the two portions of the profile and the values of diffusion coefficient, D, so obtained were plotted as the Arrhenius relationship log D vs 1/T. The result was that Wilcox et al. were able to fit three different lines to various of the data points. The lines are shown in figure 3.2 although the data points are not included. These diffusion coefficients, when related to the diffusion profiles and material used, were interpreted as:

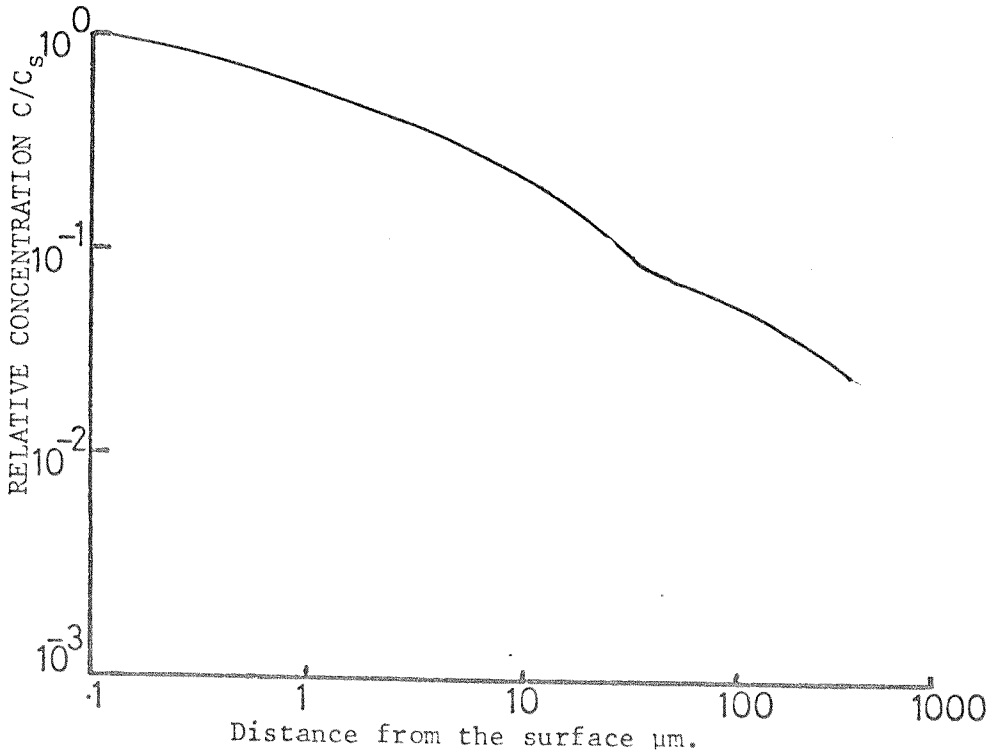


FIG. 3.1: Typical profiles for Au diffusion into thick Si specimens at 1000°C (Wilcox et al. 1964).

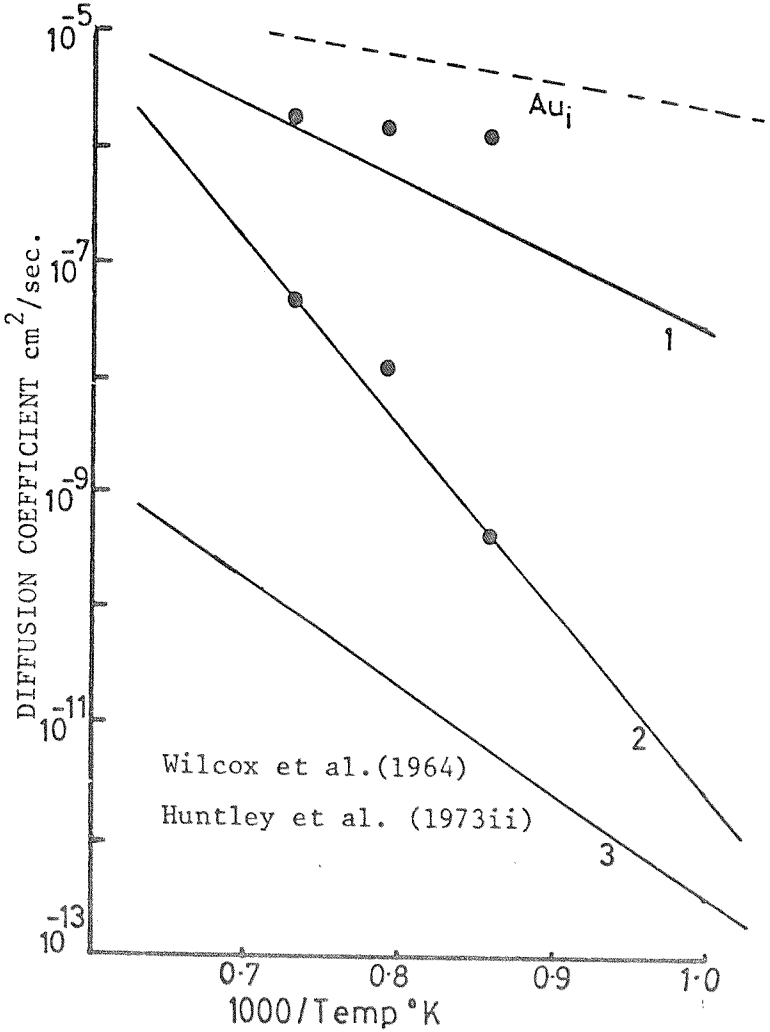


FIG. 3.2: Diffusion coefficients for Au in Si.

(1) D_1 = Gold interstitial diffusion controlled dissociative diffusion

(2) D_2 = Vacancy controlled dissociative diffusion

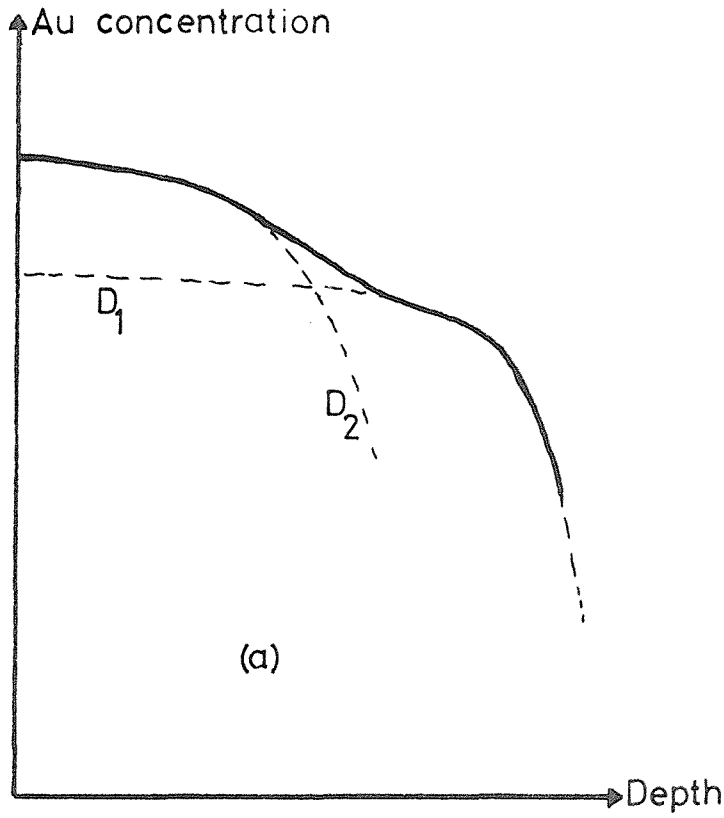
(3) D_3 = Pure gold substitutional diffusion.

(Note that the numbers on these curves are not the same as those used in the original figure in Wilcox et al's paper). D_1 and D_2 are the diffusion coefficients predicted theoretically by the Frank-Turnbull model.

Data from previous authors (cited earlier in this chapter) were shown to be on either the D_1 or the D_2 curves depending on the experimental conditions and silicon material quality (it being assumed that dislocations in some samples provided a source of vacancies). Silicon with a high dislocation content gave rise to points lying on the D_1 line and silicon with a low or zero dislocation content gave rise to points lying on the D_2 line.

The results of Wilcox et al., although valuable, are open to certain criticisms. Huntley et al. (1973 ii) point out that the technique used for the measurements is not considered to be reliable by most radiotracer workers. In addition all of the profiles presented are normalised to the surface concentration making the precise boundary conditions rather uncertain. For example, it is not known if the gold surface concentration was always maintained at the solubility limit (i.e. was the gold source always infinite?). The effect of varying the dislocation density was not assessed quantitatively.

Huntley et al. (1973 ii) also carried out gold diffusion experiments in thick silicon samples. Gold profiles were measured in dislocated and dislocation free silicon wafers and were found to be similar in shape to those reported by Wilcox et al. (see figure 3.1). Their analysis was based on the Frank-Turnbull model with account taken of vacancy generation by the grown-in dislocations. The predictions showed that the part of the profile close to the surface should have an error function complement shape controlled by D_2 (i.e. vacancy controlled dissociative diffusion) and that the tail of the profile should also have an error function complement shape controlled by D_1 (i.e. interstitial controlled dissociative diffusion). Huntley et al. also predicted that the transition region between the surface and the tail of the profile should depend on the



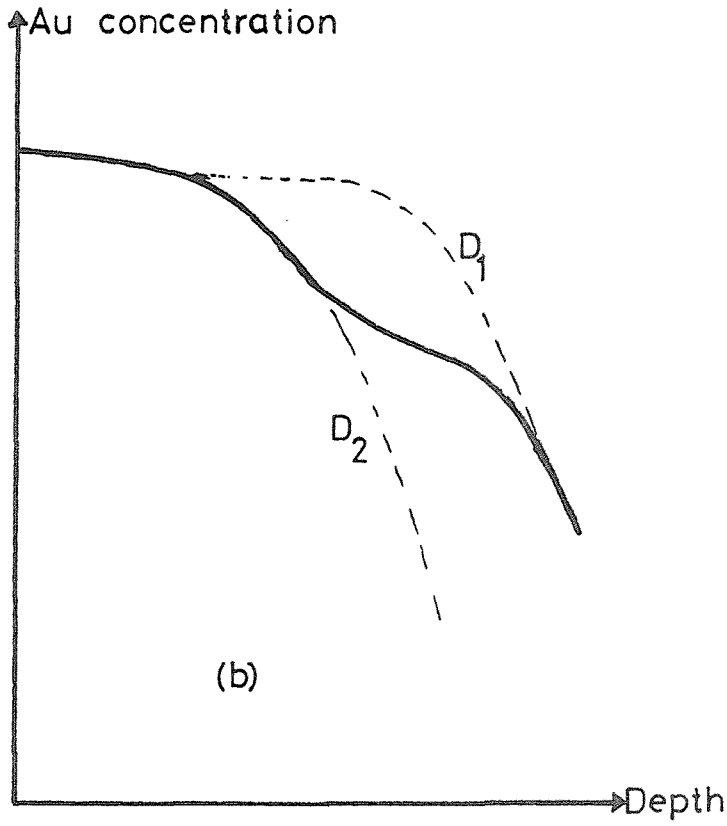


FIG.3.3: Methods of fitting erfc. shapes to Au profiles in thick Si specimens to obtain D_1 and D_2
(a) Wilcox et al.(1964).
(b) Huntley et al. (1973ii).

dislocation density (or, by inference from this, the internal vacancy generation rate) as well as time and temperature.

There was one significant difference between the analysis of Huntley et al. and that of Wilcox et al. Wilcox et al. assumed that the tail of the profile shown in figure 3.1 was interstitial gold only. The error function complement profile which was fitted to the tail was extrapolated back to the surface to give the interstitial gold solubility as shown in figure 3.3 (a). The surface concentration was an adjustable parameter used to obtain the best fit to the tail profile. The D_1 and D_2 profiles were, therefore, extrapolated to different surface concentrations. Indeed, the data obtained in this way for the interstitial gold solubility is the only such data available and if Huntley et al's analysis, described below, is correct, considerable doubt arises about the use of the values obtained by Wilcox et al.

Huntley (1972) argues that, if the equilibrium concentration of substitutional gold is greater than the equilibrium concentration of interstitial gold then radiotracer profiles will yield the substitutional gold distribution only even though the atoms may have entered the crystal interstitially. In addition the reaction expressed in 3.1 "always maintains the substitutional concentration > the interstitial concentration" if the rate is high. Huntley et al. analyse their data on the assumption that the tail of the profile is not interstitial gold but substitutional gold the concentration of which is controlled by the interstitial diffusion rate. The error function complement fits to the surface and bulk parts of the profiles are both extrapolated to the same surface concentration; that of substitutional gold, see figure 3.3 (b). The D_2 values agree well with those obtained by Wilcox et al. but the D_1 values do not. These values are plotted as points on figure 3.2. The identification of these two effective diffusion coefficients is the strongest evidence for describing gold diffusion in silicon by the dissociative mechanism.

If Huntley's assertion that 'the ratio of substitutional to interstitial gold is maintained at the equilibrium value' is correct, then the values commonly used for the interstitial gold solubility limit are very dubious. There appears to be no consensus on this point amongst authors who have considered the dissociative process. However, there is good, although indirect, evidence to suggest that the solubility limit of Au_i is much less than that of Au_s. Such evidence is mainly inferred by comparison of the

behaviour of gold in silicon on cooling from diffusion temperatures with that of other fast diffusing metals (e.g. copper, and iron). During cooling, large fractions of the total amounts of these metals precipitate onto lattice defects, gold does not behave in this way. Precipitation suggests that most of the atoms of the species in question are in a mobile (interstitial) state. The absence of precipitation of gold in silicon suggests that, at the diffusion temperature, the gold is on substitutional sites and therefore relatively immobile.

3.3 The diffusion profile of gold in thin silicon wafers

The diffusion of gold into thin silicon wafers is of greatest importance since in most cases of practical interest silicon planar devices are manufactured on such thin wafers. Unfortunately the presence of shallow diffusions of other dopants can modify the distribution of diffused gold drastically; it is such situations that are of main interest here since real device processes are based on shallow diffusions of various dopants. The behaviour of diffusing gold in thin silicon wafers is only discussed briefly here since it is dealt with in detail in chapters 6, 8 and 9.

Sprokel et al. (1965) measured the first gold profiles in thin silicon wafers. All of the profiles they obtained were characterized by a nearly horizontal central region, with an increased gold concentration on the side to which the gold was diffusing as well as that on which it was deposited. A profile typical of their results is illustrated in figure 3.4. Sprokel (1965),analysed the dissociative diffusion mechanism in thin wafers with particular reference to the results described in the earlier paper. The analysis was based on the assumption that the rate of rise of gold concentration in the centre of the wafer was determined by the reaction rate constant k_2 in equation 3.1. Similar experimental results have been obtained by Martin, Haas and Raithel (1966);Yoshida and Saito (1970); Brotherton and Rogers (1972); Huntley and Willoughby (1973 i) and in this thesis, chapter 8. Sprokel et al. suggest that the rear surface "tip-up" is due to vacancy generation at that face (the amount of tip-up was found to be greater if the back was less well prepared -- i.e. contained lattice imperfections. This could have been due to precipitation however - see chapter 6). Martin et al. attributed the tip-up on the wafer face opposite

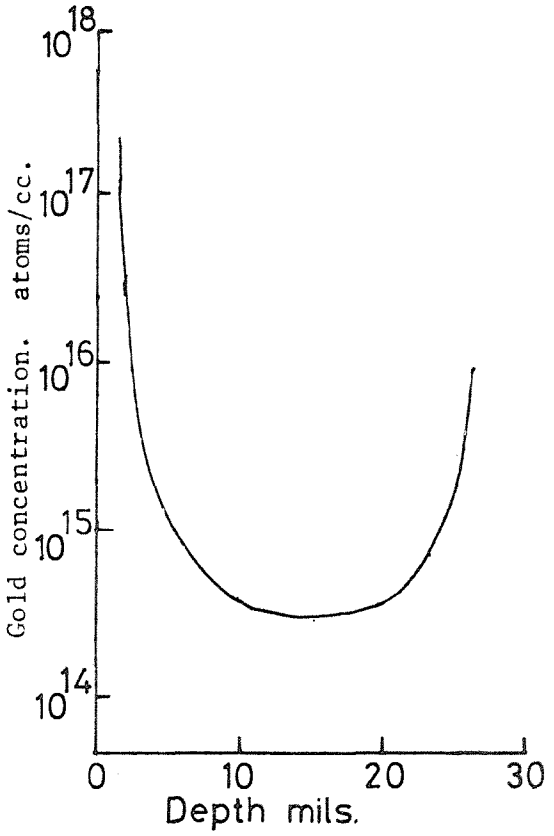


FIG.3.4: Gold profile in thin Si wafer after 30 mins diffusion at 1000°C. (Sprokel et al., 1965).

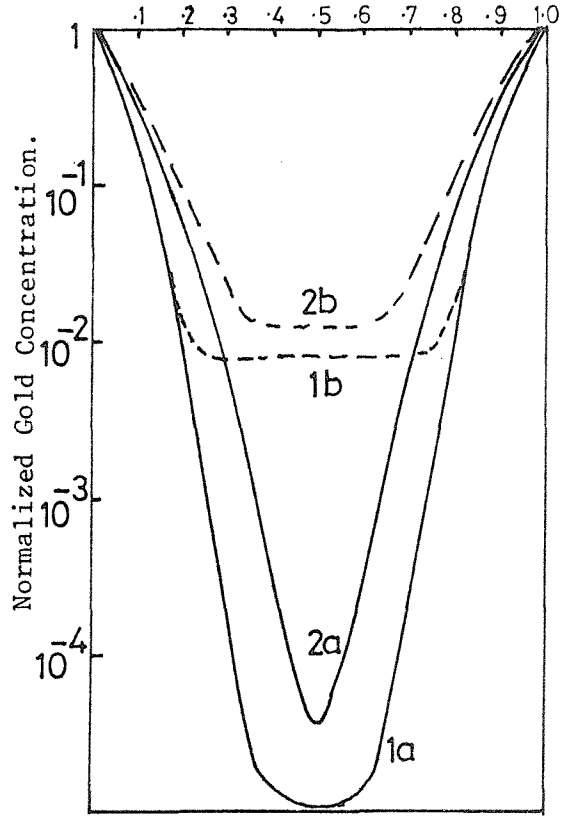


FIG.3.5: Calculated diffusion profiles of Au in thin Si wafers Huntley et al. (1971).

the gold source to surface diffusion of gold. Huntley and Willoughby (1971) have shown, however, that surface diffusion is inhibited by an oxide layer on the wafer but the opposite face tip-up is not.

Several attempts, since Sprokel (1965), have been made to explain the rate of rise of gold concentration in the centre of the thin wafers. In common with Sprokel, Malkovich (1968) assumes that the process is rate limited by the reaction of interstitial gold with vacancies and that the rate of gold diffusion by the interstitial mechanism is fast enough for the interstitial concentration to remain constant. Yoshida (1969) establishes criteria for the assumption that the thermal equilibrium of interstitial gold atoms or lattice vacancies is maintained. These criteria involve the relative rates of the various reactions described by equation 3.4 and produce no surprising results (e.g. k_v, must be greater than k_2 if the vacancy concentration is to remain in equilibrium). The results take no account of concentration gradients but are of interest in thin specimens where diffusants with high diffusion coefficients would be expected to establish flat profiles.

Yoshida et al. (1970) apply the criteria to experimental results for gold diffusions in silicon. The rate of increase of gold concentration with time is explained by the generation of vacancies from fixed and variable sources (e.g. climbing dislocations, dislocation loops etc.). Gold concentration profiles were not measured however, only average gold concentrations were obtained by counting large parts of the samples after neutron activation. It is not known, therefore, if the concentration profiles were flat so the presence of unknown concentration gradients could affect the conclusions.

Huntley and Willoughby (1970) presented a new theoretical analysis of the diffusion of gold in thin silicon wafers. Based on the model established by Sturge (1958), a term describing the generation of vacancies from jogs on climbing dislocations was used in the dissociative diffusion expressions. The resulting equations were solved for two cases: defect free material in which the only source of vacancies was the wafer surface and dislocated wafers in which the vacancies were emitted by dislocation climb.

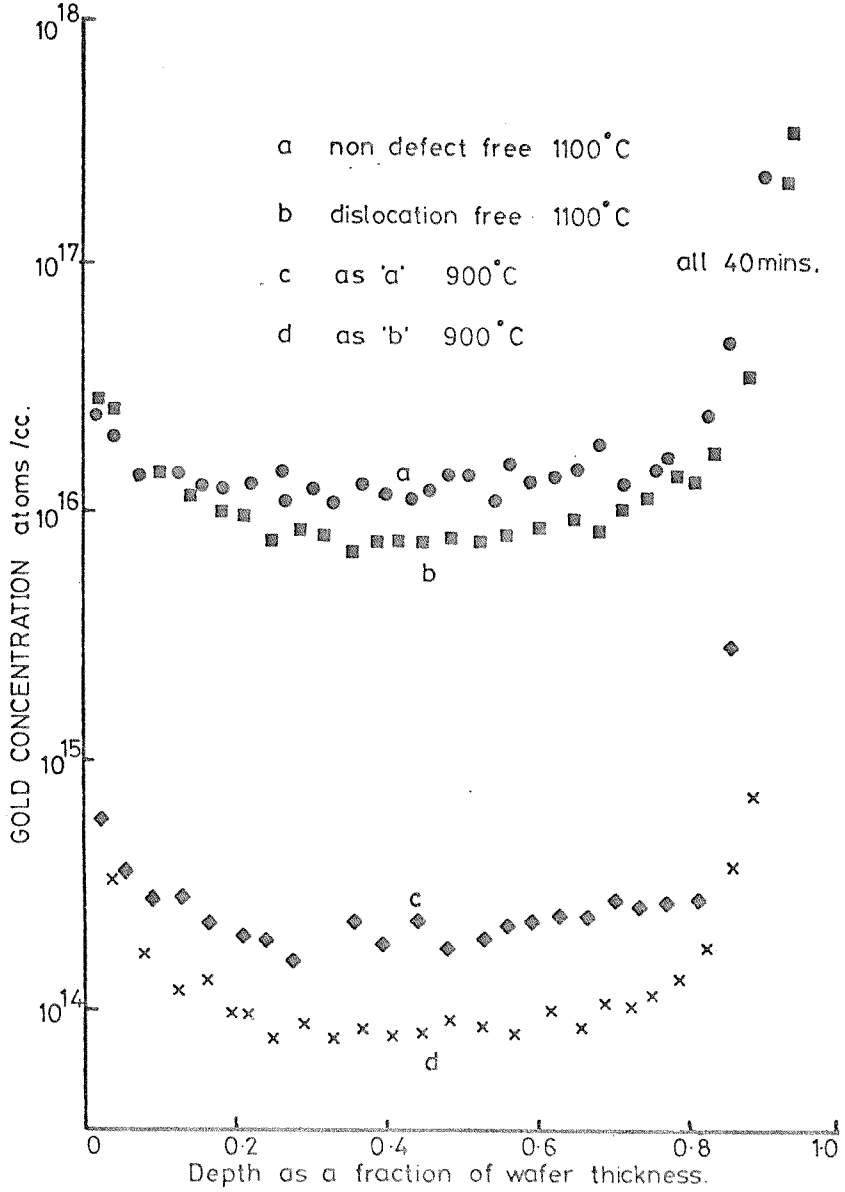


FIG. 3.6: Gold profiles in thin silicon wafers measured by Huntley et al.(1973i).

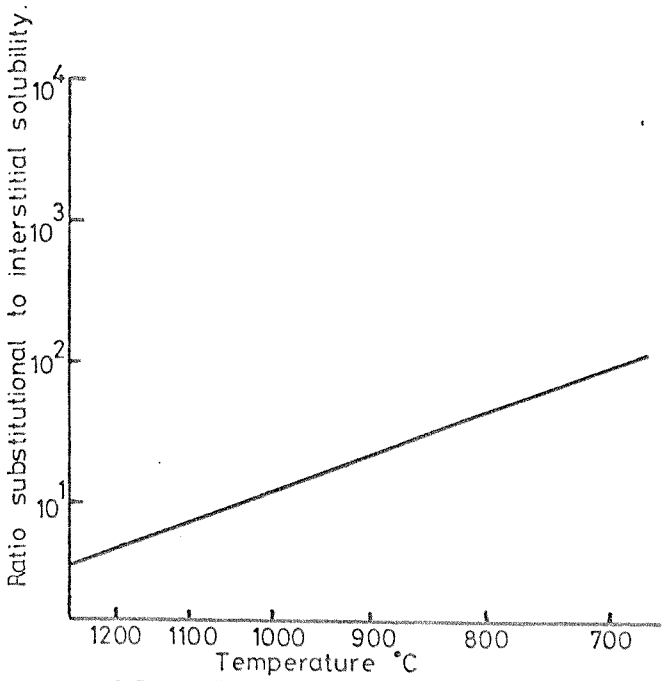


FIG.3.7: Ratio of substitutional to interstitial gold solubility from the data of Wilcox et al.(1964).

Figure 3.5 compares the basic shapes of the predicted profiles. It was assumed that the reaction rate of $Au_i + V$ (i.e. k_2 in equation 3.4) was very high and that the whole process was limited by the vacancy supply. As one might expect, the former case results in gold profiles which mirror the vacancy diffusion process from the wafer surfaces whereas the latter case results in flat diffusion profiles.

Brotherton and Rogers (1972) measured the rate of rise of central concentration in dislocated and dislocation free wafers and noted that the effect due to the vacancies was not very great - in contradiction to the Huntley et al. model. Brotherton et al. concluded that the rate of rise was limited by the reaction rate of $Au_i + V$ and not the vacancy supply although no explanation of the small effect of the dislocations was given.

More recently, Huntley et al. (1973) published results of their own comparisons of gold diffusion profiles in thin dislocated and dislocation free silicon wafers. Significant differences between the profile shapes in the two materials are noted, but they are not at all like the predicted shapes. Figure 3.6. illustrates some of these profiles. This result may support the conclusion that the profile shape is determined by the interstitial gold - vacancy reaction rate or may indicate, as suggested by Huntley et al. that there are sources of vacancies in the wafer bulk, other than climbing dislocations. Recent availability of silicon crystals which are much more 'perfect' than those used by Brotherton et al. and Huntley et al. has prompted a repeat of the experiments described above on a limited scale in this work. The results, along with a more detailed discussion of the possible vacancy sources, are described in chapters 8 and 9. The indications are that the Huntley et al. model describes the results obtained here better than the reaction rate limited model.

3.4 Concluding remarks

The dissociative or interstitial-substitutional mechanism, which has been shown by many authors to describe the diffusion of gold in silicon has been outlined. The diffusive flux is carried almost solely by

interstitial atoms but as these have a low solubility compared to the substitutional atoms, measured gold concentrations are generally thought to be predominantly substitutional gold. The precise proportion of gold in the interstitial state at the solubility limit - as measured by Wilcox et al. (1964) and illustrated in figure 3.7 - is called into question by the more recent analysis of Huntley et al. (1973) although the interstitial solubility is not actually discussed by them. It will be seen, later in this thesis, that this lack of knowledge about the interstitial gold species is something of a problem in predictions of various effects of gold on the properties of doped silicon.

There are still unanswered questions concerning the rate limiting factors which control the gold dissociative process, particularly in thin silicon wafers. Evidence has been published for both a vacancy supply rate limited process and a gold-vacancy reaction rate limited process. There is no consensus amongst authors over the question of the relative concentrations of substitutional gold and interstitial gold during a diffusion process. Many assume that the two species behave independently and that Au_i achieves an equilibrium independently of Au_s unless the reaction rate is high. In a situation where there is a lack of vacancies, one might therefore expect gold concentrations measured by the radiotracer technique to be predominantly interstitial. Early works and recently, Huntley et al. (1973) suggest that the equilibrium ratio Au_s : Au_i is maintained at all times (unless an extremely high reaction rate depresses Au_i). Obviously the relative concentrations are very dependent on the reaction rate but there is no clear evidence to indicate precisely what they are.

The data obtained on thin wafers could certainly be taken to indicate a reaction rate limited process, however the difficulty of obtaining material that is totally free of internal sources of vacancies (whatever they may be) may explain why the differences observed between dislocation free and dislocated silicon are so small. This is certainly indicated by results reported in chapter 8.

A problem which does arise, particularly if a significant proportion of the gold is interstitial, is that of gold redistribution during cooling.

The highly mobile interstitial species could undergo significant redistribution if gold diffused samples are not cooled very rapidly from the diffusion temperature.

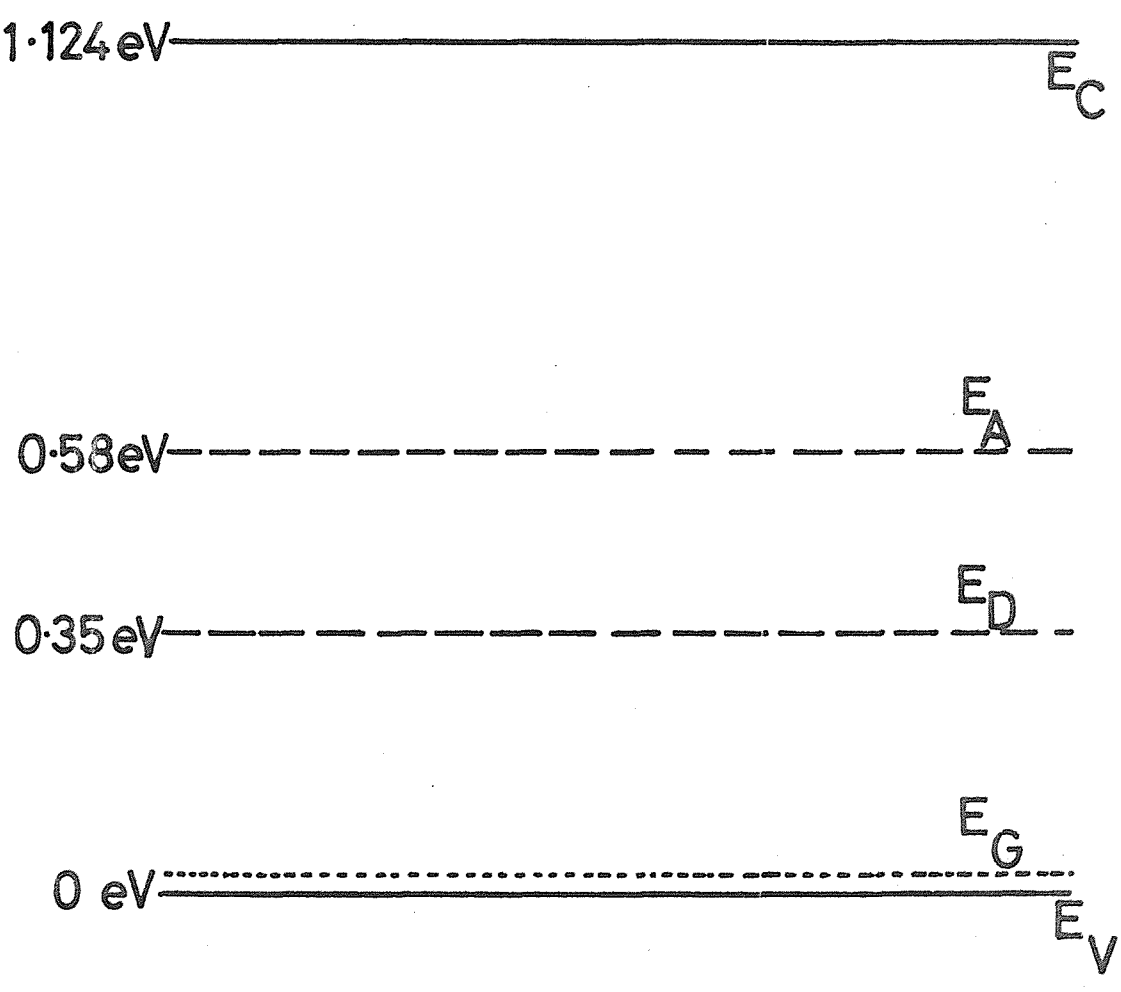


FIG. 4.1: Positions of the Gold Energy Levels in the Silicon Band Gap. E_A = Gold Acceptor, E_D = Gold Donor, E_G = Gold-coupled Shallow Acceptor.

4. GOLD ENERGY LEVELS AND DEGENERACIES IN SILICON

4.1 Gold Energy Level Positions

Energy levels in the silicon energy band gap (forbidden gap) are generally divided into two types : shallow levels and deep levels. Shallow energy levels are those which lie within 0.1 eV of either of the band edges and deep levels are those which occupy the rest of the silicon energy band gap (Band gap at $300^{o}K \simeq 1.12$ eV).

Gold introduces two deep lying energy levels, an acceptor and a donor, into the silicon energy band gap. These deep levels were first reported in 1954 and 1957 by Taft and Horn and Collins, Carlson and Gallagher respectively. Subsequently a number of other authors have used different methods to determine their exact positions in the silicon band gap (see Section 4.5). The gold acceptor energy level is situated in the upper half of the energy band gap, about 0.54 eV below the conduction band edge and the donor level is situated in the lower half of the energy band gap about 0.35 eV above the valence band edge. They are illustrated in figure 4.1.

Recently a shallow acceptor energy level associated with gold and positioned 0.033 eV from the valence band edge has been reported by Bruckner (1971) and Thurber et al. (1973); this will be discussed in Section 4.41.

Knowledge of the precise position of the gold energy levels in the silicon energy band gap is, as will be seen in later chapters, of great importance in the explanation of a number of phenomena associated with them. (Chapters 5, 6 and 7).

The occupancy of any energy level by an electron is given by the Fermi probability function where the probability, F(E), that a quantum state at an energy E_x contains an electron is given by: (Nichols and Vernon, 1966, p.75)

$$F(E) = \frac{1}{1 + \exp\left[\frac{E_x - E_F}{kT}\right]} \quad \ldots 4.1$$

where E_F is the fermi energy level. For a density N_a of acceptors at an energy level E_a in the silicon energy band gap, the number of ionized acceptors (i.e. those which are occupied by electrons) is:

$$N_a^- = N_a \times F(E) \quad \ldots 4.2$$

For a density N_d of donors at energy E_d, the number of ionized donors (i.e. those which are NOT occupied by electrons) is:

$$N_d^+ = N_d \times [1 - F(E)] \quad \ldots 4.3$$

In either case, provided E_F is either several times (kT) larger or smaller than $E_x - E_F$, the value of F(E) is either very large or very small. In the case of shallow lying energy levels this is generally true (except at high doping concentrations and high temperatures) and the exact value of $(E_x - E_F)$ is not very important. However, if $E_x \simeq E_F$ then the value of the exponential term in equation 4.1 is comparable with unity and the value of F(E) is very strongly dependent on the precise value of $(E_x - E_F)$. In most cases of practical interest for gold doped silicon (which is not heavily doped with shallow donors or shallow acceptors), E_F as well as the gold energy levels will be deep lying in the energy band gap and, therefore, calculation of the number of ionized gold levels is a strong function of the exact position of the gold energy levels. For this reason the position of the gold energy levels must be considered very carefully. (See Sections 4.4, 4.5 and 4.6).

4.2 Impurity Level Spin Degeneracy

Equation 4.1, the Fermi probability (or Fermi-Dirac) function, F(E), is obtained by considering the distribution amongst available energy levels of particles (in this case electrons) which obey Pauli's exclusion principle. The exclusion principle allows each quantum state to be occupied by no more than one electron.

If quantum states are offered, over some range of energies, the probability that a given state is occupied is derived by considering

the total number of possible distributions of the electrons available in all of the states offered and then finding the most statistically probable distribution. In general, if ω is the total number of possible distributions of n electrons amongst the available states and n_r is the number of occupied states at a given energy level E_r (Σn_r for all r, equals n, the total number of electrons, which is constant), then the most probable distribution occurs when:

$$\frac{\partial \omega}{\partial n_r} = 0 \qquad \ldots 4.4$$

that is, when ω is a maximum subject to the constraints that Σn_r is constant and the total energy $\Sigma n_r E_r$ remains constant. This equation is solved rigorously to obtain the maximum of ω (McKelvey, 1966, page 149). The result of equation 4.1 for F(E) is obtained with the additional constraint that each impurity offers only one quantum state for the neutral or ionized configuration.

If a given impurity at some energy E_r offers more than one quantum state for the neutral or for the ionized configuration, the probability of occupancy must be statistically weighted in favour of this state. An example will illustrate the effect that this will have on the occupancy of such states at E_r and how more than one state may be offered without violating the exclusion principle.

A simple monovalent donor impurity in silicon is considered, for which all electrons apart from the least tightly bound outer electron are paired in valence bonds with the surrounding silicon atoms. The probability of it being unionized (i.e. containing the outer electron and therefore being neutral), as already shown, is given by equation 4.1 if in the ionized condition (when it does not contain the outer electron) it offers only one state for the trapping of the outer electron. This outermost electron can however be trapped in two ways, spin up or spin down, without violating the exclusion principle as these two spin states constitute separate quantum states at the donor energy level. Although the donor cannot actually trap two electrons (since when one electron is trapped the valency requirements of the donor ion are satisfied and electrostatic forces raise the energy for the remaining spin possibility to a very high value) it does offer two quantum states for trapping the electron. If there are N_d^+ ionized donors (i.e. N_d^+ atoms which offer some

possibility of becoming occupied by electrons). There will be $2 \times N_d^+$ quantum states actually available. The maximum number of these states which may become occupied is only N_d^+ because as soon as a given donor has trapped one electron of either spin in the appropriate state occupancy of the other state is precluded for the reason already given. Although only half of the available states may actually become occupied, the value of ω, which was the total number of possible distributions of n electrons amongst the states, will be increased by some factor due to the doubling of the spin states available. Consequently the form of F(E) which is derived will be changed such that the probability of occupancy of the states by electrons (neutral donors) is greater because more states for occupancy are offered. The way in which F(E) is modified in this simple case may be illustrated as follows without the rigorous solution of equation 4.4:

The number of ionized donors, N_d^+ is given by equation 4.3:

$$N_d^+ = N_d[1 - F(E)] \qquad \ldots 4.3$$

Since N_d (neutral)= $[N_d$ (total)$]- N_d^+$; the ratio of N_d^+ to N_d(neutral) may be obtained, however because twice as many states as were originally assumed in obtaining F(E) are now available, the neutral (occupied) state is weighted by a factor of 2, giving:

$$\frac{N_d^+}{N_d \text{ (neutral)}} = \frac{1 - F(E)}{2 \times F(E)} \qquad \ldots 4.5$$

From equation 4.1:

$$\frac{N_d^+}{N_d(\text{neutral})} = \frac{1}{2F(E)} - \frac{1}{2} = \frac{1}{2} \exp\left[\frac{E_d - E_F}{kT}\right] \qquad \ldots 4.6$$

For this example of a monovalent donor which can have an electron of either spin in the unionized state, the Fermi probability function becomes:

$$F'(E) = \frac{1}{1 + \frac{1}{2} \exp\left[\frac{E_d - E_F}{kT}\right]} \qquad \ldots 4.7$$

from equations 4.3 and 4.6.

As expected, this probability that the level is occupied by an electron, and is therefore unionized, is slightly greater for given values of E_d and E_F than that given by equation 4.1 which neglects the spin degeneracy.

The case of a monovalent acceptor impurity may be visualised similarly. The acceptor in the unionized state comprises three covalent bonds with surrounding silicon atoms and one bond which lacks an electron.

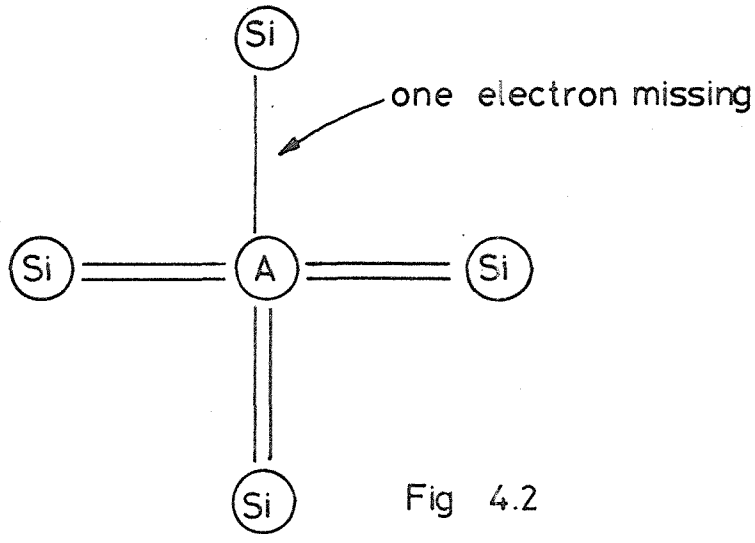


Fig 4.2

For the acceptor to be ionized, an electron must complete the fourth bond, however as the bond consists of two paired electrons when it is complete, the electron which completes the bond must be of the correct spin. Thus unlike the donor impurity which offered two states for electrons, the acceptor only offers one. The ionized state for the acceptor (in which an electron is present in the bond) is therefore less likely and F(E) must be smaller. A similar argument to that presented for the donor will lead to:

$$F'(E) = \frac{1}{1 + 2 \exp\left[\frac{E_a - E_F}{kT}\right]} \qquad \ldots 4.8$$

which is the probability that the acceptor is ionized, since the "absence of an electron" condition now has a statistical weight of 2 over the "presence of an electron" condition.

These probability expressions arise from a simple statistical view of the behaviour of electrons when they have a choice of occupying various quantum states. Equations 4.7 and 4.8 simply predict which states the electrons are most likely to be found in when the various possibilities and their likelihoods have been assessed. If one energy level, as in the case of a donor, offers more than one quantum state for electrons then the probability is weighted in its favour. F(E) may be written in general as:

$$F(E) = \frac{1}{1 + g \exp\left[\frac{E_x - E_F}{kT}\right]} \qquad \ldots 4.9$$

where g is called the degeneracy factor.

In general, for a system of energy levels and particles which obey the exclusion principle, the value of g for a given energy level E_x may be found by considering the statistical weighting in favour of, or against, occupation of the states at E_x. With the shallow monovalent donor there are two ways in which an electron filling the available site can be described (spin up or spin down) and only one way in which its absence can be described. The degeneracy factor, g, of 1/2 is the ratio of the number of ways it can be empty to the number of ways it can be filled. For the simple shallow monovalent acceptor considered above there is only one way of describing the presence of the electron since it must be of the correct spin, but there are two ways of describing its absence since an electron of either spin may be lost. Once again the value of g is given by the ratio of the number of ways of describing the level as empty (absence of the electron) to the number of ways of describing it as full, i.e. 2:1 or g = 2.

By definition (Shockley and Read, 1952 appendix B) the spin degeneracy of a given condition (or charge state) of an impurity is the number of ways in which that condition may be described. The degeneracy factor which appears as 'g' in equation 4.9 is the ratio of the degeneracies of the occupied and unoccupied conditions. Once again, for the simple monovalent donor example, the degeneracy of the unoccupied

(ionized) condition is 1 and the degeneracy of the occupied (unionized) condition is 2, giving rise to a degeneracy factor of g = 1:2.

Having established the meaning of degeneracy and its overall effect on the occupancy of a given level, it should be possible to assess the degeneracy factor of any energy level if the number of ways it can be occupied or unoccupied are calculated. ("the number of ways" is synonomous with "degeneracy"). In practice this is not always an easy task because the degeneracy ("number of ways") depends not only on the valency of the impurity concerned but also on the band structure of the semiconductor under consideration. The effect of considering the silicon valence band structure on the simple monovalent acceptor impurity will serve to illustrate this:

The degeneracy of the "filled" condition of the acceptor is clearly 1 as there is only one way it can be occupied (by an electron of the correct spin). The number of ways it can be described as empty, however, is modified by the form of the silicon valence bands. At the valence band maximum in silicon, which occurs at k = 0 on an E - k diagram, there are in fact two coincident valence bands both of which are doubly degenerate (i.e. they may both be occupied by two electrons of opposite spins or just one electron of either spin). There is, therefore, a total of four ways in which the "empty" condition of the acceptor may be described since the absent electron can be of either spin and may reside in either of the valence bands. The degeneracy of the "empty" condition is 4 and the degeneracy factor (the ratio of the "empty" and "full" degeneracies) is 4:1 and not 2:1 as assumed in the previous discussion. The probability that the acceptor level is filled (unionized) from equation 4.9 with g = 4, is now even smaller since the level is offered more choices for being empty.

The consideration given above to the two monovalent impurities may be extended to multivalent and/or multilevel impurities, such as gold, in silicon. The specific degeneracies and degeneracy factors (g) associated with the gold energy levels in silicon will be discussed in section 4.6.

The importance of knowing the values of g for the gold energy levels is equal to that of knowing the precise positions of the gold energy levels in the silicon energy band gap. In every occupancy calculation for the gold levels which includes expressions derived from F(E), the factor 'g' will appear as the coefficient of the exponential, thus: $g \exp\left[\frac{E_x - E_F}{kT}\right]$;

this may be rewritten:

$$\exp\left[\frac{(E_x + kT \ln g) - E_F}{kT}\right] \qquad \ldots 4.10$$

The effect of the degeneracy factor g is clearly the same as moving the position of the energy level by (kT ln g) the result of which will be very significant if E_F is close to E_x (see section 4.1).

In the literature the problem of the degeneracy factor is often avoided by including it in E_x (as in 4.10) by calling E_x an "effective energy level". This approach is very misleading and can give rise to a great deal of inaccuracy since estimates of E_x are often made by a method which eliminates the effect of g. This is especially so in measurements of Hall coefficient, resistivity and electron emission probabilities as function of temperature in gold doped silicon. The measured values are plotted against temperature and the slope of the curve is measured to obtain E_x. The degeneracy factor, g, is normally expected to be independent of temperature (see sections 4.21 and 4.6) and will therefore not contribute to the slope of the curve or hence to E_x, but to its intercept at zero temperature. These methods and measurements are discussed in some detail in section 4.52.

There is also a certain amount of confusion in the literature as to which number is quoted for the degeneracy factor of a particular energy level. This arises because the most commonly used expression for occupancy is that which gives the number of ionized atoms of a particular impurity. For acceptors this is equation 4.2 which gives

$$N_a^- = F(E) \times N_a = \frac{N_a}{1 + g_a \exp\left[\frac{E_a - E_F}{kT}\right]} \quad \ldots 4.11$$

For ionized donors, however, equation 4.3 yields:

$$N_d^+ = \left[1 - F(E)\right] \times N_d = \frac{N_d}{1 + \frac{1}{g_d} \exp\left[\frac{E_F - E_d}{kT}\right]} \quad \ldots 4.12$$

It is very common for $1/g_d$ in this equation to be quoted as "the degeneracy factor g_d", although in reality it is the reciprocal of the true degeneracy factor which arises in equation 4.9 from the definition given in this section.

It is important not to confuse the concept of impurity level spin degeneracy discussed here with the concept of a "degenerate semiconductor" which is defined as one in which the number of majority carriers is approximately equal to the density of available quantum states for majority carriers in the appropriate band.

For absolute clarity in this work, impurity level spin degeneracy will be referred to as 'degeneracy' and the numerical value quoted for the degeneracy factor will be that of g in equation 4.9, with an appropriate subscript (e.g. g_d for a shallow donor impurity and g_a for a shallow acceptor impurity). Equation 4.9 refers to the probability that a given quantum state at an energy level E_x is occupied by an electron, regardless of its charge state, with the degeneracy factor g defined as the ratio of the number of ways that the state may be empty (degeneracy of the unoccupied condition) to the number of ways that it may be filled (degeneracy of the occupied condition).

The great importance of the degeneracy factors of the gold energy levels will become apparent from the theoretical calculations of gold solubility and gold-doped silicon resistivity which are presented in chapters 6 and 7. The specific values of the degeneracies and degeneracy factors which may be assigned to the gold energy levels in silicon are discussed in section 4.6.

4.21 Temperature Dependence of Degeneracy

The possibility that the degeneracy of a given charge condition of an impurity may be temperature dependent arises from the possible ionization of excited quantum states of the impurity at higher temperatures. If the temperature is high enough to make ionization of the excited states possible, the "number of ways" available for a given occupancy (charge condition) of the impurity will increase and the degeneracy and hence degeneracy factor of that condition will change.

The presence, at high temperatures, of excited quantum states in shallow level impurities (such as phosphorus) in silicon is well known. It is not expected however, that deep level impurities such as gold will exhibit any effect due to excited states since the separation of the excited states from the ground state is comparable with the silicon energy band gap (Wong and Penchina, 1975). Some effects noted in gold doped silicon have been attributed to temperature dependent degeneracy factors; it will be shown in later section of this chapter that such effects are probably due to temperature variations of the gold energy levels and not the degeneracies.

4.3 Temperature Dependences of Gold Energy Levels in Silicon

Measurements of the energy levels of gold in silicon by most of the methods to be described in section 4.5 generally result in the position of one or both of the gold energy levels relative to one silicon energy band edge at a particular temperature. For this information to be useful in the analysis of other phenomena associated with gold in silicon, either the measurement must have been carried out at the temperature of interest or the variation of the positions of the gold energy levels with temperature must be known.

4.31 Variation of Silicon Energy Band Gap with Temperature

The silicon energy band gap shrinks with increasing temperature due to dilation of the crystal lattice (thermal expansion) and the increasing strength of thermal vibrations in the crystal. Most evidence points to this

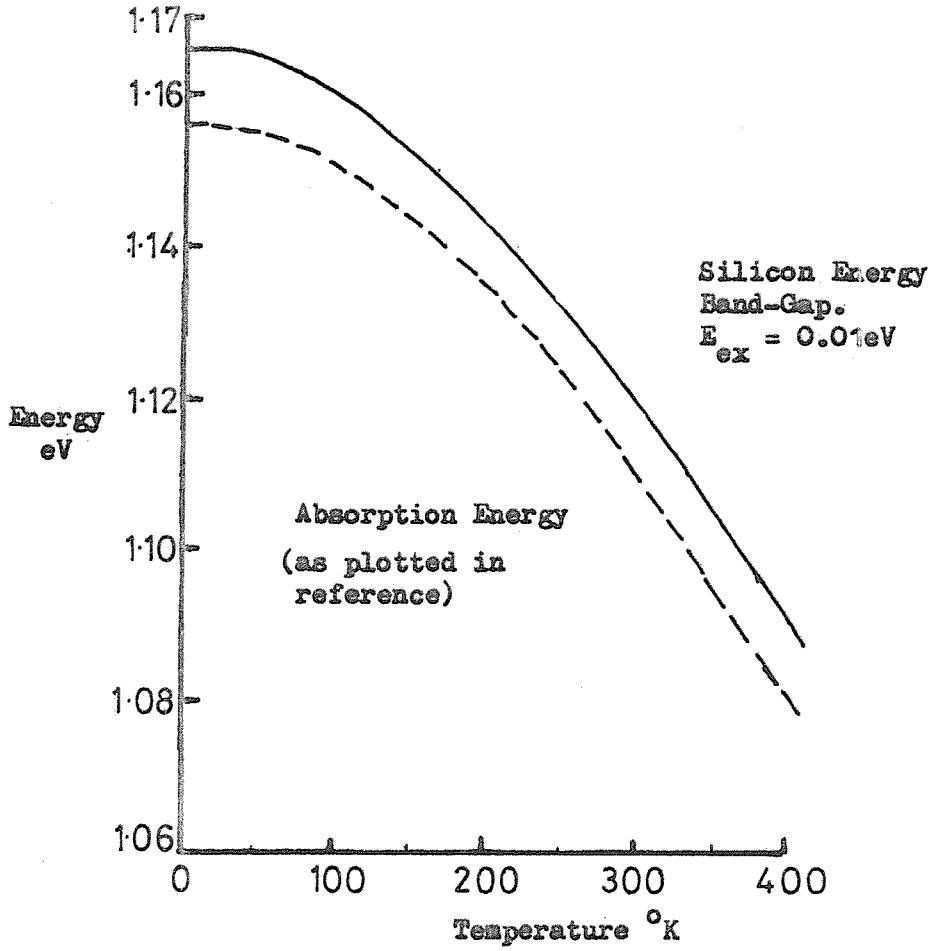


FIG. 4.3: Silicon Energy Band Gap Data from MacFarlane et al., 1958.

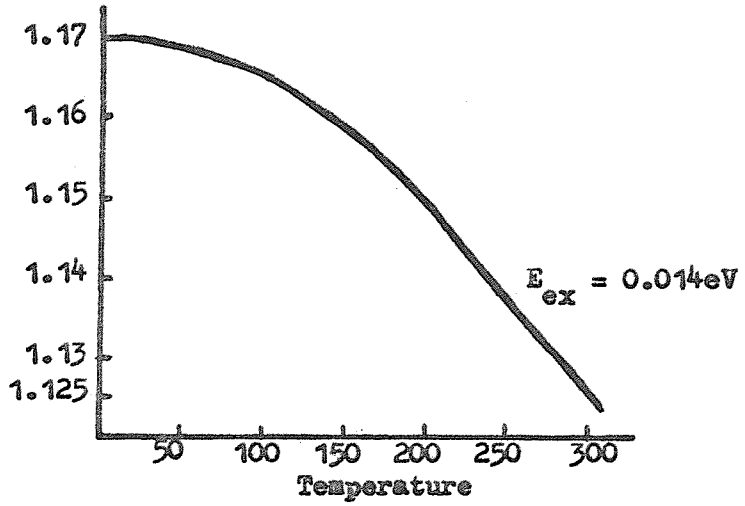


FIG. 4.4: Silicon Energy Band-Gap Data from Bludau et al., 1975.

variation with temperature being nonlinear below 300°K and linear above 300°K. (Macfarlane et al, 1958; Haynes et al, 1959; Bludau et al, 1974).

Measurements of the temperature dependence of the silicon energy band gap have been made by each of the authors cited. Macfarlane et al. made measurements of the absorption spectrum of silicon at temperatures between 4.2°K and 415°K. The main absorption edge observed by them is that indicative of the transition of electrons from the silicon valence band to the conduction band. The energy measured at the absorption edge, however, does not correspond directly to the energy band gap but to the energy required to form excitons. Excitons are electron-hole pairs which after generation are still loosely held together by coulomb attraction. Movement of an exciton does not constitute a conduction process as two oppositely charged particles are moving together. The "energy band gap" is equal to the energy needed to form an electron-hole pair (exciton) plus the energy needed to break the binding of the exciton, so that electron and hole are free to move independently in the conduction and valence bands respectively.

$$E_g = \overline{E} + E_{ex} \qquad \ldots 4.13$$

where E_g is the energy band gap, $\overline{E}$ is the absorption energy and E_{ex} is the exciton binding energy. Macfarlane et al, in their figure 9, present a plot of absorption energy as a function of temperature which is reproduced here as a broken line in figure 4.3. To obtain the silicon energy band gap at a particular temperature, the exciton binding energy has to be added to the values taken from their plot. In Macfarlane et al's paper, the value for E_{ex} is taken from one estimated by Dresselhaus (1956), and is approximately 0.01 eV. Using this value, the silicon energy band gap as a function of temperature is shown as a solid line on figure 4.3. It appears that many authors have misinterpreted figure 9 in Macfarlane et al's paper as showing the silicon energy band gap as a function of temperature because the generally quoted value of 1.11 eV at 300°K (room temperature), which is usually attributed to Macfarlane et al. is in fact the measured absorption energy at 300°K (see for example: Grove, 1967; Bullis, 1966; Thurber et al, 1973); a value of 1.12 eV, however, would be the correct

interpretation of their results.

The non-linearity of the energy band gap below 300°K is apparent in figure 4.3.

Varshni (1967) has proposed an expression for $E_g(T)$ of the form:

$$E_g = 1.557 - \frac{\alpha T^2}{T + \beta} \qquad \ldots 4.14$$

from theoretical considerations of the factors which cause energy band gap shrinkage. For silicon, values of $\alpha = 7.021 \times 10^{-4}$ and $\beta = 1108$ are proposed empiricially, to make the expression fit data supplied by McLean et al. (1960). The origin of McLean's data is not at all clear. It probably is taken from Macfarlane et al, since McLean is one of the co-authors. Varshni's expression gives the absorption energy in silicon and not the energy band gap.

Haynes et al. (1959) inferred values of the silicon energy band gap from an analysis of intrinsic (band to band) recombination in silicon. At lower temperatures their values are similar to those of Macfarlane et al. but they differ by up to 0.015 eV at higher temperatures. These results are probably less accurate than those of Macfarlane, especially at higher temperatures, because of recombination centres (impurities) in the silicon energy band gap also taking part in the recombination process.

Recently Bludau, Onton and Heinke (1974) have carried out measure--ments, similar to those of Macfarlane et al, of the exciton absorption energy in silicon. They used the more accurate technique of wavelength modulation spectroscopy on silicon samples that were probably much purer than those of Macfarlane et al. and Haynes et al. The range of temperature used was 2°K to 300°K. Bludau et al's results are shown in figure 4.4. The exciton binding energy used was 0.0147 eV (± 0.0004 eV), taken from the recent work of Shaklee and Nahory (1970). If this value is used in Macfarlane et al's data, the agreement between Bludau et al's and their work is extremely good (within 0.001 eV at 300°K).

The values obtained by Bludau et al. have been used in this work for calculations of resistivity in gold doped silicon at room temperature

(see Chapter 7). Previous calculations (Bullis, 1966; Bullis and Strieter, 1968; Thurber, 1973) have all used the incorrect value of E_g = 1.11 eV at 300°K which, as will be shown, causes very large differences in the final results.

4.32 Variation of the Gold Energy Level Positions with Temperature

In view of the variation of the silicon energy band gap with temperature, it is clear that in order to predict the position of the gold energy levels at any temperature, their variation with temperature relative to the band gap edges must also be known.

Shallow impurities are described, quantum mechanically, by simple wave functions from one energy band only. Shallow donor states are made up of conduction band wave functions and shallow acceptors are made up of valence band wave functions (Smith R.C., 1964). The ionization energy of such shallow levels is found to be essentially independent of temperature, i.e. as temperature varies the shallow level remains a fixed distance from the energy band with which it is associated. Multiply charged deep impurities (such as gold which can be an acceptor or a donor in silicon) cannot be described by the same simple theory. The impurity levels are not in general considered to be made up of wave functions from a single band; combinations of wave functions from both energy bands are probably needed to describe them. As a result, a deep donor state is not necessarily expected to remain a fixed ionization energy from the conduction band edge nor a deep acceptor to remain a fixed ionization energy from the valence band edge as the band edge varies with temperature. No theory has yet been proposed to explain the temperature dependences of deep energy levels. In the case of gold, it is only recently that any reliable experimental evidence has been presented and there is a certain amount of conflict.

Parrillo and Johnson (1972) proposed, on the basis of measurements of the temperature dependences of electron and hole emission rates from the gold acceptor energy level, that the gold energy levels vary in proportion to the variation of the silicon energy band gap with

temperature ($E_g(T)$). Their interpretation of their results and indeed their results are open to considerable doubt (see Section 4.53.2).

Engstrom and Grimmeiss (1974) carried out detailed experiments on the optical emission rates of the gold acceptor energy level as a function of temperature. Their results show that the variation in distance of the acceptor energy level from the valence band edge is the same shape as the variation in the silicon energy band gap with temperature measured by Bludau et al. (1974). This result suggests that the gold acceptor level remains a fixed distance from the silicon conduction band edge (i.e. the gold acceptor level is pinned to the conduction band edge) as temperature varies.

Wong and Penchina (1975) have re-analysed photoconductivity measurements in p-type gold doped silicon made by Newman (1954). Their conclusion is that the gold donor energy level is also fixed to the conduction band edge as temperature varies.

In this work, calculations of resistivity in gold doped silicon as a function of shallow impurity and gold concentrations have been carried out for all of the possible simple models of gold energy level variation with temperature. These are:

1) Each gold level is fixed to the energy band edge to which it is closest. The gold acceptor energy level is a constant distance from the conduction band edge and the gold donor energy level is a constant distance from the valence band edge.

2) The gold energy levels remain fixed to the energy band with which they mainly interact. The gold acceptor level, which is involved with holes, remains fixed to the valence band edge and the gold donor, which is involved with electrons is fixed to the conduction band edge.

3) Both gold levels remain fixed to the conduction band edge (in accordance with the measurements of Engstrom and Grimmeiss and Wong and Penchina).

4) Both gold levels remain fixed to the valence band edge.

5) Both gold levels are influenced by both bands in proportion to their proximity to them. This gives rise to the position of the gold levels varying in proportion to the variation of the silicon energy band gap (as suggested by Parrillo and Johnson).

The results presented in Chapter 7 indicate good agreement with Model 3.

4.4 Shallow Energy Levels Associated with Gold in Silicon

4.41 Shallow Acceptor Energy Level

Prior to 1971, it was generally assumed that gold only gave rise to the two deep lying energy levels in the silicon band gap already mentioned. It was apparent however from the few resistivity measurements made by Wilcox et al. (1964), on p-type silicon, that there might be an additional energy level or levels associated with gold. The measure--ments showed order of magnitude discrepancies between theoretical and measured values of resistivity in heavily gold-doped silicon.

Bruckner (1971) carried out Hall coefficient and resistivity measurements as a function of gold doping and temperature on gold doped p-type silicon samples. The investigation revealed the existence of a shallow acceptor energy level, 0.033 eV from the valence band edge, for samples in which the density of gold, N_{Au}, was much greater than the density of shallow acceptors, N_a, (N_a was 1.5×10^{14} atoms. cm^{-3}). Repeated experiments proved the reproducibility of the shallow acceptor level and low temperature annealing experiments indicated that it was stable. The nature of the shallow acceptor level was suggested to be that of an electrically active complex of gold with other defects in the silicon. A similar type of energy level has been found by Tweet (1958)

who carried out heat treatment experiments on copper doped germanium.

The presence of this shallow acceptor level - generally referred to as the gold-coupled shallow acceptor - has been confirmed by Thurber et al. (1972, 1973). From measurements of resistivity in heavily gold doped silicon, Thurber et al, suggest that the density of the gold-coupled shallow acceptors increases as the third power of the gold concentration, with a shallow acceptor concentration of 4.5×10^{15} atoms. cm^{-3} at a gold concentration of 1×10^{17} atoms. cm^{-3}. In this work it has been found that Thurber et al's predicted gold-coupled shallow acceptor density adequately describes the behaviour of resistivity in heavily gold-doped silicon (see Chapter 7).

4.42 Shallow Donor Energy Level

It has been suggested (Bullis, 1966; Brown et al. 1975) that interstitial gold may give rise to a shallow energy level, in this case a donor level. Bullis draws on an analogy with the behaviour of lithium and copper in silicon. The interstitial gold concentration in silicon is generally accepted to be about ten times lower than the substitutional gold concentration (see Chapter 3) but the interstitial atoms are thought to be the dominant diffusing species at high temperatures.

Experiments on the diffusion of gold in the presence of an electric-field, which were carried out by Boltaks et al. (1961) indicate that gold was accumulated at the cathode side of the sample. This is very strong evidence for gold being in a predominantly positive (ionized donor) charge state at the diffusion temperature. The possibility that this charge state is due to the substitutional species is unlikely since substitutional gold atoms are relatively immobile and so a substantial build up of positively charged gold diffusing by a substitutional mechanism would take an extremely long time.

Brown et al. carried out experiments on the solubility of gold in lightly doped and heavily doped p-type silicon at 1000°C, 1100°C, 1200°C and 1300°C. They also measured the gold concentration as a function of diffusion time and observed a difference in the rate of increase of gold concentration in heavily doped over that in lightly doped p-type silicon. This was explained in terms of interstitial gold behaving as a shallow donor. However, in the theoretical analysis of their measured enhancement of gold solubility in heavily doped p-type silicon, they ignore any possible contribution to the enhancement by the interstitial gold donors. The solubility enhancement is explained solely in terms of the behaviour of

substitutional gold atoms. In Chapter 6 of this work, it will be shown that if interstitial gold atoms do behave as shallow donors at diffusion temperatures, the solubility enhancement in heavily doped p-type silicon measured by Brown et al. is almost certainly due to the enhancement of the interstitial and not the substitutional species. If Brown et al's explanation of the rate of increase of gold concentration in terms of a gold interstitial shallow donor is to be accepted, their explanation of gold solubility enhancement in terms of substitutional gold is not.

The effect of this shallow donor level on the resistivity of gold-doped silicon at room temperature is calculated in Chapter 7. The results are incompatible with the measured values of resistivity as a function of gold concentration. The only conclusion to be reached, therefore, is that if interstitial gold does indeed exhibit shallow donor properties, it only does so at high temperatures.

4.5 Measurement of the Position and Temperature Dependence of the Gold Energy Levels in Silicon

4.51 Introduction

A number of works on the measurement of gold energy levels in silicon have been published. The methods fall into two distinct groups.

The first group consists of methods which measure a property of gold at one temperature, which directly yields the ionization energy of one of the gold levels at that temperature. The second group comprises measurements over a range of temperature of some other property of gold-doped silicon. The property chosen is one which does not give the gold energy level directly but one for which the activation energy is determined by the ionization energy of one of the gold levels. A study of the variation of the property as a function of temperature will yield its activation energy which can be related to the appropriate gold energy level.

Table 4.1 lists references of measurements which fall into the first group and table 4.2 lists those which fall into the second group.

The interpretation of results from group one, is generally straightforward at the temperature of measurement. Comparison of results obtained at different temperatures, by the same method, can indicate the temperature dependence of the energy level in question.

TABLE 4.1 Measurements of Gold Energy Levels at a Single Temperature

Newman (1954)
Badalov (1970)
Braun and Grimmeiss (1974)
Engstrom and Grimmeiss (1974)
Wong and Penchina (1975)

TABLE 4.2 Measurements of Gold Energy Levels from Activation Energies

Taft and Horn (1954)
Carlson (1956)
Collins, Carlson and Gallagher (1957)
Boltaks, Kulikov and Malkovich (1960)
Senechal and Basinski (1968)
Sah, Forbes, Rosier, Tasch and Tole (1969)
Tasch and Sah (1970)
Bruckner (1971)
Parrillo and Johnson (1972)
Thurber, Lewis and Bullis (1973)
Braun and Grimmeiss (1973)
Pals (1974)
Kassing and Lenz (1974 i)
Kassing and Lenz (1974 ii)
Engstrom and Grimmeiss (1975)

Results from the second group need much more careful interpretation because in all cases a number of the parameters involved in obtaining the energy level under investigation exhibit differing temperature dependences of their own over the range of measurement.

4.52 Energy Level Measurement Methods over a Temperature Range

4.52.1 Equivalence of results

The three methods which fall into the second group are:

1) measurement of Hall coefficient,

2) measurement of resistivity,

and 3) measurement of the thermal emission rates of electrons and holes,

all as a function of temperature. These three methods may be considered in a similar fashion as follows:

Samples are prepared and experiments carried out under appropriate conditions so that only one of the gold energy levels is dominant during the measurements. Details of the individual methods and conditions will be given later. Initially it is assumed, for simplicity, that the experiment has been set up to investigate the position of the gold acceptor energy level with respect to the conduction band edge. (The comments and results of this discussion are equally applicable to measurements of either gold energy level with respect to either band edge).

i) Hall Coefficient

In n-type silicon, if the gold concentration is greater than the shallow dopant concentration, a temperature range may be defined within which the fermi level in the silicon is very close to the gold acceptor energy level. In this case the electron concentration is determined by the position of the gold acceptor level. The Hall coefficient, R_H, is given by: (Smith R. A., 1964, p.102)

$$R_H \propto \frac{1}{nq} \qquad \ldots 4.15$$

where n is the number of electrons and q is the electronic charge. n is given by:

$$n = N_c \exp\left[\frac{E_F - E_c}{kT}\right]$$

where N_c is the density of available states in the conduction band, E_F is the fermi level, E_c is the conduction band energy, k is Boltzmann's constant and T is the absolute temperature. The conditions of the experiment have been chosen so that E_F is coincident with E_A, the gold acceptor energy level. 4.15 may be rewritten:

$$R_H^{-1} = g_A \, Kq \, N_c \exp\left[\frac{E_A - E_c}{kT}\right] \qquad \ldots 4.16$$

where K is a constant of proportionality related to the ratio between the Hall mobility and conductivity mobility of the carriers; its exact value is unimportant in this context provided it is temperature independent.

(ii) Resistivity

The resistivity of a sample prepared in the same way as above will be given by:

$$\rho = \frac{1}{n \, q \, \mu_n} \qquad \ldots 4.17$$

where n and q are defined above and μ_n is the conductivity mobility of electrons in the sample. Equation 4.17 may be rewritten:

$$\rho^{-1} = g_A \, q \, \mu_n \, N_c \exp\left[\frac{E_A - E_c}{kT}\right] \qquad \ldots 4.18$$

(iii) Thermal Emission Rate of Electrons

The thermal emission rate of electrons from the gold acceptor level to the conduction band is given by: (see Appendix D and Shockley and Read, 1952)

$$e_n^t = g_A \sigma_n^t v_n^t N_c \exp\left[\frac{E_A - E_c}{kT}\right] \quad \ldots 4.19$$

where g_A is the degeneracy factor for the gold acceptor level, σ_n is the capture cross section of electrons at the gold acceptor level and v_n is the average thermal velocity of electrons. The superscript 't' is used to indicate that these quantities are related to a thermal capture process rather than an optical one, for which a superscript 'o' would be used.

In equations 4.16, 4.18 and 4.19 there are a number of terms which may be described as simple functions of temperature if their temperature dependences are known.

In 4.16, if N_c varies as T^m, then:

$$R_H^{-1} = AT^m \exp\left[\frac{E_A - E_c}{kT}\right] \quad \ldots 4.20$$

In 4.18, if μ_n varies as T^n, then:

$$\rho^{-1} = BT^{m+n} \exp\left[\frac{E_A - E_c}{kT}\right] \quad \ldots 4.21$$

In 4.19, if v_n varies as T^p, σ_n varies as T^q and g_A is temperature independent (see Section 4.2), then:

$$e_n^t = CT^{m+p+q} \exp\left[\frac{E_A - E_c}{kT}\right] \quad \ldots 4.22$$

where A, B and C are temperature independent constants.

It can now be seen that each of these three measurements exhibit similar relationships with the position of the gold acceptor energy level, E_A, relative to the conduction band edge, E_c. Equations 4.20, 4.21 and 4.22 may be considered to be of the form:

$$F = DT^r \exp\left[\frac{-\Delta E}{kT}\right] \quad \ldots 4.23$$

where $\Delta E = E_c - E_A$ is the ionization energy of the gold acceptor level, D is a temperature independent constant and r provides the appropriate temp--erature dependence.

The natural logarithm of this expression may be obtained,

$$\ln (FT^{-r}) = \ln D - \left[\frac{\Delta E}{kT}\right] \qquad \ldots 4.24$$

If the property F is measured as a function of temperature and $\ln(FT^{-r})$ plotted against T^{-1}, the slope of the plot will be $\Delta E/k$ if ΔE is temperature independent.

4.52.2 Temperature variation of the activation energy, ΔE.

If ΔE is not temperature independent, modifications to the above analysis must be made. The temperature variation of ΔE may be approximated, over the temperature range of measurements, to be linear:

$$\Delta E(T) = \Delta E_o(T_m) + \left[G(T_m)\right].T \qquad \ldots 4.25$$

where $G(T_m)$ is the slope of this linear approximation to ΔE vs T, T_m is the 'temperature of measurement' and $\Delta E_o(T_m)$ is the intercept of the straight line at T = 0. In reality, the variation of ΔE over a wide range of temperature may not be linear, in which case the straight line described by 4.25 approximates to a tangent to the real ΔE vs T at the temperature of measurement T_m. In general, therefore, $\Delta E_o(T_m)$ is the intercept at T = 0 of a tangent to ΔE vs T at $T = T_m$ and is described as the value of ΔE linearly extrapolated from the temperature of measurement (T_m) to absolute zero. See figure 4.5.

Equation 4.24 now becomes:

$$\ln(FT^{-r}) = \ln D - \frac{\Delta E_o(T_m) + G(T_m)T}{kT} = \ln D^1 - \frac{\Delta E_o(T_m)}{kT} \qquad \ldots 4.26$$

where D^1 now includes the constant $G(T_m)/k$. The slope of $\ln(FT^{-r})$ vs T^{-1} is now:

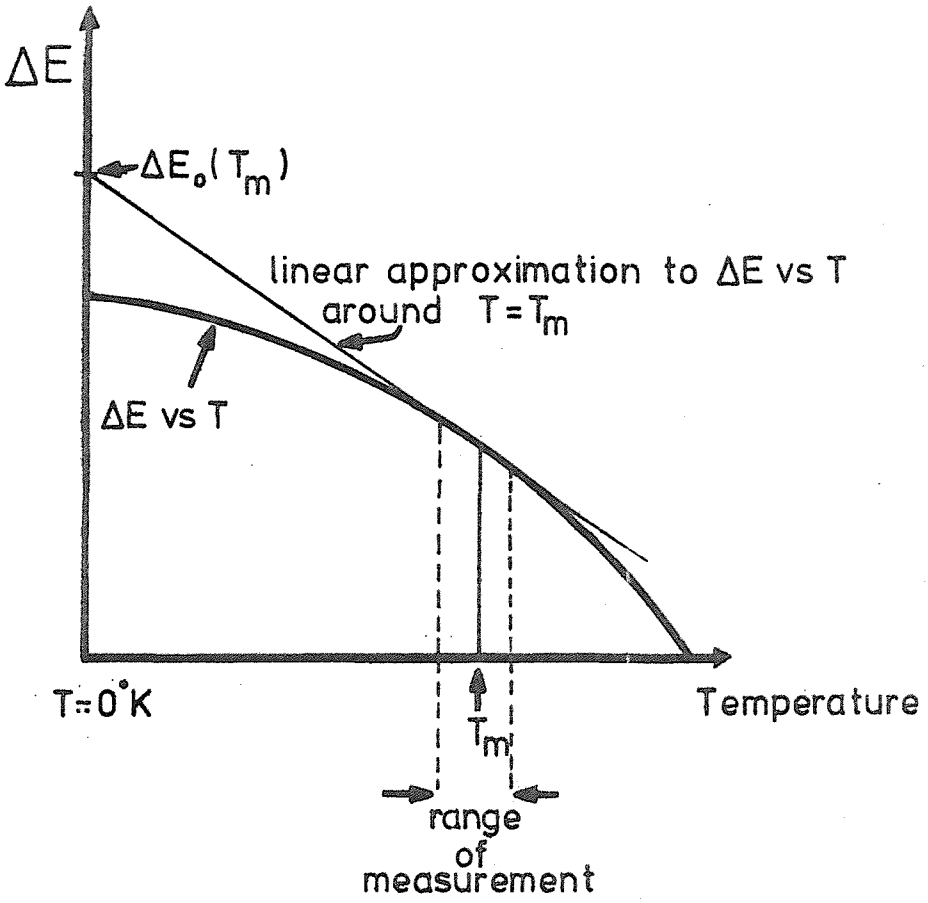


FIG. 4.5: Linearly extrapolated absolute zero energy level ΔE_o T_m.

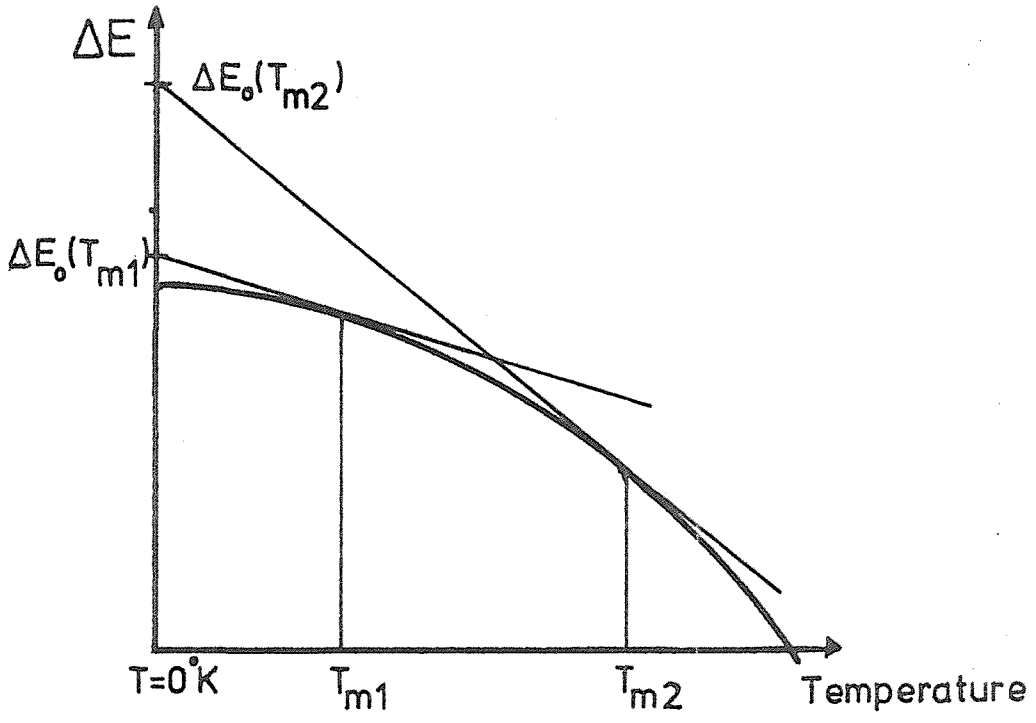


FIG. 4.6: Variation of $\Delta E_o(T_m)$ with T_m.

$$\frac{-\Delta E_o(T_m)}{k}$$

and hence $\Delta E_o(T_m)$ as defined above (equation 4.25) may be obtained. ($\ln D^1$ is the intercept at T = 0).

In some of the works listed in table 4.2, the value $\Delta E_o(T_m)$ is not the result obtained, but the "activation energy" at the temperature of measurement, T_m, of the property F is calculated. The activation energy is obtained by plotting ln(F) against T^{-1} instead of $\ln(FT^{-r})$. From equation 4.26:

$$\ln F = \ln D^1 - r \ln T - \frac{\Delta E_o(T_m)}{kT} \qquad \ldots 4.27$$

The slope of this plot will be:-

$$\left.\frac{\partial(\ln F)}{\partial(T^{-1})}\right|_{T = T_m} = -\frac{\Delta E_o(T_m)}{k} - rT_m \qquad \ldots 4.28$$

which yields the activation energy of F at T_m

$$E_{act}(T_m) = \Delta E_o(T_m) + r\,k\,T_m \qquad \ldots 4.28A$$

$\Delta E_o(T_m)$ is simply related to the activation energy by this equation and may be obtained from it for appropriate values of r and T_m. It is common for ln F to be plotted against $(kT)^{-1}$, which has units of $(\text{energy})^{-1}$, in which case equation 4.28A is obtained directly from the slope.

Clearly there are several problems involved in obtaining the actual value of ΔE at any temperature T from the type of measurements listed in table 4.2. Initially the value of r, the exponent of T in equation 4.23, must be known. Subsequently a careful interpretation of the value $\Delta E_o(T_m)$, obtained from the results, must be made. From this, the value of ΔE may only be obtained if its temperature dependence is known.

If the variation of ΔE with temperature were linear at all temperatures, $\Delta E_o(T_m)$ would be the same for all T_m and equation 4.25 would not be an approximation near a particular T_m but would be valid over the whole temperature range. However, if ΔE does not vary linearly with temperature, the value of $\Delta E_o(T_m)$ will depend on the actual value of T_m, the temperature of the original measurements.

In carrying out the analysis of experimental results which have been obtained over a range of values of T_m, it has been assumed that at least within the range of T_m the variation of ΔE is linear (equation 4.25). This assumption is reasonable for small temperature ranges and will be seen to be so if the points plotted as $\ln(FT)^{-r}$ vs T^{-1} provide a good straight line. In practice this is so in the measurements listed in table 4.2. Measurements in the range 100 - 200°K, however, will not yield the same value of $\Delta E_o(T_m)$ as measurements made in the range 150 - 200°K if $\Delta E(T)$ is non-linear. This is illustrated in figure 4.6.

4.52.3 Interpretation of results for various models for $\Delta E(T)$

As suggested in Section 4.3, there are several reasonable models for the variation of the gold energy levels with temperature. For the example of the gold acceptor level considered above ($\Delta E = E_c - E_A$) these are:

i) That it remains a fixed distance from the conduction band edge as temperature varies,

ii) that it remains a fixed distance from the valence band edge as temperature varies,

or iii) that ΔE varies in proportion to E_g, the silicon energy band gap, as temperature varies.

The value obtained experimentally, $\Delta E_o(T_m)$ may be related to the desired value of ΔE at any temperature for each of these cases as follows:

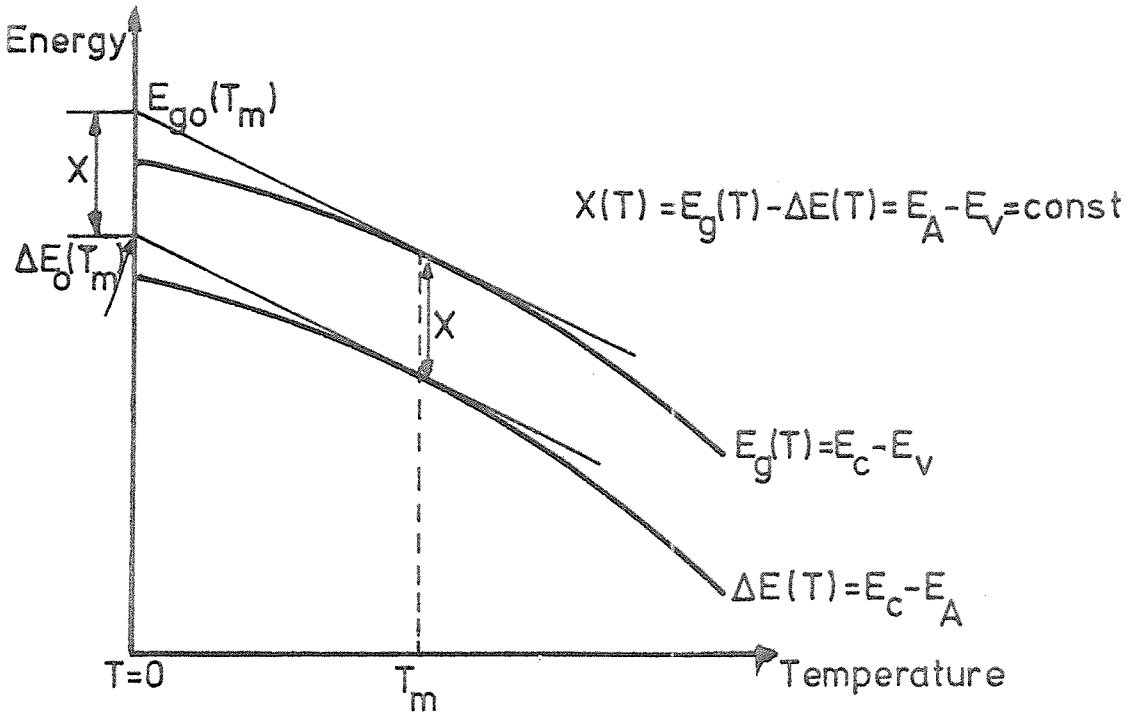


FIG. 4.7: Relationship between $\Delta E_o(T_m)$ and $E_{g_o}(T_m)$ when ΔE remains fixed to the valence band edge as temperature varies.

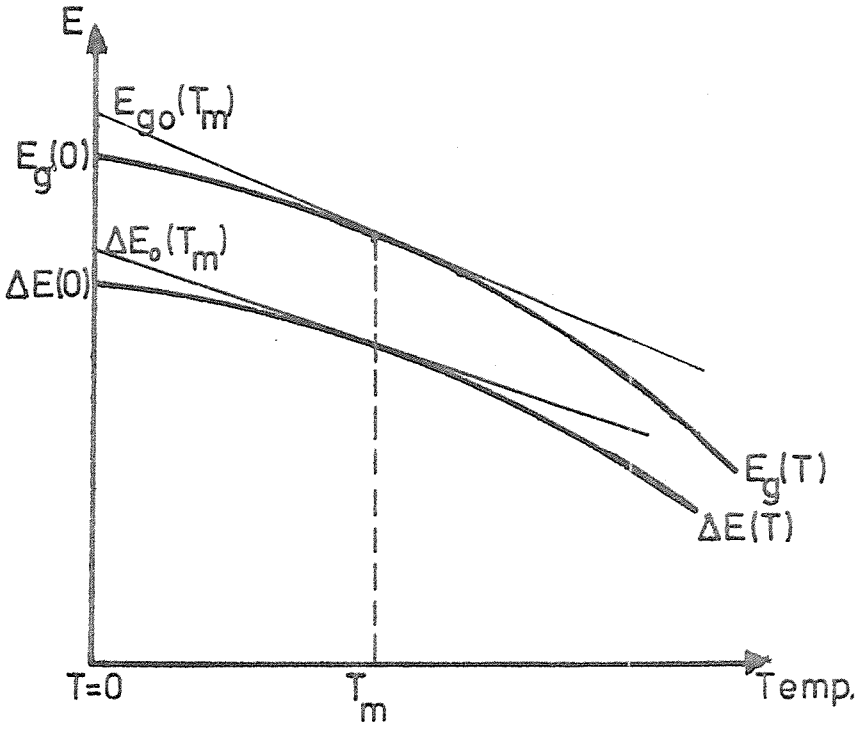


FIG. 4.8: Relationship between $E_{g_o}(T_m)$ and $\Delta E_o(T_m)$ when ΔE varies in proprotion to $E_g(T)$.

Case i If the gold acceptor level remains a fixed distance from the conduction band edge, $\Delta E = E_c - E_A$ is constant. Equation 4.25 is valid for all temperatures and $G(T_m)$ is zero because ΔE plotted against T is a horizontal straight line:

$$\Delta E(T) = \Delta E_o(T_m) = \text{constant} \qquad \ldots 4.29$$

Thus the slope of (FT^{-r}) against T^{-1} directly yields the activation energy of the gold acceptor level relative to the conduction band edge only if the gold acceptor remains a fixed distance from the conduction band edge as temperature varies.

Case ii If the gold acceptor level remains a fixed distance from the valence band edge, $E_A - E_v$ is constant and as temperature increases $E_c - E_A$ ($= \Delta E$) is reduced at the same rate as $E_c - E_v = E_g$, the silicon energy band gap.

In figure 4.7, the temperature variations of E_A relative to E_c, ($= \Delta E(T)$) and E_g ($= E_c - E_v$) are shown. The energy difference between the two curves at any temperature is $(E_c - E_v) - (E_c - E_A)$; which is equal to $E_A - E_v$ (shown as X on the figure). For this case, $E_A - E_v = X$ is constant by definition. The quantity obtained from the experimental results, $\Delta E_o(T_m)$, defined in equation 4.25 and shown in figure 4.7, must be related to ΔE at any temperature if a useful activation energy value is to be obtained. ΔE at any temperature may be seen in figure 4.7 to be:

$$\Delta E = E_g(T) - X \qquad \ldots 4.30$$

$E_g(T)$ has been measured accurately by Bludau et al., (1974) and X may be obtained, as shown in figure 4.7, by subtracting the measured value of $\Delta E_o(T_m)$ from $E_{go}(T_m)$, where $E_{go}(T_m)$ is the intercept at 0^oK of a tangent drawn at $T = T_m$ to $E_g(T)$. This value may be obtained from the data of Bludau et al., (see figure 4.9). For a given set of measurements made around a temperature T_m, which yield a value $\Delta E_o(T_m)$, the general result for $\Delta E(T)$, if the gold acceptor energy level is fixed to the valence band edge, is:

$$\begin{aligned} \Delta E(T) &= E_g(T) - X \\ &= E_g(T) - \left[E_{go}(T_m) - \Delta E_o(T_m)\right] \end{aligned} \qquad \ldots 4.31$$

Case iii Figure 4.8 illustrates the situation which occurs if ΔE varies in proportion to E_g as temperature changes.

The ratio $\Delta E(0) : E_g(0)$ is constant by definition, and the value of ΔE at any temperature T is:

$$\Delta E(T) = E_g(T) \left[\frac{\Delta E(0)}{E_g(0)}\right] \qquad \ldots 4.32$$

Once again it is necessary to relate $E(T)$ to the experimentally obtained value $\Delta E_o(T_m)$ and to $E_{go}(T_m)$ which can be obtained from published data. If the variation of E_g with temperature is represented by a polynomial:

$$E_g(T) = a + bT + cT^2 + dT^3 \ldots \qquad \ldots 4.33$$

then,

$$E(T) = E_g(T) \cdot K = K \cdot (a + bT + cT^2 \ldots) \qquad \ldots 4.34$$

where K is the constant obtained from the ratio in equation 4.32.

The tangents at $T = T_m$ to the curves represented by equations 4.33 and 4.34, may be obtained by differentiation with respect to temperature:

$$\frac{\partial}{\partial T} E_g(T)\Big|_{T = T_m} = b + \frac{1}{2} cT + \frac{1}{3} dT^3 \ldots \qquad \ldots 4.35$$

and

$$\frac{\partial}{\partial T} \Delta E(T)\Big|_{T = T_m} = |K| \left[b + \frac{1}{2} cT + \frac{1}{3} dT^3 \ldots\right] \qquad \ldots 4.36$$

The intercept with $T = 0$ of the tangent to $E_g(T)$ at $T = T_m$ is given by 'b' from equation 4.35, and the intercept with $T = 0$ of the tangent to $\Delta E(T)$ at $T = T_m$ is equal to $(K) \cdot b$ from equation 4.36. The required constant of proportionality, K, is given by the ratio of these two intecepts, shown on figure 4.8:

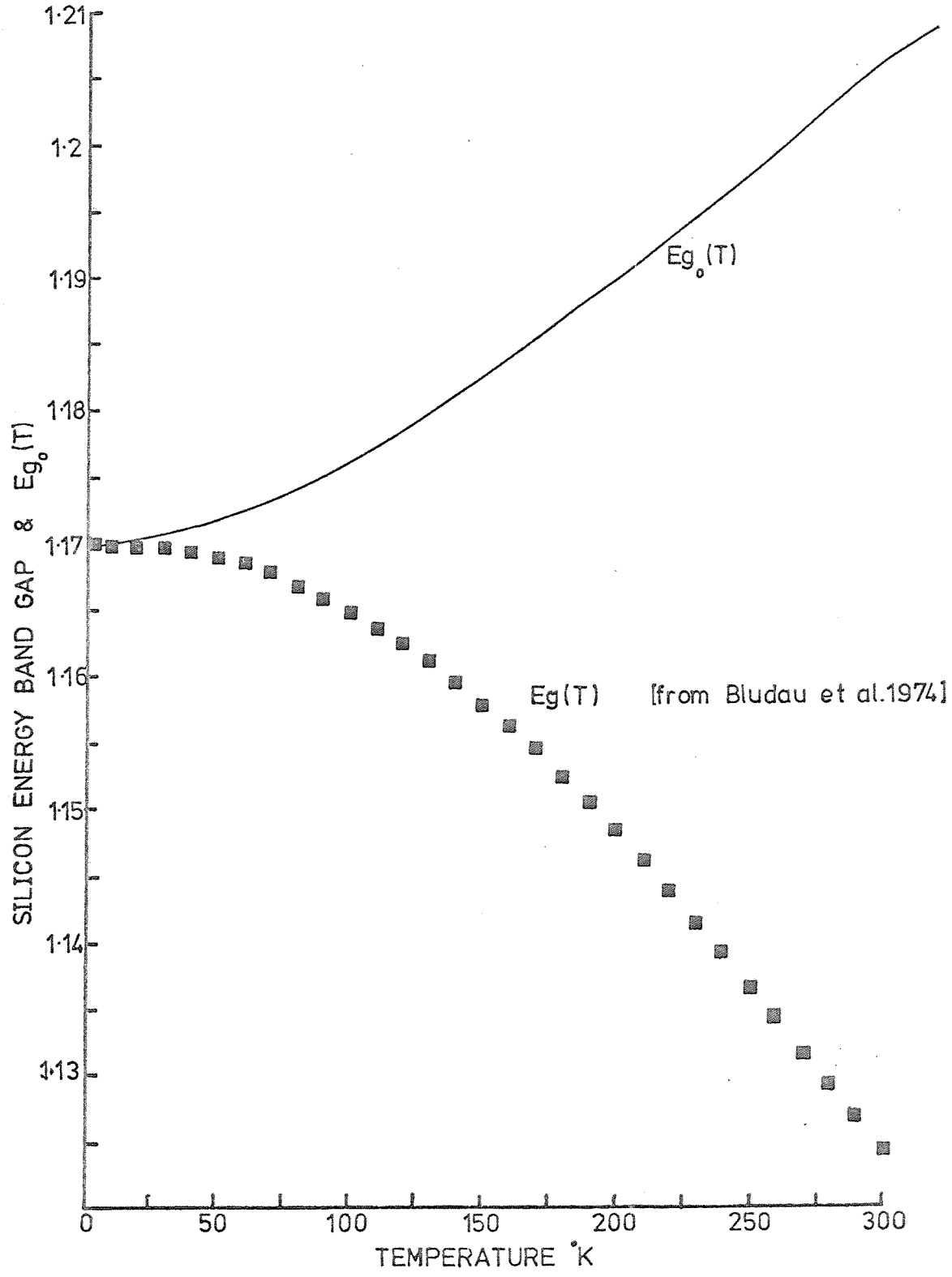


FIG. 4.9: Eg(T) vs T and $Eg_o(T_m)$ vs T_m.

$$K = \frac{\Delta E(0)}{E_g(0)} = \frac{\Delta E_o(T_m)}{E_{go}(T_m)} \qquad \ldots 4.37$$

The required relationship for case iii, is:

$$\Delta E(T) = E_g(T) \left[\frac{\Delta E_o(T_m)}{E_{go}(T_m)}\right] \qquad \ldots 4.38$$

4.52.4 Concluding remarks

Equations 4.29, 4.31 and 4.38 show that the value of ΔE at any temperature (including T = 0) depends on the model of energy level temperature dependence chosen. For a given set of measurements from which one value of $\Delta E_o(T_m)$ is obtained, there are three possible values of ΔE at a given temperature for the three models considered here. In order to obtain $\Delta E(T)$ for the second and third cases knowledge of $\Delta E_o(T_m)$ alone is not adequate, the value of T_m is also necessary so that $E_{go}(T_m)$ may be obtained (see below).

The measurements of activation energies and linearly extrapolated absolute zero gold energy levels which are reported in the literature and are listed in table 4.2, have not been analysed in this way. Unless $\Delta E(T)$ is linear, or all of the different measurements have been made at the same T_m (which is not the case), these reported gold energy levels cannot be compared validly. Reanalysis as described above must first be carried out.

The values of $E_{go}(T_m)$, which are required, have been obtained, in this work, from the results of E_g vs T given by Bludau et al., (1974). Figure 4.9 shows E_g plotted as a function of temperature and also shows $E_{go}(T_m)$ on the same axes. In order to obtain $\Delta E(T)$, the value of $E_g(T)$ and $E_{go}(T_m)$ may both be taken directly from this figure.

The reanalysis of data obtained from references listed in table 4.2 may be carried out using equations 4.29, 4.31 and 4.38. $\Delta E_o(T_m)$ is provided directly by plots of $\ln(FT^{-r})$ against T^{-1}. Where plots of lnF against T^{-1} are given, equation 4.28A, with an appropriate value of 'r' may be used to obtain $\Delta E_o(T_m)$.

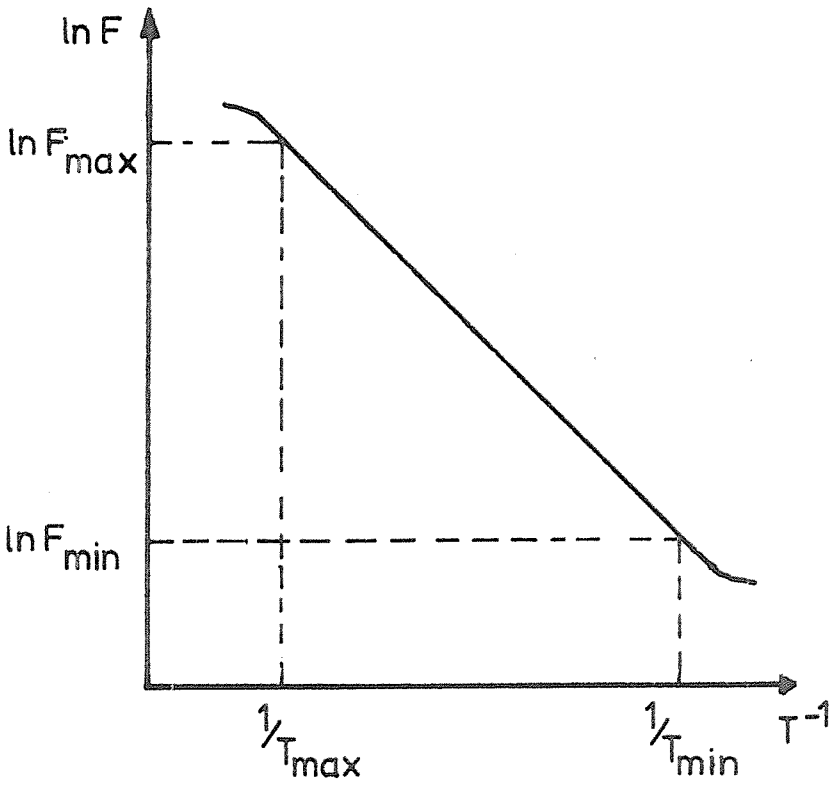


FIG. 4.10: Illustration of $\partial(\ln F)/\partial (T^{-1})$

The only approximation made in the analysis is that the temperature variation of the energy level in question is linear over the range of measurement (equation 4.25). If this temperature range is T_{min} to T_{max}, a single value of T_m, which gives an $E_{go}(T_m)$ characteristic of the range, must be obtained. Equation 4.23 gives the general form of the measured property, F, thus:

$$F = DT^r \exp\left[-\frac{\Delta E}{kT}\right] \qquad \ldots 4.23$$

After a linear approximation to the temperature variation of ΔE between T_{min} and T_{max} had been made (equation 4.25) it was shown in equation 4.28, that the slope of lnF against T^{-1} is given by:

$$\left.\frac{\partial(\ln F)}{\partial\, T^{-1}}\right|_{T = T_m} = -\frac{\Delta E_o(T_m)}{k} - rT_m \qquad \ldots 4.28$$

Figure 4.10 shows a typical plot of some lnF against T^{-1}, linear over the range T_{min} to T_{max}. The slope of the linear portion of this plot is:

$$\frac{\partial(\ln F)}{\partial\, T^{-1}} = \frac{\ln(F_{max}) - \ln(F_{min})}{T_{max}^{-1} - T_{min}^{-1}} \qquad \ldots 4.39$$

Substitution from equation 4.23 gives:

$$\frac{\partial(\ln F)}{\partial\, T^{-1}} = \frac{\ln\left[DT_{max}^{r} \exp\frac{-\Delta E_{max}}{kT_{max}}\right] - \ln\left[DT_{min}^{r} \exp\frac{-\Delta E_{min}}{kT_{min}}\right]}{T_{max}^{-1} - T_{min}^{-1}} \qquad \ldots 4.40$$

$$= \frac{r\ln T_{max} - r\ln T_{min} - \left[\frac{\Delta E_{max}}{kT_{max}}\right] + \left[\frac{\Delta E_{min}}{kT_{min}}\right]}{T_{max}^{-1} - T_{min}^{-1}} \qquad \ldots 4.41$$

and from equation 4.25, where

$$\Delta E(T) = \Delta E_o(T_m) + \left[G(T_m)\right] T \qquad \ldots 4.25$$

The following may be obtained:

$$\frac{\partial(\ln F)}{\partial T^{-1}} = \frac{r\left[\ln T_{max} - \ln T_{min}\right] - \left[\frac{\Delta E_o(T_m)}{kT_{max}}\right] - \left[\frac{G(T_m)}{k}\right] + \left[\frac{\Delta E_o(T_m)}{kT_{min}}\right] + \left[\frac{G(T_m)}{k}\right]}{T_{min}^{-1} - T_{min}^{-1}} \qquad \ldots 4.42$$

$$= \frac{r\left[\ln T_{max} - \ln T_{min}\right] - \frac{\Delta E_o(T_m)}{k}\left[\frac{1}{T_{max}} - \frac{1}{T_{min}}\right]}{T_{max}^{-1} - T_{min}^{-1}} \qquad \ldots 4.43$$

$$= -r\left[\frac{\ln T_{min} - \ln T_{max}}{T_{max}^{-1} - T_{min}^{-1}}\right] - \frac{\Delta E_o(T_m)}{k} \qquad \ldots 4.44$$

This is the same as the desired expression in equation 4.28, if T_m is given by:

$$T_m = \frac{\ln T_{min} - \ln T_{max}}{T_{max}^{-1} - T_{min}^{-1}} \qquad \ldots 4.45$$

because substitution of 4.45 into 4.44 gives equation 4.28:

$$\left.\frac{\partial(\ln F)}{\partial\, T^{-1}}\right|_{T = T_m} = -\left[\frac{\Delta E_o(T_m)}{k} + r\, T_m\right] \qquad \ldots 4.28$$

Equation 4.45 may be used to obtain the appropriate value of T_m for a set of measurements which are linear over the range T_{min} to T_{max}. Using this value (an effective temperature of measurement) the value of $E_{go}(T_m)$ may be obtained from figure 4.9 and used in equations 4.31 and 4.38.

4.53 Review and Reanalysis of Reported Energy Level Measurements

The possible variation with temperature of the position of the gold energy levels in silicon, has not been considered by all of the authors listed in tables 4.1 and 4.2, in the analysis of their results. Collins et al. (1957), in one of the earliest measurements of the gold energy level positions, did mention the possibility of such an effect. The brief reference states that such a temperature variation, if linear, would yield absolute zero gold energy levels from the reported measurements of Hall coefficient and resistivity as a function of temperature. No attempt was made to consider higher order temperature dependences, such as those considered in 4.52.2, which yield the energy levels linearly extrapolated to absolute zero from the temperature of measurement ($\Delta E_o(T_m)$). In many subsequent publications there is either no mention of temperature dependences or only a brief reference, such as that made by Collins et al. In 1972, Parrillo and Johnson proposed that the position of both gold energy levels varied in proportion to the variation of the silicon energy band gap with temperature. The theoretical results calculated by Parrillo and Johnson are consistent with their measured values of electron and hole emission probabilities from the gold acceptor level. Their experimental results, however, do not agree with recently published ones (Engstrom and Grimmeiss, 1974 and 1975) and certain values used in their calculations are open to question (see 4.53.1).

The problem of gold energy level variation with temperature was spotlighted by Parrillo and Johnson and recently attempts have been made to measure them directly (Engstrom and Grimmeiss, 1975; Wong and Penchina, 1975).

Measurements listed in tables 4.1 and 4.2 will now be considered, the latter first. The results and their interpretation will be reconsidered and compared where appropriate. It will become clear that the frequently quoted values of 0.54 eV for the gold acceptor relative to the conduction band edge and 0.35 eV for the gold donor relative to the valence band edge, can only be considered to be approximate.

4.53.1 Resistivity and Hall effect measurements

Taft and Horn (1954) and Carlson (1956), respectively, first reported the presence of a deep acceptor and a deep donor energy level in gold doped silicon. The first comprehensive report of a search for energy levels due to gold in the silicon energy band gap was made by Collins, Carlson and Gallagher (1957). Measurements of resistivity and Hall effect as a function of temperature were made in silicon samples doped with varying concentrations of shallow impurities (boron and phosphorus) and gold.

In n-type silicon, the addition of gold resulted in a reduction in the number of electrons in the conduction band. When sufficient gold had been added, the resistivity approached that of intrinsic silicon, indicating the presence of an acceptor energy level due to gold. Similarly a decrease in the number of holes in the valence band after the addition of gold to p-type silicon, suggested the presence of a donor level due to gold, nearer to the valence band than the gold acceptor.

For appropriate concentrations of gold relative to n-type dopant (phosphorus) it can be shown that the carrier concentration is a simple function of the position of the gold acceptor energy level relative to the conduction band edge over a certain temperature range (see appendix C). An expression of the form:

$$n \simeq K_1 T^m \exp - \left[\frac{E_c - E_A}{kT}\right] \qquad \ldots 4.46$$

is obtained, when $N_d > N_a$, $n >> p$, $n << N_d - N_a$, $n << N_{Au} - (N_d - N_a)$ and $N_{Au} > (N_d - N_a)$; n = electron concentration, p = hole concentration, N_d = shallow donor (phosphorus) concentration, N_a = shallow acceptor (boron) concentration, N_{Au} = gold concentration, E_c = conduction band edge energy, E_A = gold acceptor energy and K_1 is a constant.

Similarly, in p-type silicon, if E_v is the valence band edge energy and E_D is the gold donor energy level, the expression obtained under appropriate conditions is:

$$p \simeq K_2 T^m \exp - \left[\frac{E_D - E_v}{kT}\right] \qquad \ldots 4.47$$

for $N_a > N_d$, $p >> n$, $p << N_a - N_d$, $p << N_{Au} > (N_a - N_d)$ and $N_{Au} > (N_a - N_d)$.

Both resistivity and Hall coefficient are dependent on the carrier concentration (n or p) and, as shown in 4.52.1 and 4.52.2, measurement of the temperature dependence of either will result in a quantity ($\Delta E_o(T_m)$) related to the appropriate gold energy level. Collins et al. carried out measurements of both Hall coefficient and resistivity on samples for which the conditions described above were appropriate. The slopes of their plots of the logarithm of Hall coefficient or resistivity against reciprocal temperature were corrected for the temperature dependences included in 'r' in equation 4.23 (section 4.52.1). The results which were thus obtained are the commonly quoted values for the gold energy level positions: $E_c - E_A = 0.54$ eV and $E_D - E_v = 0.35$ eV. The analysis presented in 5.52.2 has shown, however, that these are in fact the absolute zero values of the energy levels linearly extrapolated from the temperature of measurement. (A reanalysis of these results, using more recent values for 'r', may be found in section 4.53).

Collins et al. also searched for another acceptor energy level closer to the conduction band edge by gradually changing the ratio of

N_{Au} to $(N_d - N_a)$ from $N_{Au} > (N_d - N_a)$ to $2N_{Au} > (N_d - N_a) > N_{Au}$

over several samples. Measurements of resistivity as a function of temperature showed a change in dependence from 0.54 eV, characteristic of the gold acceptor level $(E_c - E_A)$, to 0.045 eV characteristic of the shallow donor level $(E_c - E_d)$. There was no behaviour which suggested the presence of another energy level. A similar experiment in p-type silicon (boron doped) carried out by Taft and Horn (1954) exhibited similar results. The indications of these experiments were that any possible shallow energy levels due to gold in n-type or p-type silicon would have ionization energies comparable to phosphorus or boron if they existed at all. The maximum gold concentration used in these experiments was 1.1×10^{16} atoms. cm^3. The recent discovery of a shallow acceptor energy level associated with gold (see section 4.4) was made for gold concentrations greater than 4.0×10^{16} atoms. cm^{-3}.

The deep gold acceptor level found by Collins et al. is located very near to the middle of the silicon energy band gap and as a result the addition of enough gold to initially n-type silicon can shift the Fermi level so far towards the valence band that the sample becomes p-type, (see chapter 7). Under such conditions the carrier concentration, (now predominantly holes) begins to increase and the behaviour of resistivity or Hall coefficient as a function of temperature becomes dependent on the position of the gold acceptor level relative to the valence band edge, $(E_A - E_v)$. Collins et al. carried out measurements for samples in which the gold concentration (N_{Au}) was approximately eight hundred times the shallow donor concentration (N_d). The value obtained for $E_A - E_v$ of 0.62 ± 0.02 eV must once again be interpreted as the linearly extrapolated absolute zero value. (A similar effect is not noted in initially p-type silicon because the gold donor level is much further from the silicon conduction band edge than the gold acceptor is from the valence band edge).

A full reanalysis of these results, and comparison with other results may be found in 4.53.4. It is noteworthy, however, that the differences in the results obtained by Collins et al., from measurements of

resistivity and then Hall coefficient on the same samples, are almost eliminated when the results are reanalysed with correct temperature dependences as described in 4.52. For example: in figure 3 of Collins et al's paper, Hall coefficient measurements as a function of temperature give a slope of 0.37 eV in p-type gold doped samples ($E_D - E_v$). Resistivity measurements over the same temperature range give a slope of 0.33 eV - a difference of 0.04eV. After reanalysis the difference is only 0.003 eV (the absolute value depending on which model is chosen for the variation of the gold energy level with temperature).

Boltaks, Kulikov and Malkovich (1960) report similar measurements to those of Collins et al. The results, $E_D - E_v = 0.36 \pm 0.02$ eV and $E_c - E_A = 0.54 \pm 0.02$ eV are given but no details of the experiments or the way in which the data was analysed are mentioned. It is not clear, therefore, whether any corrections were made for the temperature dependences of various other factors (see equations 4.18 and 4.20). Any further consideration of these results would be of little value.

Bruckner (1971) also describes measurements of resistivity and Hall coefficient on p-type gold doped silicon samples. An expression for p_o, the carrier concentration in the silicon valence band for a shallow acceptor concentration N_a and a gold donor concentration N_D is given. Limiting cases for various ratios of N_{Au} to N_a are derived.

The expression (see appendix C) is:

$$\frac{p_o(p_o + N_D - N_a)}{N_a - p_o} = N_v \frac{1}{g_D} \exp\left[\frac{-(E_D - E_v)}{kT}\right] \quad \ldots 4.48$$

where g_D is the degeneracy factor for the gold donor level (see 4.2 and 4.6). At high temperatures, the carrier concentration is dominated by N_a and p_o is given by:

$$p_o = N_a. \quad \ldots 4.49$$

At low temperatures, the situation is similar to that described for the experiments of Collins et al.; where p_o has an exponential temperature dependence and the slope of p_o (inferred from resistivity or Hall coefficient measurements) against reciprocal temperature gives the linearly extrapolated absolute zero value of E_D (ie: $\Delta E_o(T_m)$). The result is quoted, by Bruckner, as being the "activation energy of the gold donor". No allowance is made for any variation of $E_D - E_v$ with temperature. Bruckner carried out experiments for ratios of N_{Au} to N_a of 1, 5 and 100 to 1000. The results for $N_{Au} = N_a$ should have resulted in a special case for equation 4.44 at low temperatures, for which the slope of p_o against reciprocal temperature should have equalled $(E_D - E_v) \times \frac{1}{2}$. The condition of $N_{Au} = N_a$ was found to be impossible to obtain experimentally due to inhomogeneities in the distribution of gold in any given sample. The case of $N_{Au} = 5N_a$ was the one for which the slope of p_o against T^{-1} at low temperatures gave "the activation energy of the gold donor". This value was then used to fit equation 4.44 to the experimental results with g_D as a parameter. The "best fit" was obtained for $g_D^{-1} = \frac{1}{16}$, however as the value used for $E_D - E_v$ was incorrectly interpreted from the experimental results (no allowance for temperature variation was made), this value of g_D^{-1} is hardly reliable (for further discussion see sections 4.2 and 4.6).

The most significant result obtained by Bruckner, is due to the effect of increasing the concentration of gold relative to the shallow acceptor to a very high value. The slope of p_o against T^{-1} at low values of $N_{Au} : N_a$ was characteristic of the gold donor level, but as the gold concentration became greater than 2×10^{16} atoms. cm^{-3} (for $N_a = 1.5 \times 10^{14}$ atoms. cm^{-3}), the behaviour changed rapidly. The curves assumed a slope characteristic of an energy level 0.033 eV above the valence band edge, behaving as a shallow acceptor. The addition of gold to a p-type sample is expected to compensate it (under the action of the gold donor level) and for a given temperature and shallow acceptor concentration, an increase in the amount of gold should be reflected by a decrease in the number of carriers (holes) in the valence band, p_o.

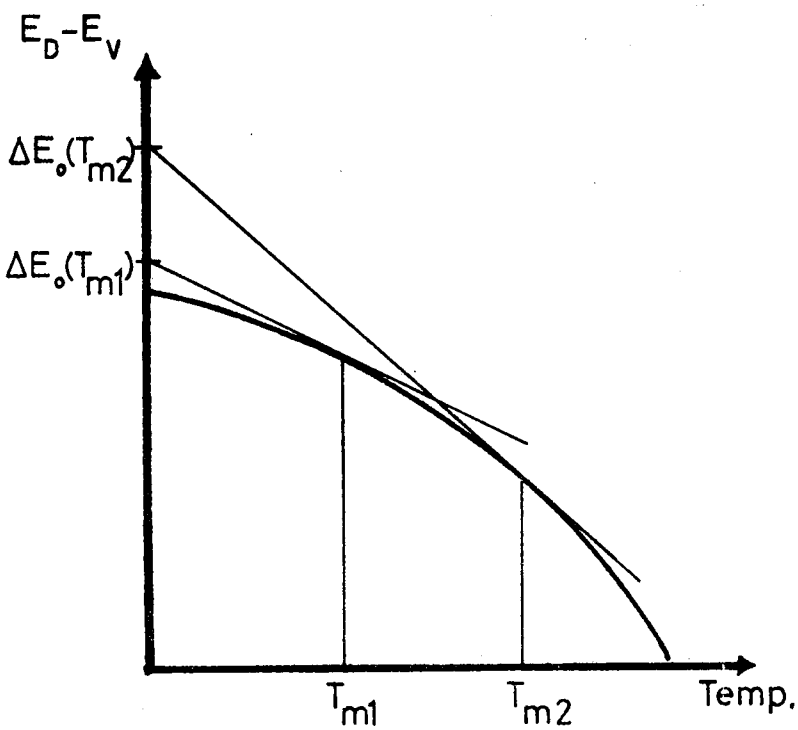


FIG. 4.11: A possible interpretation of the different results obtained by Brückner, 1971.

This behaviour was observed until the gold concentration reached 2×10^{16} atoms. cm^{-3}, beyond which p_o began to increase, hence the conclusion that the gold was introducing large numbers of shallow acceptors. No definite explanation of this level was offered by Bruckner, who did uggest however, that it might be caused by gold forming an electrically active complex with other impurities during the cooling of the samples from the diffusion temperature. This suggestion was prompted by the discovery that the concentration of this "gold-coupled shallow acceptor" increased in samples doped with more than 2×10^{16} atoms. cm^{-3} of gold when they were annealed at $600^{o}C$ for several hours.

Bruckner also observed that when the ratio of gold atoms (N_{Au}) to shallow acceptors (N_a) was increased, there was an apparent increase in the slope of plots of Hall coefficient against reciprocal temperature from 0.35 to 0.37 eV. This was explained by the possible influence of excited quantum states of the gold donor energy level, (see section 4.21). The measurements were carried out over different temperature ranges and a more probable explanation is that a higher value of $\Delta E_o(T_m)$ was being obtained for a higher T_m (see 4.52.1). Reinterpreting this result in terms of the variation of the position of the gold donor relative to the valence band edge, suggests that $E_D - E_v$ is not constant but becomes smaller as temperature increases; this is illustrated in figure 4.11 and has been discussed in section 4.52.3.

Thurber, Lewis and Bullis (1973) also carried out a series of experiments similar to those of Collins et al. (1957). The gold concentrations relative to the shallow acceptor and donor concentrations were chosen in various samples to enable the positions of all three energy levels associated with gold to be studied. These three levels - the deep acceptor at E_A, the deep donor at E_D and the gold-coupled shallow acceptor at E_G - were obtained by measurements of Hall coefficient and resistivity as a function of temperature. The equations governing the limiting cases are similar to those of Collins et al. and are discussed in appendix C. The gold coupled shallow acceptor level was measured in lightly boron doped

silicon samples (10^{13} atoms. cm^{-3}) which were heavily doped with gold (10^{17} atoms. cm^{-3}) at 1200°C and subsequently annealed at 600°C. Resistivity measurements as a function of temperature gave an activation energy of 0.0349 eV, similar to that reported by Bruckner. Thurber et al. recognised that the values of "activation energy" obtained for the deep gold energy levels were in fact linearly extrapolated absolute zero values, but carried out no further analysis. Their measurements of resistivity as a function of temperature were not corrected for the temperature dependence of mobility (see equation 4.21). A reanalysis of Thurber et al's data for each of the gold level temperature variations discussed earlier (4.52.3) may be found in section 4.53.4.

4.53.2 Thermal emission measurements

The measurements of gold energy levels remaining to be discussed all involve the property that is exhibited by deep lying energy levels whereby they may act as carrier recombination and generation centres in the silicon energy band gap. This "trapping" behaviour has been described by Shockley and Read (1952) and Hall (1952). The theory proposed by these authors has been extremely successful in explaining many phenomena observed in semiconductors (see, for example, Grove, 1967) and is usually referred to as the S.R.H. (Shockley-Read-Hall) trapping theory. A more detailed consideration of the S.R.H. theory may be found in appendix D.

For the present, the results arising from consideration of the behaviour of deep gold traps (or energy levels) in the silicon energy band gap will be considered.

For a given trap energy level E_t (which may be the gold acceptor level or the gold donor level) S.R.H. statistics give an expression for the electron emission probability of the trap, e_n, which is the probability that an electron will jump from an occupied centre into the conduction band:

$$e_n = g_t \, V_{th} \, \sigma_n \, N_c \exp\left[\frac{-(E_c - E_t)}{kT}\right] \qquad \ldots 4.50$$

where V_{th} is the thermal velocity of an electron; σ_n is the electron capture cross section of the trap - a measure of how close an electron must come to the centre to be captured (see chapter 5 and appendix D) g_t is the degeneracy factor for the trap and N_c is the density of available states in the conduction band. Similarly, the emission probability of a hole from the trap to the valence band is given by:

$$e_p = \frac{1}{g_t} V_{th} \sigma_p N_v \exp\left[\frac{-(E_t - E_v)}{kT}\right] \quad \ldots 4.51$$

where σ_p is the hole capture cross section of the trap and N_v is the density of available states in the valence band. These equations are of the same form as equation 4.19. Measurements of the thermal emission rates of carriers from the gold levels as a function of temperature may be used to obtain the activation energies $E_c - E_t$ and $E_t - E_v$.

Certain problems arise in obtaining the activation energies by such measurements. Initially the experimental conditions must be such that it is certain which gold energy level is responsible for the observed phenomena. Some of the methods used require an accurate knowledge of the number of impurity centres present or the carrier concentration during the experiment - both of which are difficult to measure and can introduce considerable inaccuracies. Many of the methods involve very fast recombination times, the inclusion of more than one time constant or time constants with non-exponential behaviour - all of which introduce uncertainties into the results. Even when apparently accurate and unambiguous results have been obtained, the possible temperature dependences of g_t, V_{th} and σ_n or σ_p in equations 4.50 and 4.51 must be known in order to derive the required activation energy (see section 4.52.2). The degeneracy factor g_t is considered to be temperature independent (see section 4.2) and the average thermal velocity increases as the square root of temperature ($V_{th} \equiv \sqrt{3kT/m}$). The temperature dependence of σ, the capture cross section, is not well known. Bemski (1958) predicted that each of the four possible capture cross sections for gold may have different temperature dependences. Capture cross section will be considered in more detail in chapter 5.

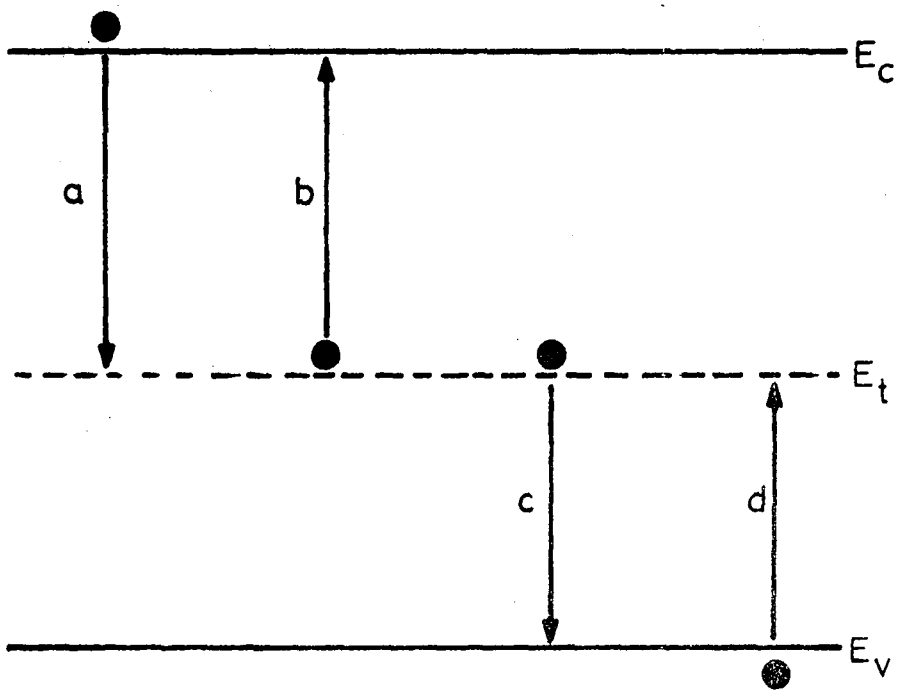


FIG. 4.12: Transitions in S.R.H. capture processes. Arrows denote electron transitions.

a) Electron capture by trap

b) Electron emission to conduction band

c) Hole capture by trap

d) Hole emission to valence band

For a single impurity (trap) level, there are four thermal transition processes, illustrated in figure 4.12, where the arrows show the directions of electron transitions. A gold centre may exist in one of three charge states; positive, neutral or negative (ionized donor, neutral or ionized acceptor). In the situations considered here, the gold centres exist either predominantly as acceptors or predominantly as donors. The two possible trapping behaviours of the acceptor or of the donor may be considered separately. S.R.H. statistics describe the behaviour of a trap which can exist in two charge states and may therefore be applied to either the gold acceptor or the gold donor separately. Each trapping level has two capture cross sections, one for electrons and one for holes. In the case of gold where there are two trapping levels being considered separately, there are four capture processes:

i) Capture of an electron by the neutral (unionized) acceptor.

ii) Capture of a hole by the negatively charged (ionized) acceptor.

iii) Capture of an electron by the positively charged (ionized) donor.

iv) Capture of a hole by the neutral(unionized) donor.

The capture cross sections associated with capture of a carrier by a neutral centre (i and iv) are generally considered to be temperature independent. There is some disagreement about the temperature dependences of the capture cross sections associated with processes ii and iii.

Details of the methods used to obtain the thermal emission rates will not be given here, a brief discussion may be found in appendix F.

Sah, Forbes, Rosier, Tasch and Tole (1969) measured thermal emission rates of electrons and holes at the gold acceptor level and of holes at the gold donor level. The measurements were made in the "high electric field" region of a reverse biased p.n. junction. It is expected that the capture cross section is electric field dependent (Lax, 1960) and the thermal emission rate is also field dependent (Frenkel, 1938). Both of these effects are expected at fields above 10^5 volts. cm^{-1} and when observed

by Sah et al. for the gold acceptor emission rates, they extrapolated the results to fit the exponential dependence of their low field data. The results for the emission rate of holes from the gold donor were also noted to be field dependent but were not corrected as all of the results were obtained at Sah et al's maximum field of 10^5 volts. cm^{-1}. More recently, Braun and Grimmeiss (1973) have shown that these "observed" field dependences were in fact due to an incomplete formulation for the reverse current in the p.n. junction. The results were fitted to an equation of the form:

$$e = A \left[\frac{T}{300}\right]^2 \exp\left[\frac{-\Delta E}{kT}\right] \qquad \ldots 4.52$$

where ΔE is the activation energy corresponding to the emission rate e. This equation, when compared with equations 4.50 and 4.51, implies that the temperature dependences of the thermal velocity V_{th}, the capture cross section σ, and the density of states N_v or N_c are all given by T^2. The temperature dependences of thermal velocity ($T^{\frac{1}{2}}$) and of the density of states ($T^{3/2}$) provide the factor of T^2. No allowance, therefore, has been made for the temperature dependence of the capture cross section. In the cases of electron emission from the gold acceptor and hole emission from the gold donor, the capture cross sections are expected to be temperature independent but in the case of hole emission from the gold acceptor the capture cross section is not. The results are not interpreted by Sah et al. in terms of any temperature variation of the gold energy levels. The value of 547.23 ± 2.17 meV for $E_c - E_A$ is probably the only reliable one from this paper if it is reinterpreted as a linearly extrapolated value of $E_c - E_A$ from the temperature of measurement to absolute zero ($\Delta E_o(T_m)$). Only if the gold acceptor remains a fixed distance from the conduction band edge as temperature varies will the experimental value be valid at all temperatures - see section 4.52.3. The value of $E_A - E_v$ is unreliable because the temperature dependence of the capture cross section for holes at the gold acceptor is ignored and the value of $E_D - E_v$ is unreliable because of the incorrect interpretation of a field dependence of the hole emission rate from E_D.

Sah et al. also comment on the degeneracies of the gold levels having fitted their energy level data back into equations 4.50 and 4.51. A similar comment to that made about Bruckner (1971) applies here - a discussion may be found in section 4.6.

Tasch and Sah (1970) used the impurity photovoltaic effect in p-i-n silicon junctions to obtain thermal and optical emission rates of holes and electrons from the gold acceptor level (p = p-type, i = intrinsic and n = n-type). The results of thermal emission rates were treated in the same way as Sah et al. (1969) by fitting them to equation 4.52. The comments already made regarding the capture cross section temperature dependences apply equally to these results. Tasch and Sah also measure the effect of high electric fields ($> 10^4$ v. cm^{-1}) on the thermal emission rates and find only a small effect, contrary to their predictions and the earlier predictions of Sah et al. (1969). The observed small dependence is explained in terms of the influence of excited quantum states of the gold acceptor but, as already stated, Braun and Grimmeiss (1973) have indicated incomplete formulation in the analysis. The presence of excited states of the gold acceptor level would lead to a temperature dependent degeneracy, see section 4.2. The reliable result for $E_c - E_A$ from this paper is 545.48 ± 1.36 meV provided, once again, it is interpreted as a linearly extrapolated value from the temperature of measurement to absolute zero.

Parrillo and Johnson (1972) proposed that the gold energy levels varied in proportion to the silicon energy band gap with temperature. This proposal was made on the basis of experimental measurements of the hole and electron emission probabilities from the gold acceptor level. Parrillo and Johnson recognised the importance of including the temperature dependences of the hole and electron capture cross sections in calculating the activation energy of the gold acceptor (see equation 4.28A) from measurements governed by equations 4.50 and 4.51. Using results due to Bemski (1958) that σ_n (capture cross section for electrons at the gold acceptor) varies as T^0 and that σ_p (capture cross section for holes at the gold acceptor) varies as T^{-4}, the linearly extrapolated absolute zero values of $E_c - E_A$ and $E_A - E_v$ were obtained by plotting emission probabilities against inverse temperature. Using a similar expression to equation 4.28A

and their proposed model of the temperature variation of the gold acceptor level, Parrillo and Johnson were able to fit equations 4.50 and 4.51 to their measured values of e_n and e_p with a degeneracy factor of 3/2 for the gold acceptor level (taken from Bullis and Streiter, 1968). Comparison of Parrillo and Johnson's experimental data with other authors' shows considerable disagreement. At 300°K, for example, although Sah et al. (1969), Tasch and Sah (1970), Engstrom and Grimmeiss(1975)all agree with Parrillo and Johnson that the emission probability of electrons from the gold acceptor is about 10^3. sec^{-1}, the value obtained by Parrillo and Johnson for the emission probability of holes from the gold acceptor is at least an order of magnitude smaller than those given by the other authors (viz., $12.sec^{-1}$ compared with 10^2, 10^2 and 1.2×10^2. sec^{-1}). Their proposal for the temperature dependence of the gold acceptor in the range 0°K to 300°K has also been disproved recently (see section 4.53.3) by Engstrom and Grimmeiss (1974). Parrillo and Johnson, in their analysis, recognise that the slope of the curve of emission probability against inverse temperature yields a linearly extrapolated absolute zero value of the activation energy, but observe - incorrectly - that this value is independent of the temperature of measurement T_m. It has been shown earlier (section 4.52.2, figure 4.11 and equation 4.38 for Parrillo and Johnson's model) that the value of $\Delta E_o(T_m)$ does depend on T_m. Parrillo and Johnson's error arises from their assumption that the linear approximation to $\Delta E(T)$ (equation 4.25) is valid for all temperatures, whereas it is only valid for a small range around each T_m. Equations 4.50 and 4.51 are quoted incorrectly, the degeneracy factors are both reciprocated in Parrillo and Johnson's paper. The temperature dependence of σ_n,(T^o) is generally accepted to be correct, however the quoted temperature dependence of σ_p (T^{-4} from: Bemski, 1958; Tasch and Sah, 1970; and Parrillo and Johnson) has been questioned by Engstrom and Grimmeiss who argue that the correct dependence is T^{-2} (see chapter 5 and also later in this section).

It is apparent that the measurements made by Parrillo and Johnson have led them, probably incorrectly, to a model of the variation of the gold energy levels with temperature. However, they were the first authors to show that ignoring the possible temperature variation of the gold levels

and ignoring the temperature dependences of the capture cross sections, makes a significant difference to the activation energies inferred from the experimental results.

Pals (1974), in studies of the properties of deep levels due to several different impurities in silicon, has published values for the activation energies of both deep gold energy levels. Pals' technique (see appendix F) is used to measure the emission probabilities of electrons from the gold acceptor and holes from the gold donor. In both cases the temperature dependences of the capture cross sections are not included in the S.R.H. equations. The two capture cross sections concerned are both for capture at neutral trapping centres and therefore are generally considered to be temperature independent: Thus the energy levels obtained by Pals are probably reliable if interpreted as linearly extrapolated absolute zero values. The values are: $E_c - E_A = 0.56$ eV and $E_D - E_v = 0.346$ eV.

Kassing and Lenz (1974 i and ii) describe measurements of capture and emission rates at the gold acceptor energy level. A value of 0.25 eV for $E_c - E_A$ is obtained from the temperature dependence of the electron emission time constant. No details of corrections for temperature dependences are given.

Senechal and Basinski (1968), in measurements of the capacitance of gold doped silicon p-n junctions, inferred on activation energy for the gold acceptor level $E_c - E_A$ of 0.545 ± 0.005 eV. The result was obtained from the slope of the time constant associated with the capacitance variation of the junction plotted against reciprocal temperature. No corrections for the temperature dependences of capture cross sections, densities of states and thermal velocities were made.

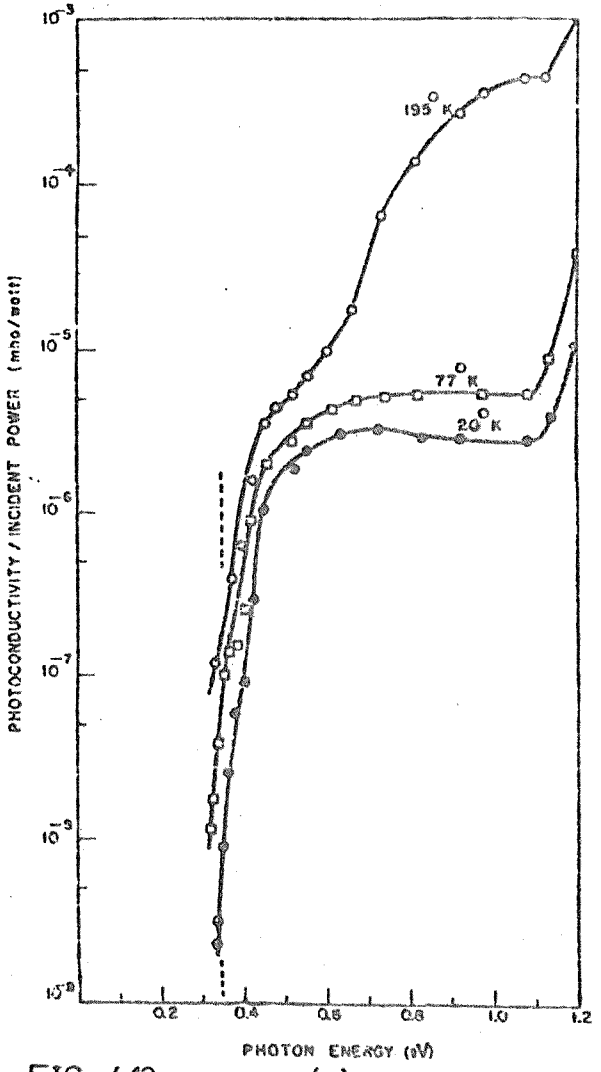
Engstrom and Grimmeiss (1975), with measurements of the thermal emission rates of electrons and holes from the gold acceptor level, have presented values for the linearly extrapolated absolute zero values of $E_c - E_A$ and $E_A - E_v$. The fact that the results are linearly extrapolated absolute zero values and not activation energies at any temperature is

discussed and comparisons made with the values obtained by other authors (all reviewed in this section). The comparisons, however, are incomplete. It has been shown here that most of the authors who have reported results of emission rate measurements have not included the correct temperature dependences in their results; before any valid comparison can be made, the results must be reanalysed. Engstrom and Grimmeiss do not attempt a reanalysis so their comparisons are not really valid. A reanalysis and complete comparison is given in section 5.53.4. Engstrom and Grimmeiss assume that the capture cross section for electrons at the neutral acceptor is not temperature dependent and they resolve the problem of the temperature dependence of the capture cross section for holes at the negative acceptor by carrying out appropriate measurements (see chapter 5). Contrary to the results of Bemski (1958) and Tasch et al (1970), they find that the temperature dependence of σ_p is $T^{-2.0 \pm 0.1}$, and not T^{-4}. The accuracy of the linearly extrapolated absolute zero values for $E_c - E_A$ and $E_A - E_v$ obtained is strengthened by the fact that their sum is almost exactly equal to the value of the linearly extrapolated absolute zero band gap from their temperature of measurement (see figure 4.8). In their measurements, Engstrom and Grimmeiss do not note any electric field dependence of capture cross sections for the field strengths used in their experiments on p-n junctions. Their results are $\left[E_c - E_A\right]_0 (T_m) =$ 0.553 eV and $\left[E_A - E_v\right]_0 (T_m) = 0.641$ eV, (sum = 1.196 eV).

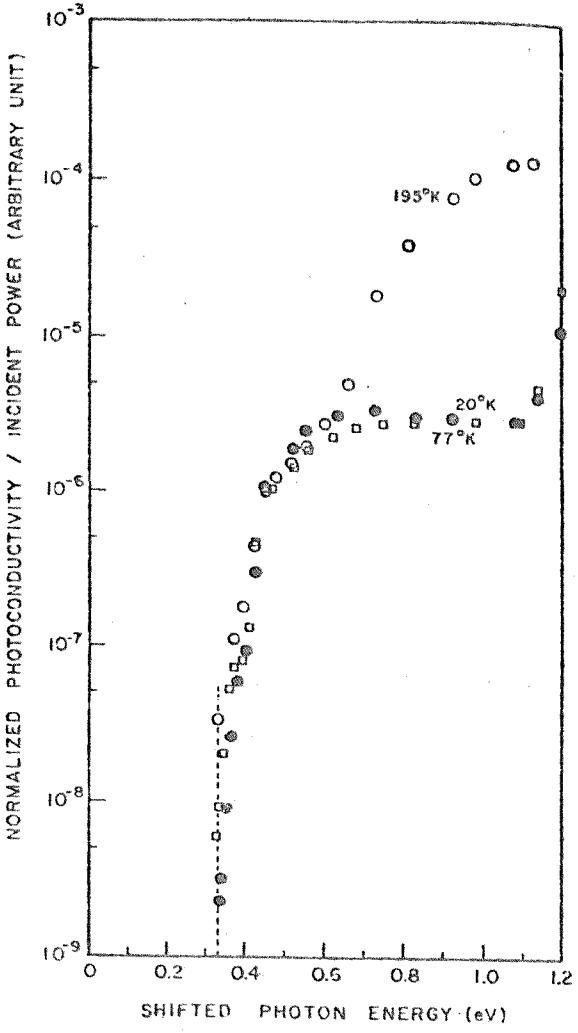
These results are considered by this author to be the most accurately obtained values for the gold energy level positions linearly extrapolated from T_m to absolute zero.

4.53.3 Single temperature measurements and temperature dependence of gold energy levels

Newman (1954), in an attempt to confirm the presence of a gold donor level located 0.33 eV above E_v, the valence band edge (originally



(a)



(b)

FIG 4.13

Data from Newman (1954) reanalysed by Wong et al. (1975).

reported by Taft and Horn, 1954), carried out measurements of the photoconductive response of p-type silicon doped with gold at 20°K, 77°K and 195°K. The results showed a distinct rise in photoconductivity for incident photons above energies of about 0.33 eV, thus confirming the presence of a hole trap energy level 0.33 eV above E_v. On closer examination, their results show a threshold shift to slightly lower energies at the higher temperatures. Wong and Penchina (1975), in a reanalysis of Newman's original data, show that this shift to lower energies is accurately predicted if it is assumed that the gold acceptor level moves towards the valence band edge by the same amount as the silicon energy band gap shrinks over the temperature range. Their conclusion is that the gold donor level remains a fixed distance from the conduction band edge as temperature varies. Figure 4.13(a) shows Newman's original figure and figure 4.13(b) shows how the results overlap after normalization to the same signal strength and shifting to the left by the amount that the silicon energy band gap has shrunk at the appropriate temperature. Wong and Penchina conclude that the gold donor level remains fixed 0.84 eV below the conduction band edge as temperature varies. This conclusion is supported, as indicated earlier, by a reinterpretation of the results of Bruckner for $E_D - E_v$ (which appeared to become smaller at higher temperatures) and was also reached from calculations of resistivity in gold doped silicon carried out in this work (see Chapter 7) before the publication of Wong and Penchina's paper.

Badalov (1970), in similar experiments to those of Newman, illustrates the photoconductive response of n-type gold doped silicon. The threshold of photoresponse, although indicative of the gold donor 0.54 eV below E_c, does not have a sharp edge and therefore does not allow an accurate measurement of $E_c - E_A$.

Engstrom and Grimmeiss (1974), in measurements of optical emission rates from the gold acceptor as a function of temperature, show that, in the temperature range 90°K to 242°K, the gold acceptor level probably remains a fixed distance from the conduction band edge. The measurements of $E_c - E_A$ are constant over the temperature range whereas the measurements of $E_A - E_v$ closely follow the shape of the variation of the silicon energy band gap with temperature measured by Bludau et al.

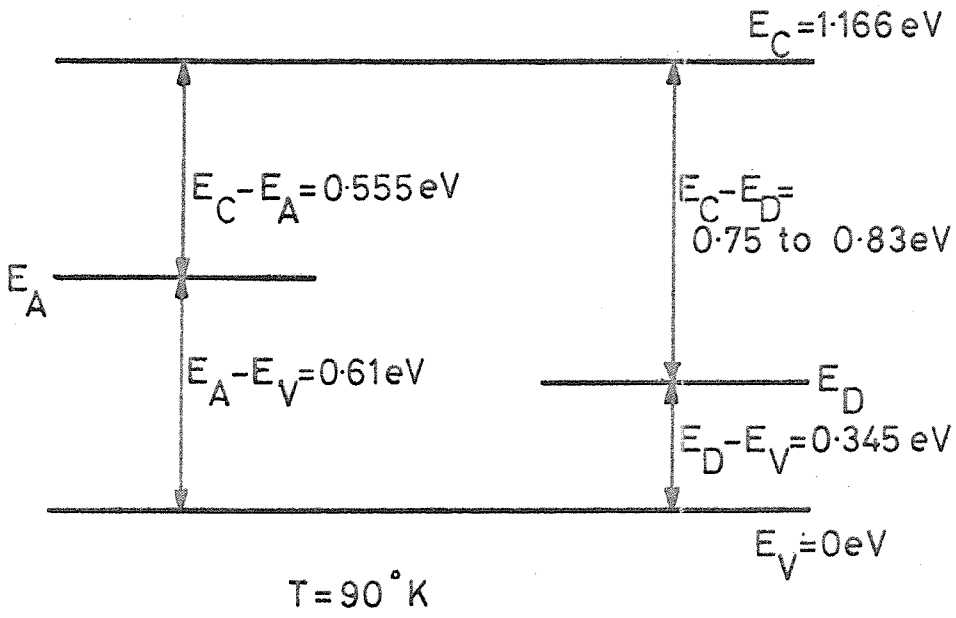


FIG. 4.14: Results from Braun and Grimmeiss, 1974.

(1974). This conclusion is, once again, confirmed in this work for calculations of resistivity as a function of temperature (Chapter 7) and calculations of solubility of gold in heavily doped n-type silicon (Chapter 6) following those of Brown et al. (1975).

Braun and Grimmeiss (1974), in a study of photon energy thresholds for electron and hole emission from both gold levels at 90°K, measure threshold energies of: 0.555 eV for $E_c - E_A$; 0.61 eV for $E_A - E_v$; 0.75 to 0.83 eV for $E_c - E_D$ and 0.345 eV for $E_D - E_v$. The sum of the two threshold energies of one of the levels should equal the silicon energy band gap of 1.166 eV at 90°K (Bludau et al., 1975), as illustrated in figure 4.14.

The sum $(E_c - E_A) + (E_A - E_v) = E_g$ is 1.165, which is very close to the expected value at 90°K. The sum $(E_c - E_D) + (E_D - E_v)$ ranges from 1.095 to 1.175 eV because of the uncertainty in $E_c - E_D$. However, if $E_c - E_v$ ($= E_g$) is taken as 1.166 eV and $E_D - E_v$, for which only one value of 0.345 eV is given, the resultant $E_c - E_D$ is 0.821, which is fairly close to the value predicted by Wong and Penchina, (0.84 eV - constant as temperature varies).

4.53.4 Reanalysis and comparison of energy level measurements

Three problems arising from energy level measurements in the literature have been highlighted. Firstly, many of the measurements have been incorrectly interpreted as being the actual gold energy levels at any temperature, whereas it has been shown that the result usually obtained is the energy level linearly extrapolated from the temperature of measurement to absolute zero. Secondly, a number of the results reported have not been analysed correctly; temperature dependences of various factors, particularly capture cross sections, have frequently not been considered. Thirdly, use of the results in predicting other energy level dependent properties has frequently been made without any consideration of the variation of the gold energy level positions with temperature.

The results of Engstrom and Grimmeiss (1974) and Newman (1954) - the latter reanalysed by Wong and Penchina (1975) - show that for temperatures up to 300°K, both deep gold energy levels are fixed to the conduction band edge.

In this section, a reanalysis of data from the literature sources already reviewed is presented. Reference to the equations presented in sections 4.52.1 and 4.52.2 are made. The linearly extrapolated absolute zero values of the gold energy levels (in general $\Delta E_o(T_m)$) will each be denoted thus:

$$(E_c - E_A)_o \; T_m \;, \quad (E_A - E_v)_o \; T_m \;, \quad (E_c - E_D)_o \; T_m \quad \text{and}$$

$$(E_D - E_v)_o \; T_m .$$

Data which had been presented as $\ln(F)$ against T^{-1} (equation 4.27) was reanalysed using equation 4.28A, with appropriate values of r and T_m, to obtain $\Delta E_o(T_m)$. Data which had been presented as $\ln(FT^{-r})$ against T^{-1} did not require reanalysis, the slopes obtained were reinterpreted as linearly extrapolated absolute zero values ($\Delta E_o(T_m)$). Data for which some of the temperature dependences had been corrected for and others had not (eg. Sah et al., 1969 and Tasch et al., 1970) was also reanalysed with equation 4.28A, where the factor r accounted for the temperature dependences not already included. The value of T_m, the effective temperature of measurement, was obtained from equation 4.45 using values of T_{max} and T_{min} from the original data.

The results of this reanalysis are displayed in table 4.3. The temperature dependences corrected for in the original publications and those corrected for here are listed in separate columns.

Hall coefficient measurements were corrected for the $T^{3/2}$ dependence of the density of states (see equation 4.20), the density of states in one of the bands being given by, (Nichols and Vernon, 1966):

Author	Type of Measurement	Gold Level Investigated	Temperature Dependences Corrected for	Additional Temperature Dependences Corrected here	T_m °K	Authors Result eV	$\Delta E_o(T_m)$.eV calculated here (see col.3 for ΔE)	Energy level position at 300°K assuming both levels fixed to E_c eV
Collins (1957)	Hall effect	$(E_c - E_A)_o T_m$	-	N_c	261	0.56	0.53	$E_c - E_A = 0.53$
	Hall effect	$(E_A - E_v)_o T_m$	-	N_v	261	0.66	0.62	$E_c - E_A = 0.53$
	Hall effect	$(E_D - E_v)_o T_m$	-	N_v	209	0.37	0.34	$E_c - E_D = 0.84$
	Resistivity	$(E_c - E_A)_o T_m$	-	N_c, μ_n	233	0.54	0.56	$E_c - E_A = 0.56$
	Resistivity	$(E_A - E_v)_o T_m$	-	N_v, μ_p	250	0.59	0.61	$E_c - E_A = 0.58$
	Resistivity	$(E_D - E_v)_o T_m$	-	N_v, μ_p	198	0.33	0.34	$E_c - E_D = 0.85$
Thurber et al. (1973)	Hall effect	$(E_D - E_v)_o T_m$	N_v	-	202	0.3609	-	$E_c - E_D = 0.8293$
	Hall effect	$(E_D - E_v)_o T_m$	N_v	-	221	0.3620	-	$E_c - E_D = 0.8310$
	Hall effect	$(E_D - E_v)_o T_m$	N_v	-	197	0.3613	-	$E_c - E_D = 0.8281$
	Hall effect	$(E_c - E_v)_o T_m$	N_c	-	274	0.5373	-	$E_c - E_A = 0.5373$
	Resistivity	$(E_D - E_v)_o T_m$	-	N_v, μ_p	171	0.3469	0.3587	$E_c - E_D = 0.8267$
	Resistivity	$(E_D - E_v)_o T_m$	-	N_v, μ_p	184	0.3438	0.3565	$E_c - E_D = 0.8310$
	Resistivity	$(E_D - E_v)_o T_m$	-	N_v, μ_p	171	0.3528	0.3646	$E_c - E_D = 0.8207$
Sah et al. (1969)	Emission rates	$(E_A - E_v)_o T_m$	N_v, V_{th}	T^{-2} for σ_p	263	0.5898	0.634	$E_c - E_A = 0.56$
		$(E_c - E_A)_o T_m$	N_c, V_{th}	-	-	0.547		$E_c - E_A = 0.547$
Tasch et al. (1970)	Emission rates	$(E_c - E_A)_o T_m$	N_c, V_{th} $\sigma_n = T^o$	-	-	0.5455	-	$E_c - E_A = 0.5455$
Engstrom et al. (1975)	Emission rates	$(E_c - E_A)_o T_m$	N_c, V_{th} $\sigma_n = T^o$	-	247	0.533	-	$E_c - E_A = 0.553$
		$(E_A - E_v)_o T_m$	N_v, V_{th} $\sigma_p = T^{-2.0}$	-	243	0.641	-	$E_c - E_A = 0.555$
Parrillo et al. (1972)	Emission rates	$(E_c - E_A)_o T_m$ $n = T^o$	N_c, V_{th}	-	329	0.49	-	$E_c - E_A = 0.49$
		$(E_A - E_v)_o T_m$	N_v, V_{th} $\sigma_p = T^{-4.0}$	-	332	0.72	-	$E_c - E_A = 0.488$
Pals (1974)	Capture rates	$(E_c - E_A)_o T_m$	N_c, V_{th}	-	-	0.56	-	$E_c - E_A = 0.56$

Table 4.3.

Re-analysis of Energy Level Measurements

$$N_{c \text{ or } v} = \frac{2 \left(m^*_{e \text{ or } h} \, kT \right)^{3/2}}{h^2} \qquad \ldots 4.53$$

The small temperature dependences of m_e^* or m_h^* (the electron and hole effective masses) were ignored (Barber, 1967).

Resistivity measurements were corrected for the temperature dependence of the density of states and also for the temperature dependence of mobility (see equation 4.21). Smith (1964) gives values for the temperature dependences of electron and hole conductivity mobilities of $T^{-2.5}$ and $T^{-2.3}$ respectively. The overall temperature dependences used for resistivity measurements (r in equation 4.28A) were $T^{-1.0}$ and $T^{-0.8}$ in n-type and p-type silicon respectively.

Emission probability measurements were corrected for the $T^{-3/2}$ temperature dependence of the density of states and the $T^{1/2}$ temperature dependence of carrier thermal velocity (see section 4.53.2). The temperature dependence of the cross section for electron capture at the gold acceptor(σ_{no}) was taken as T^{o} (Bemski, 1958; Engstrom and Grimmeiss, 1974 and others) and for hole capture at the gold acceptor (σ_p) was taken as $T^{-2.0}$. This latter value is due to Engstrom and Grimmeiss, 1974. The previously accepted value of T^{-4} due to Bemski (1958) was also tried but, as will be shown, the results obtained suggest that $T^{-2.0}$ is probably correct. The value given by Bemski for the temperature dependence of the capture cross section for holes at the gold donor of $T^{-2.5}$ was also used but not with any great confidence in the light of Engstrom et al's., disagreement with Bemski's other result.

Where results are given of both $(E_c - E_A)_o T_m$ and $(E_A - E_v)_o T_m$ for the gold acceptor, provided the value of T_m is roughly the same for both measurements, the sum of the two results should equal the linearly extrapolated band gap from T_m to absolute zero, $E_{go}(T_m)$ - see figure 4.8. If this result is indeed obtained, it is a fairly good indicator to the

Author	T_m. ^oK	$Eg_o(T_m)$	Sum of linearly extrapolated energy levels eV
Collins et al. (1957)	261	1.99	$(E_c - E_A)_o T_m + (E_A - E_v)_o T_m = 1.22$
Sah et al. (1969)	263	1.2	$(E_c - E_A)_o T_m + (E_A - E_v)_o T_m = 1.181$ for $\sigma_p \alpha T^{-2}$ $= 1.23$ for $\sigma_p \alpha T^{-4}$
Tasch et al. (1970)	239	1.196	$(E_c - E_A)_o T_m + (E_A - E_v)_o T_m = 1.174$ for $\sigma_p \alpha T^{-2}$ $= 1.22$ for $\sigma_p \alpha T^{-4}$
Engstrom et al. (1975)	245	1.197	$(E_c - E_A)_o T_m + (E_A - E_v)_o T_m = 1.194$ for $\sigma_p \alpha T^{-2}$
Parrillo et al. (1972)	330	1.21	$(E_c - E_A)_o T_m + (E_A - E_v)_o T_m = 1.21$ for $\sigma_p \alpha T^{-4}$ $= 1.153$ for $\sigma_p \alpha T^{-2}$

Table 4.4 Extrapolated absolute zero energy level values which should add up to $Eg_o(T_m)$

accuracy of the individual values of $(E_c - E_A)_o T_m$ and $(E_A - E_v)_o T_m$, :

$$(E_c - E_A)_o T_m + (E_A - E_v)_o T_m = (E_c - E_v)_o T_m = E_{go}(T_m) \quad \ldots 4.54$$

Table 4.4 lists the results for which this should occur. The values for $(E_A - E_v)_o T_m$ were recalculated from the results of Sah et al., (1969) and Tasch et al., (1970) using both of the proposed temperature dependences for σ_p , viz., $T^{-2.0}$ (Engstrom and Grimmeiss) and $T^{-4.0}$ (Bemski). Clearly if it is assumed that the measurements are accurate the value of $T^{-2.0}$ gives a sum which is closer to the expected value in both cases.

In Engstrom and Grimmeiss's own paper, they contrast the values obtained from equation 4.54 when applied to the results of previous authors. Consequently Sah et al's., and Tasch et al's., results apparently fail to satisfy equation 4.54 and disagree markedly with those of Engstrom and Grimmeiss. Had Engstrom and Grimmeiss first noted that the other results had not been corrected for the temperature dependence of the hole capture cross section, σ_p , and then corrected for it with their own proposed value of $T^{-2.0}$, the agreement would, as shown here, have been better than they indicated, (the sums for Sah et al's., uncorrected data was 1.136 eV and for Tasch et al's., was 1.133 eV).

The results obtained by Parrillo and Johnson are also shown in tables 4.3 and 4.4. With a temperature dependence of $T^{-4.0}$ for σ_p , the result for equation 4.54 with their values is close to the expected $E_{go}(T_m)$, however the individual values of $(E_c - E_A)_o T_m$ and $(E_A - E_v)_o T_m$ are quite different from other reported values. When the temperature dependence of σ_p proposed by Engstrom and Grimmeiss is used in a reanalysis of Parrillo and Johnson's data, the value obtained for $E_{go}(T_m)$ is incorrect.

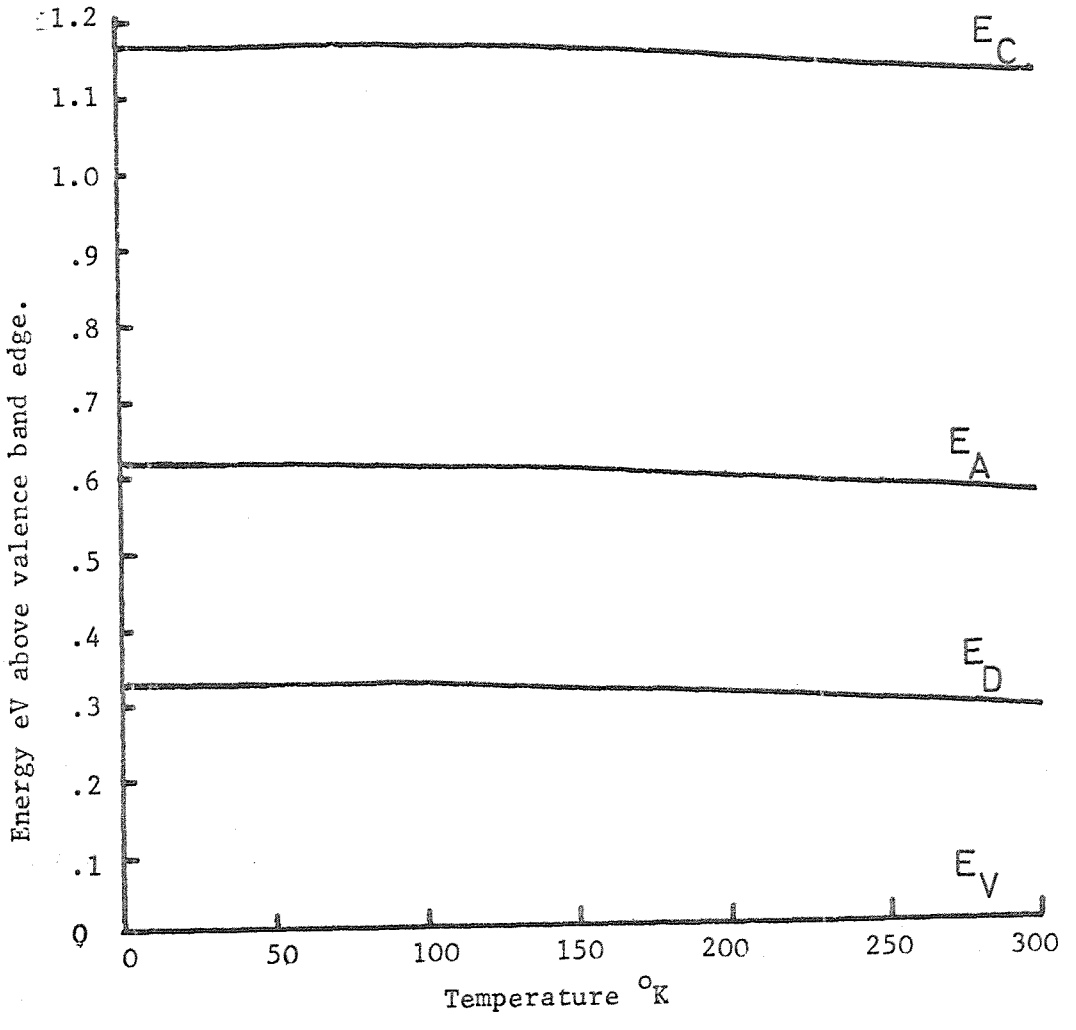


FIG 4.15: Probable temperature variation of the energy levels of gold in silicon.

4.54. Summary and Conclusion

A thorough reconsideration and, where appropriate, reanalysis of the large number of reported measurements of the gold deep level positions and their temperature dependences has been carried out. The generally accepted values of $E_c - E_A = 0.54$ eV and $E_D - E_v = 0.35$ eV at all temperatures have been shown to be very approximate only.

The results indicate that the gold deep acceptor energy level remains a constant distance from the conduction band edge as temperature varies. The value of $E_c - E_A = 0.553$ eV given by Engstrom and Grimmeiss is probably the most reliable. The evidence points to the gold donor energy level also remaining a fixed distance from the conduction band edge as temperature varies, with $E_c - E_D = 0.84 \pm 0.05$ eV. At 0^oK this gives a value of $E_D - E_v = 0.33$ eV and at 300^oK it gives a value of 0.284 eV which is very different from the previously accepted value of 0.35 eV.

These deep energy levels due to gold are plotted as a function of temperature between 0^oK and 300^oK in figure 4.15 using values of $E_g(T)$ taken from Bludau et al., (1974).

4.6 Degeneracies of Gold Energy Levels in Silicon

The concepts of degeneracy and degeneracy factors which were introduced and defined in section 4.2 may be extended to impurities, such as gold, which introduce more than one energy level into the silicon energy band gap. There is a degeneracy factor g_A associated with the gold acceptor level and a degeneracy factor g_D associated with the gold donor level.

Several different sets of degeneracy factors for the deep gold energy levels have been suggested in the literature. Some of these reported values arise from occupancy probability considerations and others from experimental evidence.

4.61 Degeneracy Values Reported in the Literature

Bullis (1966) and Bullis and Strieter (1968) discuss two models for gold in silicon and derive two different sets of degeneracy factors. There are some inconsistencies between the two models and between the two papers, which have given rise to confusion in many subsequent works. The two models proposed are:

i) Gold is monovalent in silicon. In the neutral state a substitutional gold atom, which would normally take part in four covalent bonds with the nearest neighbour silicon atoms, has only one valence electron. Loss of this electron results in a positive gold ion (ionized donor) and capture of another electron results in a negative gold ion (ionized acceptor).

ii) Gold is trivalent in silicon. In the neutral state, a substitutional gold atom has three valence electrons taking part in covalent bonds with silicon atoms. Loss of an electron and capture of an electron once again constitute donor and acceptor behaviour respectively.

A simple count of the number of ways that the electrons may distribute themselves amongst the available states for these models should give rise to the degeneracies of each charge state, and hence the degeneracy factors g_A and g_D. Bullis obtains degeneracy factors of $g_A = 2$ and $g_D = \frac{1}{2}$ for the trivalent model. These arise, by considering that an electron of either spin may leave to form a positive gold ion but that the captured electron forming a negative gold ion must be of the correct spin, in much the same way as simple donors and acceptors were considered in section 4.2. This consideration does not satisfy the definitions of degeneracy and degeneracy factor given in section 4.2. The ratios of the number of ways of

distributing electrons in the various empty and filled conditions should have been obtained, (see section 4.62). For the monovalent model however Bullis does count up the number of ways of distributing electrons amongst the various available states and obtains degeneracy factors of $g_A = \frac{2}{3}$ and $g_D = \frac{1}{4}$. These latter degeneracy factors were used almost exclusively in subsequent publications until Bruckner (1971) proposed a value of $g_D = \frac{1}{16}$ and Brown et al., (1975) used degeneracy factors obtained correctly for the trivalent model. Brown et al's., values of $g_A = 4$ and $g_D = 3/2$ will be derived in section 4.62. The inconsistency between the Bullis 1966 paper and the Bullis and Streiter 1968 paper is that from precisely the same considerations in each, the former gives values of $g_A = 2$ and $2/3$ for the two gold valency models whereas the latter gives $g_A = \frac{1}{2}$ and $\frac{3}{2}$. This reciprocation of the original values does not agree with the definition of degeneracy factor given in section 4.2 and is illustrative of the confusion that has arisen in the past. The reciprocated values are found in a number of other papers, notably Boltaks et al., (1960) and Parrillo and Johnson (1972) in which it is never quite clear which number (g_A or g_A^{-1}) is being used for the degeneracy factor.

The experimental values which have been reported for degeneracy factors all arise from measurements of gold energy levels described in section 4.53. These values, which are summarized in table 4.5, were obtained by substituting the experimental values of energy level positions back into the expressions describing the measured quantity (see equations 4.16, 4.18 and 4.19.) and using the degeneracy factor as an adjustable parameter. Since in all of these cases, except Engstrom and Grimmeiss (1975), the energy levels were incorrectly interpreted and no allowance was made for their temperature dependences, the value of degeneracy factor obtained includes not only the true degeneracy factor but also all of the temperature dependences and errors (see equation 4.10).

4.62 Degeneracies of the Gold Energy Levels from Theoretical Considerations

The degeneracies of each charge state may be obtained for the two models of substitutional gold bonding in silicon which have been mentioned

Table 4.5. Gold degeneracy factors suggested in the literature.

Bullis (1966) : $g_A = 2$, $g_D = \frac{1}{2}$ or $g_A = {}^2/3$, $g_D = {}^1/4$

Sah et al. (1969) : $g_A = 11.5$ or $g_A = 22.6$.

Sah et al. (1969) : from data of Fairfield and Gokhale (1965) $g_D = 51$.

Bruckner (1971) : $g_D = 16$.

Parrillo and Johnson (1972) : $g_A = {}^3/2$.

Thurber et al. (1973) : $g_D = 16$ or $g_D = 8$.

Engstrom et al. (1973): $g_A = 3.2$.

MONOVALENT				TRIVALENT			
excluding split-off band		including split-off band		excluding split-off band		including split-off band	
g_A	g_D	g_A	g_D	g_A	g_D	g_A	g_D
$^2/_3$	$^1/_4$	$^4/_3$	$^3/_4$	4	$^3/_2$	6	$^5/_2$

Table 4.6. Degeneracy factors for gold levels derived in this section

(monovalent and trivalent). If g^x, g^- and g^+ are the numbers of ways that electrons may be described in the neutral, negative and positive charge states respectively, the degeneracy factor for the gold acceptor is $g_A = g^x/g^-$ (ratio of no., ways empty to no., ways filled) and the degeneracy factor for the gold donor is $g_D = g^+/g^x$ (again the ratio of no., ways empty to no., ways filled).

In obtaining the three degeneracies, the way in which electrons are distributed amongst the available valence band states must be considered.

In a perfect silicon lattice, each silicon atom is covalently bonded to four nearest neighbour silicon atoms. Each band comprises two electrons of opposite spins which are shared between the two atoms concerned. A given atom, therefore, has eight electrons participating in the covalent bonds around it, four of which it has contributed itself and four of which are contributed, one from each, by the surrounding silicon atoms. A substitutional impurity replacing a silicon atom in the lattice is generally considered to participate in this tetrahedral bonding. The number of electrons provided to the bonds by the impurity atom depends, at any instant, on its valency and charge state. For example, a monovalent substitutional impurity will, in its neutral state, contribute one electron to the covalent bonds, it may gain up to three more electrons (acceptor action) or lose the one electron (donor action). Thus the monovalent impurity may be either singly donor ionized or singly acceptor ionized, doubly acceptor ionized or triply acceptor ionized. A similar consideration may be given to substitutional impurities with different valencies. The possibility of being more than singly ionized does not necessary imply that this state of affairs will occur. In the case of a group III shallow acceptor impurity, the single acceptor action is observed but the multi-charge donor action is not since a very large amount of energy would be needed to break the other bonds. Some deep level impurities however, do indeed offer more than one charge state, gold in silicon offers two

(positive and negative) and in germanium it offers four (single donor, single acceptor, double acceptor and triple acceptor). Clearly the arrangement of the electrons around the deep multi-level impurities is such that the energy required to break the bonds falls within the silicon energy band gap, (see section 4.63).

In order to predict the degeneracies of the various charge conditions of the substitutional impurity, the way in which the bonding and valence electrons are distributed amongst the silicon energy bands must be considered. Band structures of semiconductors have been the subject of many theoretical studies and silicon has received much attention with the result that the form of its energy bands is well known, (Phillips, 1973; Cohen et al., 1966; Kane, 1956, and many others).

A unit (primitive) cell of the silicon lattice contains eight valence electrons (4 covalent bonds) unless thermal excitation has removed one or more of them to the conduction band. In order to accommodate these electrons there are four valence bands which may each be occupied by two electrons of opposing spins, i.e. each band is doubly degenerate. There are two bands coincident at $k = 0$ with the highest energy. The next lowest band is split off from these by spin orbit coupling (Phillips, 1973, page 111) and lies below them by about 0.035 eV (McKelvey, 1966, page 299). The last band lies considerably lower in energy.

Teitler and Wallis (1960) presented a model for the occupancy of the valence bands in germanium. This model was used by Shtivel'man (1974) to calculate the degeneracies of the charge states of gold in germanium and extended by Brown (1976) to calculate the degeneracies of the charge states of trivalent gold in silicon. Brown used the Teitler and Wallis scheme because the valence bands of silicon are qualitatively similar to those of germanium. This model will be considered in a little more detail here.

The Teitler and Wallis model proposes that the substitutional impurity be imagined as a box which may contain as many as eight electrons

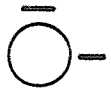
or as few as four depending on the valency and charge state of the impurity. It is assumed that the eight possible electrons are split up between the four bands. Two electrons paired in some sense are described in terms of the lowest band and are at the lowest energy of the eight electrons. Two electrons paired in some sense are described in terms of the next highest (split-off) band and are at a somewhat higher energy. There may be as many as four electrons or as few as no electrons described in terms of the two highest energy valence bands. Since these two highest bands are coincident and are degenerate at $k = 0$, it is assumed that all of the electrons present at these levels have the same energy. If the impurity in the box is a substitutional gold atom and it is monovalent, the number of electrons in the two highest bands when the gold atom is neutral will be one. When the gold atom is in the negative (ionized acceptor) charge state there will be two electrons and in the positive (ionized donor) charge state there will be none. If the gold atom is trivalent, there will be three electrons present in the two highest bands for the neutral state, four for the negative state and two for the positive state. The degeneracies of the charge states may be obtained from the number of ways of distributing the electrons between the two bands thus:

i) Monovalent substitutional gold:

The number of electrons in the two highest bands are represented schematically in figure 4.16:

FIG 4.16

Au^x, neutral, degeneracy g^x.



Au^-, negative, degeneracy g^-.

Au^+, positive, degeneracy g^+.

These electrons are distributed amongst four possible states (2 bands each with 2 spins). The number of ways of distributing one electron amongst the four states (the degeneracy of the neutral state) is:

$$g^x = \frac{4!}{1!\ (4-1)!} = 4 \qquad \ldots 4.55$$

since the number of ways of distributing n indistinguishable objects amongst r receptacles is given by:

$$\frac{n!}{r!\ (n-r)!} \qquad \ldots 4.56$$

Similarly, the degeneracies of the negative and positive charge states are:

$$g^- = \frac{4!}{2!\ (4-2)!} = 6 \qquad \ldots 4.57$$

and,

$$g^+ = \frac{4!}{0!\ (4-0)!} = 1 \qquad \ldots 4.58$$

The degeneracy factors for the two gold levels are obtained from the original definition (section 4.2). For the gold acceptor level:

$$g_A = \frac{g^x}{g^-} = \frac{4}{6} = \frac{2}{3} \qquad \ldots 4.59$$

and for the gold donor level:

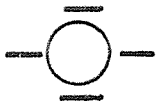
$$g_D = \frac{g^+}{g^x} = \frac{1}{4} \qquad \ldots 4.60$$

ii) Trivalent substitutional gold:

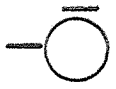
The number of electrons in the two highest bands are represented in figure 4.17:

FIG 4.17

Au^x, neutral, degeneracy g^x.



Au^-, negative, degeneracy g^-.



Au^+, positive, degeneracy g^+.

The degeneracies arising from the distribution of these electrons amongst the four states in the highest two bands are:

$$g^x = \frac{4!}{3!\,(4-3)!} = 4 \qquad \ldots 4.61$$

$$g^- = \frac{4!}{4!\,(4-4)!} = 1 \qquad \ldots 4.62$$

$$g^+ = \frac{4!}{2!\,(4-2)!} = 6 \qquad \ldots 4.63$$

which give rise to the degeneracy factors:

$$g_A = 4 \qquad \ldots 4.64$$

and $$g_D = \frac{3}{2} \qquad \ldots 4.65$$

The former set of degeneracy factors ($g_A = \frac{2}{3}$, $g_D = \frac{1}{4}$) is that proposed by Bullis for monovalent substitutional gold in silicon and the latter set is the correct version used by Brown et al., for trivalent gold in silicon ($g_A = 4$, $g_D = \frac{3}{2}$).

There has been no attempt in the literature to identify which set of degeneracy factors are correct. O'Shaughnessy et al., (1974) used a model of monovalent gold with its energy levels varying in proportion to the silicon energy band gap with temperature to explain measurements of the enhanced solubility of gold in heavily doped n-type silicon. Brown et al., (1975) used a model of trivalent gold with its energy levels both a constant distance from the conduction band edge as temperature varies to explain measurements of the enhanced solubility of gold in heavily doped p-type silicon. In this work (chapter 6) it will be shown that under the conditions of the two experiments (temperature, dopant concentrations etc.) neither model is distinguishable from the other. In chapter 7, calculations of resistivity in gold doped silicon at $300^{o}K$, using the corrected energy levels derived in section 4.53, various sets of values for g_A and g_D and each of the gold energy level temperature variation models, are presented. Slight differences between the two models mentioned above are shown by these calculations. The results tend to support the trivalent gold model (Brown et al.,) if the gold energy levels are fixed to the conduction band edge as temperature varies. This result is further supported by the conclusions drawn by Engstrom and Grimmeiss (1974) and Wong and Penchina (1975) that both gold energy levels do remain a constant distance from the conduction band edge for low temperatures ($< 300^{o}K$).

In addition to the two degeneracy models described above, a third model, proposed here, must be considered. The degeneracy factors already derived from the Teitler and Wallis model are obtained by distributing the electrons which take part in the electrical behaviour of the substitutional gold atom between the two highest energy valence bands. The four electrons occupying the other two valence bands are considered to be of sufficiently low energies to preclude their interaction. The split-off band in germanium is 0.28 eV below the two highest valence bands but in silicon it is only 0.035 eV below. Compared with the ionization energies of the gold levels relative to the two highest valence bands ($\approx$0.35 eV and $\approx$0.58 eV), the spin orbit splitting in silicon is very small and it is conceivable that the split-off band is involved in the ionization behaviour

of the gold atoms. If this is the case, the distribution of the electrons will be different. Once again there will be two electrons in the lowest valence band but now there may be as many as six or as few as two electrons distributed between the split-off and the two highest valence bands. For monovalent gold in silicon the degeneracies obtained if the split-off band has a probability of interaction (i.e. taking part in the ionization processes) equal to that of the two highest bands are: $g^x = 20$ (3 electrons amongst 6 states), $g^- = 15$ (4 electrons amongst 6 states) and $g^+ = 15$ (2 electrons amongst 6 states). The resultant degeneracy factors are $g_A = \frac{4}{3}$ and $g_D = 3/4$. For trivalent gold in silicon the same conditions result in $g_A = 6$ and $g_D = \frac{5}{2}$. The real situation, if the split-off band does indeed contribute to the degeneracies of the levels, would probably be even more complex as it is unlikely that the split-off band has a probability of interaction which is equal to that of the higher bands since it is slightly lower in energy. The examples above, however, do show that the possible interaction of the split-off band does modify the degeneracies considerably. If the split-off band is assigned some lower probability for interaction during ionization processes, degeneracy factors between those given immediately above and those obtained by ignoring the split-off band will be obtained.

Table 4.6 summarizes the degeneracy factors derived and discussed in this section. The degeneracy factors which include an effect due to the split-off band have not been considered before but will be compared with other reported values in the calculations presented in chapter 7, where it is shown that the trivalent model including the split-off band coupled with the gold energy levels fixed to the conduction band edge gives much better results than the monovalent model including the split-off band. It is not possible, however, to distinguish the trivalent degeneracies including the split-off band from those which exclude it.

Whether gold is monovalent or trivalent in silicon is open to some discussion and has not received a great deal of attention in the literature. From the basic arrangement of electrons around the gold atom, one might expect it to be monovalent since the outermost electron shell (the P-shell) contains one 6s electron only. The remaining unsatisfied

bonds after single acceptor ionization lead one to question why there are not more charge states apparent by analogy with gold in germanium which may be singly, doubly or triply acceptor ionised (Sze and Irvin, 1968). The monovalent bonding arrangement (figure 4.16) would certainly seem to lend itself to such behaviour. The trivalent bonding arrangement seems a little more consistent from this point of view, (although one could ask, in this case, why there are not more donor ionized states!). Consideration of how this bonding arrangement arises - if there is only one outer electron on the gold atom - tends to support the trivalent model. Haldane and Anderson (1976) modelled the multiple charge states which arise from transition metal impurities in semiconductors. They show that in order to give rise to closely spaced energy levels with different charge states, all of the electrons must occupy the same sub-shell of the metal atom - if not, the closeness of the ionization energies of for example the gold levels cannot be easily explained. In order for this to occur, the bonding arrangement is suggested to involve hybridization of the orbitals. This could occur for gold if, as suggested by Haldane et al., the bonding electrons which are derived from the d sub-shell of the O-shell, are promoted to form 's-p' hybrid orbitals. The occupation of the d-states varies as electrons are accepted or donated. This model could give rise to trivalent gold in silicon, but it must be stressed that this is speculative as no satisfactory theory has yet been published.

The strength of the model, as used here, lies in its ability to predict the solubility and conductivity properties of gold in silicon when coupled with the gold energy level temperature variations measured by Engstrom et al. (1974) and Newman (1954).

5. GENERATION-RECOMBINATION PROPERTIES OF GOLD IN SILICON

A detailed discussion of this topic is outside the intended scope of this thesis, however a brief review of the available literature is given below since the prime use of gold as a minority carrier lifetime killer is dependent on an understanding of the parameters of recombination processes at the gold energy levels. The basic Shockley-Read-Hall processes which are assumed to govern the behaviour of generation-recombination centres are described in chapter 4 and appendix D.

Gold acts as an efficient recombination centre in both n-type and p-type silicon due to its ability to behave as a deep donor or deep acceptor in silicon. When minority carrier lifetime killing is desired, gold is generally added in smaller quantities than the shallow level dopant concentration (otherwise the material would be compensated - see chapter 7). In n-type silicon, with the shallow donor concentration N_d considerably greater than the gold concentration N_{Au}, the Fermi-level will be close to the shallow levels and in equilibrium the gold atoms will be predominantly ionised as acceptors and therefore be negatively charged. This negatively charged centre has a large capture cross-section for minority carrier holes and, if they are present in excess, will capture them readily. As soon as the gold centre has trapped a minority carrier hole and become neutral, an electron drops from the conduction band to reionise it (for n-type material with $N_d \gg N_{Au}$ and low level injection of holes, the electron concentration, n, is high and this latter process will be extremely rapid - see appendix D equation D1). In this way recombination of the excess minority carrier holes occurs. The process in p-type material, where the shallow acceptor concentration $N_a \gg N_{Au}$, is similar. The gold levels are predominantly ionised as donors in equilibrium and present a large capture cross section for minority carrier electrons.

The parameter of interest to the device designer is the minority carrier lifetime τ. This is a measure of the mean time between generation and recombination of a minority carrier (Grove, 1967, p 120 ff). In cases involving recombination of injected excess minority carriers (for low level injection) the net recombination rate per unit volume is given by a simple relationship which states that it is proportional to the excess

minority carrier concentration. The lifetime τ, is related to the capture process by ;

$$\tau_n = \frac{1}{\sigma_n V_{th} N_t} \qquad 5.1$$

for electrons in p-type silicon, and

$$\tau_p = \frac{1}{\sigma_p V_{th} N_t} \qquad 5.2$$

for holes in n-type silicon. σ_n and σ_p are the capture cross-sections of the recombination centre for electrons and holes respectively, V_{th} is the thermal velocity of the carriers (see appendix D) and N_t is the density of recombination centres. For the case considered here, N_t is equal to the density of gold acceptors in n-type silicon and the density of gold donors in p-type silicon. In order to be able to predict the minority carrier lifetime in a given sample, the capture cross section and the recombination centre density need to be known. The thermal velocity, $V_{th} = \sqrt{3kT/m}$, where m is the carrier effective mass.

Measurements of minority carrier lifetime in gold doped structures have been carried out by many authors. Few, however, have attempted to relate them to the basic parameter of the capture process, viz. the capture cross section σ. (The four possible capture cross sections for gold in silicon have been described in chapter 4). The main problem arising from trying to establish the relationship between τ and σ is the measurement of the recombination centre density in the region in which the phenomena are taking place.

5.1 Measurements of capture cross sections

Values of the capture cross sections of the gold levels were obtained by Bemski (1958) from measurements of the lifetime of excess minority carriers in p-type and n-type gold doped silicon. Measurements were carried out over a range of temperatures to yield data for the temperature dependences of the capture cross sections for electrons and holes at the gold acceptor and of electrons at the gold donor. Bemski's data, which related τ to σ by equations 5.1 and 5.2, must be treated with caution

for two reasons. The first concerns the measurement of gold recombination centre densities. These were obtained by measuring the change in conductivity of uniformly doped n-type and p-type silicon doped with gold from the melt. The measurements of τ, however, were carried out in diode junction regions the fabrication of which could have had a dramatic effect on the gold distribution (see chapters 6 and 8) making inference of σ very dubious. It is also curious that Bemski reports a very close correlation between the gold concentration measured by the conductivity change with that measured by neutron activation. Many authors (e.g. Thurber et al. 1973) have noted consistent differences between total and electrically active gold concentrations. The second reason for suspecting the accuracy of Bemski's results is due to the fact that many of the measurements were made under conditions of high level injection of carriers. Under such conditions capture processes via both gold levels may occur rather than, as assumed, only via the gold level which would have been operative under low level injection conditions. Bemski's results, at $300^{o}K$, are as follows:

Hole capture cross section at the gold acceptor $= 1 \times 10^{-15}\ cm^2$, temperature dependence : T^{-4}.

Electron capture cross section at the gold acceptor $= 5 \times 10^{-16}\ cm^2$, temperature independent.

Electron capture cross section at the gold donor $= 3.5 \times 10^{-15} cm^2$, temperature dependence : $T^{-2.5}$.

Hole capture cross section at the gold donor $\geqslant 10^{-16} cm^2$ no temperature dependence measured.

Davis (1959) measured lifetimes in gold doped silicon at $77^{o}K$; these results are not particularly applicable to temperatures of interest and did not agree at all well with values measured by other authors.

Fairfield and Gokhale (1965) measured the four capture probabilities ($= \sigma x\ V_{th}$) by the photoconductivity decay method and the diode reverse recovery method. The former method gave rise to lifetimes obtained under low-level conditions which were related to the capture probabilities through equations 5.1 and 5.2. Gold recombination centre densities were measured by the change in conductivity, radiotracer experiments and by neutron activation;; caution must be exercised with the results. The

samples were prepared by diffusing gold into single crystals of silicon so large discrepancies in gold concentration (due to other dopants) are not expected and these results are certainly more reliable than those of Bemski. The diode recovery measurements were obtained on p^+n alloyed diodes fabricated on n-type gold doped silicon. Measured minority carrier lifetimes are plotted as a function of gold concentration although it is not stated whether the gold concentrations, which were obtained by diffusion, were corrected for the probable occurrence of the surface 'tip-up' in the gold concentration profiles (see chapter 3). Fairfield and Gokhale's data, multiplied by a carrier thermal velocity of $\sim 10^7$ atoms/sec at 300°K yields the following results:

Hole capture cross section at the gold acceptor = $1.15 \times 10^{-14} cm^2$
Electron capture cross section at the gold acceptor = $1.65 \times 10^{-16} cm^2$
Electron capture cross section at the gold donor = $6.3 \times 10^{-15} cm^2$
Hole capture cross section at the gold donor = $2.4 \times 10^{-15} cm^2$

Considering all of the possible sources of errors, it is surprising perhaps that, with the exception of hole capture cross section at the gold acceptor, those values agree so well with those of Bemski. These data are probably more reliable.

More recent publications have reported capture cross sections obtained from measurements of the capture and emission rates of the gold levels. The basis of the emission techniques is described in appendix F and equations 4.50 and 4.51, however the inference of capture cross sections and their temperature dependences requires some comment.

Equations 4.50 and 4.51 relate the capture cross section to the emission rates directly. One of these equations (4.51) is repeated below for ease of reference :

$$e_p = g_a^{-1} \sigma_p V_{th} N_v \exp\left[\frac{E_V - E_A}{kT}\right]$$

$$= AT^r \exp\left[\frac{E_V - E_A}{kT}\right] \qquad 5.3$$

As already described in chapter 4, the temperature dependences of all of the factors in this equation must be known (given by the exponent r) to obtain an activation energy for the gold acceptor level. In addition the temperature dependence of $E_V - E_A$ must be known in order to obtain E_A at any temperature. If the temperature dependences of σ_p and E_A are known, then a value of σ_p may be calculated if g_A is also known.

Engstrom and Grimmeiss (1975) considered the temperature dependence of the hole capture cross section at the gold acceptor level (the electron capture cross section being assumed to be temperature independent - see 4.53.2). Using equation 5.2, the Einstein relationship (Grove, 1967, p113) and the expression for hole diffusion length $L_p = \sqrt{D_p \tau_p}$, the hole capture cross section was expressed as :

$$\sigma_p = \frac{kT \mu_p}{qV_{th} N_t L_p^2} \qquad 5.4$$

The mobility of holes μ_p is caused by two scattering mechanisms, lattice scattering μ_L and ionised impurity scattering, μ_I, such that :

$$\frac{1}{\mu_p} = \frac{1}{\mu_L} + \frac{1}{\mu_I} \qquad 5.5$$

Using recently reported values of μ_L as a function of temperature and calculated values of μ_I (see section 7.3.2), Engstrom et al. obtained the temperature variation of μ_p in the range 150 - 250°K. Measurements of hole diffusion length as a function of temperature were made to give the final result (from equation 5.4) that :

$$\sigma_p = CT^{-2.0} \qquad 5.6$$

where C is a constant. This disagrees with Bemski's value of T^{-4}.

The hole capture cross section was then measured from equation 5.4 using a value for the gold acceptor concentration N_t. It was not stated how this concentration was measured. The value obtained for the capture cross section for holes at the gold acceptor was $1 \times 10^{-14} cm^2$ (assuming $V_{th} \sim 10^{-7}$ cm/s) at 300°K, which is very close to the value obtained by Fairfield and Gokhale (1965). Subsitution of this value in the emission rate equation (equation 5.3) yielded a value for the gold acceptor level degeneracy (assuming that $E_c - E_A$ was constant as temperature varied) of $g_A = 3.2$; this is close to

the trivalent bonding model value of $g_A = 4$. The value of g_A was used to calculate an electron capture cross section of the gold acceptor level of 2×10^{-15} cm^2. This value does not agree well with that of Fairfield et al (1965), however the uncertainties in the measurements of gold acceptor concentrations may well explain the differences.

Pals (1974) measured capture rates of electrons at the gold acceptor level and of holes at the gold donor level, both of which are expected to be temperature independent. The capture cross sections obtained were 1.3×10^{-16} for electrons at the gold acceptor level and 1.6×10^{-15} for holes at the gold donor level. These values agree well with those of Fairfield and Gokhale.

5.2 Concluding Comments

Although there is consensus within an order of magnitude for the gold level capture cross sections, many uncertainties suggest that further measurements would be valuable. Recently developed techniques (Herman, and Sah, 1972 ; Sah, 1976 ; Brotherton, 1976 and Pals, 1974) should allow measurements of the gold 'trap' density and the emission rates of electrons and holes to be obtained with much greater accuracy. The trap density (gold donor or gold acceptor) is obtainable by comparing gold-doped with non gold-doped control samples using the method of Herman et al.; the gold doped samples can then be used for emission rate measurements. The technique described by Brotherton allows the gold concentration to be obtained without the need for a control sample thereby eliminating possible errors. A combination of this measurement and an emission rate measurement, such as that described by Pals (1974) should enable the capture cross sections of majority carriers to be obtained (i.e. electrons at the gold acceptor in n-type material and holes at the gold donor in p-type material) although one of the techniques described by Sah et al.(1970) would have to be employed to obtain minority carrier capture cross sections ; this latter parameter being the most important for low level injection conditions.

6. THE SOLUBILITY AND GETTERING OF GOLD IN SILICON

6.1 Introduction

The solubility and gettering behaviour of gold in doped silicon are closely related phenomena. Heavily doped and crystalographically disordered areas of a silicon sample may have a dramatic effect on the local solubility limit of gold. This "solubility effect" and the gettering of gold into such areas may be explained by similar mechanisms.

In the literature, it has been reported that the solubility limit of gold at a given temperature is very much higher in heavily doped n-type silicon than in lightly doped or intrinsic silicon, (Cagnina, 1969; O'Shaughnessy et al., 1974; Chou et al, 1975; see 6.3). Dorward and Kirkaldy (1969) demonstrated that the gold solubility in heavily boron doped silicon was actually lower than in intrinsic silicon; however, recent experiments carried out by M. Brown (1976) and reported by Brown et al., (1975) show a small gold solubility enhancement in silicon doped with 9×10^{19} atoms/cc. of boron. Precipitation of gold has been observed by Seidel, Meek and Cullis (1975), in regions of silicon which have been crystalographically damaged by ion-implantation. Measurements of the gettering effects of heavily phosphorous doped and ion-implanted regions of silicon wafers have also been extensively reported, (Lambert and Reese, 1968; Nakamura et al., 1968; Meek et al, 1975; Seidel et al, 1975; Sigmon et al, 1976; Murarka, 1976; - see 6.5).

In this chapter, the mechanisms which are believed to cause the enhanced solubility of gold in silicon, and their relationship to the gettering effect, will be discussed. Both experimental and theoretical results which are described in the literature will be reassessed since, in many cases, errors and approximations have led to erroneous conclusions. A new model to elucidate the nature of the phosphorous gettering process is proposed in section 6.5.

6.2 Solubility enhancement through interaction of gold and other dopants

The equilibrium solubility of one species in another is generally

constrained by considerations of atomic size ratios, valency differences and crystal structures of the species (Van Vlack, 1970, pp 149-160). In a semiconductor, there are additional constraints due to the presence of other dopants. The effects which are to be considered here are the so-called "fermi-level solubility enhancement" effect, the ion-pairing effect and the formation of gold-impurity complexes. In addition some consideration is given to non-equilibrium effects which probably contribute to many of the results presented in the literature.

6.2.1 The fermi-level effect

The position of the Fermi-level in a silicon sample may have a dramatic effect on the equilibrium solubility of a charged species at a given temperature. This phenomenon has been treated in detail by Reiss et al, (1956), Shockley and Moll (1960) and Hall and Racette (1964).

For substitutional gold atoms diffused to saturation in intrinsic silicon at a given temperature, the total amount of gold will be made up of neutral, positively ionized and negatively ionized atoms :

$$N_{Au}^{TOT} = N_{Au}^{+} + N_{Au}^{x} + N_{Au}^{-} \qquad 6.1$$

in equilibrium, the ratio of charged to neutral gold atoms, neglecting degeneracy, is given by :

$$N_{Au}^{+} : N_{Au}^{x} : N_{Au}^{-} = \exp\left[\frac{E_D - E_F}{kT}\right] : 1 : \exp\left[\frac{E_F - E_A}{kT}\right] \qquad 6.2$$

where E_A and E_D are the energy levels of the gold acceptor and donor levels respectively and E_F is the position of the Fermi level. As shown in chapter 4, consideration of impurity level spin degeneracy affects the probability of occupancy of the various states such that equation 6.2 becomes :

$$N_{Au}^{+} : N_{Au}^{x} : N_{Au}^{-} = g^{+}\exp\left[\frac{E_D - E_F}{kT}\right] : g^{x} : g^{-}\exp\left[\frac{E_F - E_A}{kT}\right]$$

$$= g_D \exp\left[\frac{E_D - E_F}{kT}\right] : 1 : \frac{1}{g_A}\left[\exp\frac{E_F - E_A}{kT}\right] \qquad 6.4$$

where g^+, g^x and g^- are the degeneracies of the various charge states and g_D and g_A are the degeneracy factors defined in 4.2.

The equilibrium gold concentration, N_{Au}^{TOT}, becomes :

$$N_{Au}^{+} + N_{Au}^{x} + N_{Au}^{-} = N_{Au}^{x}\left[g_D \exp\left[\frac{E_D - E_F}{kT}\right] + 1 + \frac{1}{g_A}\exp\left[\frac{E_D - E_F}{kT}\right]\right] \qquad 6.5$$

Neglecting effects due to the silicon surfaces, in saturation the distribution of neutral atoms is uniform and in thermal equilibrium with gold on the exterior of the crystal. This is so since in equilibrium there is no driving force (such as a concentration gradient) on a neutral gold atom which can cause it to move. If the position of the Fermi level relative to the impurity energy levels varies for some reason (such as the introduction of a heavily doped region) the density of neutral gold, N_{Au}^{x} is unaffected ; however, in accordance with equation 6.4, the ratio of ionized gold to neutral gold changes since $(E_D - E_F)$ and $(E_F - E_A)$ change. Hence, N_{Au}^{TOT}, given by equation 6.5 will change.

In intrinsic silicon at a given temperature let $E_F = E_{Fi}$ and the substitutional gold concentration at the solubility limit be N_{Au}^{TOT} (i). Similarly, in extrinsic silicon at the same temperature, let $E_F = E_{Fex}$ and the subsitutional gold concentration at the solubility limit be N_{Au}^{TOT} (ex). The total change in solubility, or the solubility "enhancement" due to the Fermi level shift from E_{Fi} to E_{Fex} may be defined as the ratio:

$$\frac{N_{Au}^{TOT}\ (ex)}{N_{Au}^{TOT}\ (i)}$$

using equation 6.5 :

$$\frac{N_{Au}^{TOT}\ (ex)}{N_{Au}^{TOT}\ (i)} = \left[\begin{array}{c}\text{substitutional}\\ \text{gold solubility}\\ \text{enhancement}\end{array}\right] = \frac{N_{au}^{x}\left[1 + g_D \exp\left[\frac{E_D - E_{Fex}}{kT}\right] + \frac{1}{g_A}\exp\left[\frac{E_{Fex} - E_A}{kT}\right]\right]}{N_{au}^{x}\left[1 + g_D \exp\left[\frac{E_D - E_{Fi}}{kT}\right] + \frac{1}{g_A}\exp\left[\frac{E_{Fi} - E_A}{kT}\right]\right]} \qquad 6.6$$

The solubility enhancement can be calculated from this expression if the positions of the gold energy levels and their degeneracies are accurately known and if the values of E_{Fi} and E_{Fex}, at the diffusion temperature, are also known.

A similar Fermi level solubility effect may occur for interstitial gold atoms in the silicon lattice if they behave as a shallow donor impurity (this possibility has been discussed in chapter 4). In this case, the ionized interstitial donor concentration is : (from equation 4.3) :

$$N_{int}^{+} = \frac{N_{int}^{TOT}}{1 + \frac{1}{g_i} \exp\left[\frac{E_F - E_{Di}}{kT}\right]} = \frac{N_{int}^{TOT}}{F} \qquad 6.7$$

Now,

$$N_{int}^{TOT} = N_{int}^{+} + N_{int}^{x} \qquad 6.8$$

where N_{int}^{x} is the fraction of neutral interstitial gold (unionized donors), hence :

$$N_{int}^{TOT} = \frac{N_{int}^{x}}{1 - \frac{1}{F}} \qquad 6.9$$

The interstitial gold solubility enhancement in extrinsic (heavily doped) silicon is, therefore :

$$\frac{N_{int}^{TOT}\ (ex)}{N_{int}^{TOT}\ (i)} = \begin{array}{c}\text{Interstitial}\\ \text{gold solubility}\\ \text{enhancement}\end{array} = \frac{N_{int}^{x}\left[1 - \frac{1}{F_{ex}}\right]^{-1}}{N_{int}^{x}\left[1 - \frac{1}{F_i}\right]^{-1}} \qquad 6.10$$

where F is given by :

$$F = 1 + \frac{1}{g_i} \exp\left[\frac{E_F - E_{Di}}{kT}\right] \qquad 6.11$$

If E_{Di} is a shallow donor level close to the silicon conduction band edge, the interstitial gold solubility enhancement may be considerable in heavily doped p-type silicon because E_{Fex} is close to the valence band

edge. This possibility has not been considered by Brown et al, (1975) although it could be the dominant mechanism in their experiments on boron doped silicon.

Obviously, the Fermi level solubility enhancement effect only occurs if E_{Fex} moves appreciably from its intrinsic position, E_{Fi}, at the temperature of interest. At temperatures of 1000°C and above, the intrinsic carrier concentration, n_i, is quite high ($> 10^{18}$ carriers/cc); any extrinsic doping will have to be to a higher concentration than n_i before the Fermi level will move from the intrinsic position. Thus at normal diffusion temperatures (> 900°C) solubility enhancement of gold would only be expected in silicon which has been very heavily doped with a shallow level species. Under such conditions, there are considerable difficulties in evaluating equations 6.5 and 6.11. Not only are there uncertainties in the positions of the deep gold energy levels but there are also complications in evaluating E_{Fex}. This problem is considered in detail in section 6.4.

6.2.2 Ion-pairing of gold with other dopants in silicon

The ion-pairing phenomenon may occur when, at the diffusion temperature, two impurities of opposite charge are attracted together by forces, generally coulombic, so as to minimise the free energy of the system. Once paired, the two ions are considered to take little further part in diffusion processes since the ion-pair is much less mobile than the separate ionized atoms.

Reiss et al, (1956) consider ion-pairing in terms of a reaction :

$$D^+ + A^- \rightleftharpoons (DA) \qquad 6.12$$

where D^+ is an ionized donor, A^- is an ionized acceptor and DA is the neutral ion-pair. Application of the law of mass-action for a fixed number of ionized donors and acceptors to this reaction gives :

$$N_{DA} = K \quad (N_D^+ - N_{DA}) \; (N_A^- - N_{DA}) \qquad 6.13$$

(see: Reiss et al.)

where K is the pairing constant, which is independent of the doping concentrations but dependent on temperature.

To illustrate the way in which the ion-pairing effect leads to a solubility enhancement the specific case of silicon, heavily doped with shallow donors (e.g. phosphorus) will be considered. The expected pairing reaction is between negatively charged gold acceptor ions, N_{Au}^{-}, and positively charged shallow donor ions N_D^{+}. It has been shown by a number of authors, notably Cagnina (1969) and O'Shaughnessy et al, (1974), that at the gold solubility limit in extrinsic, n-type silicon, $N_{Au}^{TOT} >> N_D$ despite the Fermi-level and ion-pairing effects. It follows that the number of ion-pairs, N_{pairs}, is very much smaller than the number of shallow donors: $N_{pairs} << N_D$. Similarly $N_{pairs} << N_D^{+}$ since in this situation most or all of the shallow dopant atoms are ionized (see section 6.4). N_D^{+}, in equation 6.13, is therefore unaffected by the formation of ion-pairs. Consequently, the electron concentration, n (which in this heavily doped example is $>> n_i$) is also unaffected and the position of the Fermi-level, E_{Fex} is not altered. The density of ionized gold acceptors, N_{au}^{-}, is determined by the position of E_{Fex} and is therefore maintained at the value predicted by solution of equation 6.5. As a result, the formation of ion-pairs increases the gold solubility such that :

$$N_{Au}^{TOT}(ex) = \left[N_{Au}^{+} + N_{Au}^{x} + N_{Au}^{-}\right] + N_{pairs} \qquad 6.14$$

$$= \left[\begin{array}{c}\text{Density determined}\\ \text{by}\\ E_{Fex}\text{ only}\end{array}\right] + \left[\begin{array}{c}\text{Density determined}\\ \text{by}\\ \text{equation 6.13}\end{array}\right]$$

The pairing reaction between substitutional gold ions and shallow donor ions may be described using equation 6.12 thus :

$$N_{pairs} = KN_{Au}^{-} \cdot N_D^{+} \qquad 6.15$$

Since $N_D^{+} >> N_{pairs}$, a given value of N_D^{+} defines N_{Au}^{-} (through fixing the electron concentration n and hence E_{Fex}) and therefore N_{pairs}.

At a given shallow donor density (N_D) the Fermi-level (E_{Fex}) may be determined (see section 6.4) and the density N^-_{Au} inferred from equation 6.5 :

$$N^-_{Au} = N^x_{Au} \cdot \frac{1}{g_A} \exp\left[\frac{E_{Fex} - E_A}{kT}\right] \qquad 6.16$$

N^x_{Au} , as explained in 6.2.1, depends on temperature not E_F, at the solubility limit, and is the same in intrinsic and in extrinsic silicon. From equation 6.1, therefore :

$$N^x_{Au} = N^{TOT}_{Au}\ (i) - N^+_{Au}\ (i) - N^-_{Au}\ (i) \qquad 6.17$$

where (i) denotes the densities of the species in intrinsic silicon at the temperature of interest.

Using the Shockley-Last relationship expressed in equation 6.2 :

$$N^x_{Au} = N^{TOT}_{Au}\ (i) - N^x_{Au}\ g_D \exp\left[\frac{E_D - E_{Fi}}{kT}\right] - N_{Au}{}^x\ \frac{1}{g_A} \exp\left[\frac{E_{Fi} - E_A}{kT}\right] \qquad 6.18$$

Substitution of 6.18 and 6.16 into 6.15 yields :

$$N_{pairs} = \frac{N^{TOT}_{Au}\ (i).\ K.\ N^+_D\ \frac{1}{g_A}\ \exp\left[\frac{E_{Fex} - E_A}{kT}\right]}{1 + g_D \exp\left[\frac{E_D - E_{Fi}}{kT}\right] + \frac{1}{g_A} \exp\left[\frac{E_{Fi} - E_A}{kT}\right]} \qquad 6.19$$

The enhanced solubility due to both the ion-pairing and Fermi-level effects may be written from equation 6.14 :

$$N^{TOT}_{Au}\ (ex) = N^{TOT}_{Au}\ (i) \cdot \left[\text{(Equation 6.6)} + \left(\frac{N_{pairs}}{N^{TOT}_{Au}\ (i)}\right)\right] \qquad 6.20$$

which, using equation 6.19, gives :

$$\frac{N^{TOT}_{Au}\ (ex)}{N^{TOT}_{Au}\ (i)} = \left[\text{(Equation 6.6)} + \frac{K.\ N^+_D\ \frac{1}{g_A}\ \exp\left[\frac{E_{Fex} - E_A}{kT}\right]}{1 + g_D \exp\left[\frac{E_D - E_{Fi}}{kT}\right] + \frac{1}{g_A} \exp\left[\frac{E_{Fi} - E_A}{kT}\right]}\right] \qquad 6.21$$

The magnitude of the ion-pairing effect is very dependent on the value of K, the pairing constant between the two species in question. Evaluation of K is considered in section 6.2.2.1.

In gold-doped silicon, enhanced solubility is generally observed only in regions of very heavy shallow-level doping. For example, the very small enhancement of gold solubility observed in p-type silicon by Brown et al, (1975), was for a boron concentration of 9×10^{19} atoms/cc. In phosphorous and arsenic doped silicon the enhanced solubility effects noted by a number of authors were for shallow dopant concentrations in excess of 10^{19} atoms/cc. At phosphorous concentrations greater than 5×10^{19} atoms/cc, the gold solubility enhancement becomes very large indeed - much larger than predicted by the Fermi-level effect alone. It is to explain results in this concentration range that the ion-pairing reaction has been employed.

Joshi and Dash (1966), Lambert and Reese (1968), Cagnina (1969) and Chon and Gibbons (1975) have all used expressions derived from the Reiss et al (1956) theory, to predict the magnitude of the ion-pairing effect on gold in phosphorous or arsenic doped silicon. The results of Reiss et al, are not directly applicable to the examples in question. Although the mass action principle, which was used in their paper, is certainly applicable, the expression for the number of ion-pairs which was derived by Reiss et al., assumed that non-degenerate conditions prevailed in the semiconductor. That is to say, the majority carrier concentration was considered to be much smaller than the density of available states in the majority carrier band, ($n \ll N_c$ or $p \ll N_v$). This allowed certain well-known approximations to be implicit in the equation for N_{pairs}. In silicon, in the temperature range of interest (900 - 1300°C), the density of available quantum states is about 5×10^{19}/cc. If the doping level of phosphorous (which is the most common example) is greater than this value, the silicon can no longer be considered non-degenerate and the expressions obtained by Reiss et al., do not apply. Since enhanced solubility of gold is observed only in degenerate silicon, use of the Reiss et al., expressions by the authors cited above is not strictly correct. Equation 6.19 was derived without making non-degenerate approximations. As long as K and

E_{Fex} are known for the case in question, the number of ion-pairs may be calculated. Unfortunately, E_{Fex} is influenced by a number of ill-understood phenomena which occur at heavy doping levels ; however, calculations which are certainly more realistic than those available in the literature will be presented in section 6.4.

The Reiss et al. expression for n-type silicon doped with gold is briefly derived below in order to illustrate the way in which N_{pairs} is affected by the non-degenerate assumptions. The law of mass action is again used to describe the pairing reaction (see equation 6.15) :

$$N_{pairs} = K \left[N_{Au}^{-}(ex)\right] \quad N_D^{+} \qquad 6.22$$

where $N_{Au}^{-}(ex)$ is the concentration of negatively charged gold ions in extrinsic silicon. Under non-degenerate conditions the following equations are applicable (Grove 1967 p.99 ff)

$$n = n_i \exp\left[\frac{E_{Fex} - E_{Fi}}{kT}\right] \qquad 6.23$$

$$pn = n_i^2 \qquad 6.24$$

Under intrinsic conditions ($E_F = E_{Fi}$), equation 6.5 gives :

$$N_{Au}^{x} : N_{Au}(i) = 1 : \frac{1}{gA} \cdot \exp\left[\frac{E_{Fi} - E_A}{kT}\right] \qquad 6.25$$

and under extrinsic conditions ($E_F = E_{Fex}$) :

$$N_{Au}^{x} : N_{Au}^{-}(ex) = 1 : \frac{1}{gA} \exp\left[\frac{E_{Fex} - E_A}{kT}\right] \qquad 6.26$$

Thus :

$$\frac{N_{Au}^{-}(ex)}{N_{Au}^{-}(i)} = \exp\left[\frac{E_{Fex} - E_{Fi}}{kT}\right] \qquad 6.27$$

Substituting for the right hand side of this equation from equation 6.23 :

$$N_{Au}^{-}(ex) = \frac{\left[N_{Au}^{-}(i)\right]}{n_i} \cdot n \qquad 6.28$$

Therefore the pairing reaction becomes :

$$N_{pairs} = \frac{K N_D^+ \left[N_{Au}^- (i) \right] n}{n_i} \qquad 6.29$$

Since $N_D^{>>} N_{Au}$, the condition of charge balance is :

$$n = N_D^+ + p \qquad 6.30$$

From equation 6.24

$$n^2 = n N_D^+ + n_i^2 \qquad 6.31$$

and:

$$n = \frac{N_D^+ + \sqrt{(N_D^+)^2 + 4n_i^2}}{2} \qquad 6.32$$

The number of ion-pairs may now be written :

$$N_{pairs} = KN_D^+ . \left[N_{Au}^- (i) \right] . \left[\frac{N_D^+ + \sqrt{(N_D^+)^2 + 4n_i^2}}{2n_i} \right] \qquad 6.33$$

This, or a similar expression, was used by Joshi and Dash (1966), Lambert and Reese (1968), Cagnina (1969) and Chon and Gibbons (1975).

Comparison of equations 6.33 and 6.19 indicate that, if non-degenerate assumptions are incorrectly made, the number of ion-pairs increases in the manner :

$$N_{pairs} \alpha (N_D^+)^2 \qquad 6.34$$

since $N_D^+ \gg N_{Au}^- (i)$. If the approximations are not made :

$$N_{pairs} \alpha N_D^+ \exp \left[\frac{E_{Fex} - E_A}{kT} \right] \qquad 6.35$$

A comparison of the value of N_{pairs} obtained by each of these methods is given in section 6.4.

The possibility of ion-pairing between interstitial gold and substitutional acceptors should also be considered if interstitial gold does indeed behave as a shallow donor at the diffusion temperature. In the next section, however, it is shown that the ion-pairing constant, K, for this case is likely to be very small. Consequently the Fermi-level effect on interstitial gold donors in heavily doped p-type silicon is expected to be the dominant reaction.

6.2.2.1 The ion-pairing constant

The ion-pairing constant, K, in equation 6.13, has been considered theoretically in two ways. In their initial formulation of the ion-pairing effect, Reiss et al, (1956) modelled the pairing of ionized acceptor and ionized donor atoms by considering the host lattice (in this case silicon) to be a dielectric continuum, or fluid, in which the ions interacted solely through coulombic forces. The pairing constant was derived by evaluating the probability that a positive ion was situated within a sphere of radius 'a' centred at a given negative ion, where 'a' was the nearest neighbour distance of atoms in the lattice. Although the model seems to be a simple one, Wiley (1971) has pointed out that difficulties arise in solving the integral equation which results from it. In addition, ignoring the discrete nature of the host lattice means that a most important quantity required for evaluating ion-pairing effects - the concentration of nearest neighbour pairs in the lattice - is not included in the model.

Wiley (1971), using an approach based on work by Lidiard (1954), considers the way in which the interacting ions arrange themselves in order to minimise the free energy of the system. The approach is much more realistic and it will be used in a following section to calculate the ion-pairing constant. Derivation of the expression for the pairing constant is presented here for the case of gold acceptor - shallow donor pairs since Wiley's consideration, which is for zinc-oxygen pairs in gallium

phosphide, differs slightly.

A thermodynamic relationship for the free energy of the system is used (the Gibbs Free Energy equation) :

$$F = E - TS \qquad 6.36$$

where E is the internal energy of the system, F is the free energy available for reactions, T is the temperature and S is the configurational entropy or "disorder" of the system. TS is given by : (Dekker, 1952, p531)

$$TS = kT \ln W \qquad 6.37$$

where W is the number of distinguishable ways of constructing the system. The discrete nature of the silicon host lattice is introduced through W.

The internal energy, E, of the system is due to the formation of coulomb bonded nearest neighbour pairs (e.g. $\bar{Au}$- P^+ pairs) ; other interactions between unpaired ions are considered not to contribute to this energy if the number of pairs, P, is significant.

$$E = \frac{-Pe^2}{4\pi\varepsilon_o\varepsilon_{si}a} \qquad 6.38$$

where ε_{si} is the dielectric constant of the host (silicon) lattice and 'a' is the nearest-neighbour separation.

The entropy term may be obtained by evaluating W. If there are N lattice sites, N_D^+ ionised donors, N_A^- ionised acceptors, P donor-acceptor pairs and Z nearest-neighbour lattice-sites to a given lattice site, then W is given by

$$W = \left[\frac{Z^P \prod_{s=0}^{p-1} (N-2s)}{P!}\right] \times \left[\frac{(N-2P)!}{(N-N_D^+-N_A^-)!\,(N_D^+-P)!\,(N_A^- - P)!}\right] \qquad 6.39$$

It is assumed, in this equation, that unionised gold and phosphorus atoms "look like" silicon atoms if they are on a nearest neighbour site to

an ionised atom. The possible coulomb repulsion between ionised gold donors and ionised shallow donors on nearest neighbour sites is not included since, in the presence of a large shallow donor concentration, most gold atoms will be in the ionised acceptor state. The first term in equation 6.39 is the number of ways of forming the pairs and the second term is the number of ways of filling the remaining lattice sites.

Substituting 6.39 into 6.37 and 6.36, and setting $\frac{dF}{dP} = 0$ results in the minimum free energy condition.

Using Stirling's approximation :

$$\ln N! = N \ln N - N \qquad 6.40$$

and making the approximation that $N \gg S$ (typically $N = 10^{22}$ and $S = 10^{18}$), when $\frac{dF}{dP} = 0$.

$$\frac{P}{(N_D^{+}-P)(N_A^{-}-P)} = \frac{Z}{N} \exp\left(\frac{e^2}{4\pi\varepsilon_o\varepsilon_{si}akT}\right) \qquad 6.41$$

This equation may be rearranged to resemble the mass action relationship expressed in equation 6.13 :

$$P = N_{DA} = \frac{Z}{N} \exp\left(\frac{e^2}{4\pi\varepsilon_o\varepsilon_{si}akT}\right) (N_D^{+}-P)(N_A^{-}-P) \qquad 6.42$$

thus the pairing constant, K, is given by :

$$K = \frac{Z}{N} \exp\left(\frac{e^2}{4\pi\varepsilon_o\varepsilon_{si}akT}\right) \qquad 6.43$$

In silicon, Z, the number of "nearest neighbours" is equal to 4, (tetrahedral bonding).

Three effects other than coulomb binding may contribute to the energy of the ion-pair. First, the ions may not behave as point charges (equation 6.38). Inclusion of the possible multiple interactions that this implies would be very complex and would probably not affect the outcome greatly since the coulomb term is expected to dominate the difference between paired and unpaired nearest neighbours. Secondly, covalent bond formation may contribute to the problem by enhancing or inhibiting pair formation. This is unlikely to have any large effect however, since the classical picture

of donor atoms in silicon predicts that the bonding electrons assume the configuration appropriate to silicon, with the 'extra' electron in a bound state with a small binding energy (at a large radius) or in the conduction band (ionised). Because of this, covalent bonding is unlikely since there is no way for an atom to distinguish whether its nearest neighbour is an unionised phosphorus atom or a silicon atom. The third consideration is that of the ions interacting via their strain fields as well as via their coulombic fields. If one species has a much smaller ionic radius than the host lattice atoms and the other species has a much larger radius, the lattice strain would be minimised if the two ions occupied adjacent lattice sites. In the case of gold and phosphorus, this effect may be important since phosphorus atoms are smaller than silicon atoms and gold atoms are larger. The result would be an increase in the magnitude of K, the pairing constant.

A similar treatment may be given to the possibility of an ion-pairing reaction between a substitutional ion and an interstitial ion. Such a reaction might be of importance in gold diffused silicon which is heavily doped with shallow acceptors if interstitial gold atoms are shallow donors. In this case, the number of ways, W, of constructing the system will be different. If the acceptor ion is substitutional and the donor ion is the interstitial, there are Z_i equivalent positions that the donor may occupy to be paired with the substitutional acceptor (analogous to Z nearest neighbours in the substitutional-substitutional case). If there are N_i sites in the lattice which may be occupied by an interstitial ion, then :

$$W = \left[\frac{Z_i^{P} \prod_{S=0}^{P-1} (N-2s)}{P!} \right] \times \left[\frac{(N-P)!}{(N-N_A^-)!(N_A^- - P)!} \right] \times \left[\frac{(N_i - P)!}{(N_D^- - P)!(N_i - N_D)!} \right] \qquad 6.44$$

N and P have the same meanings as before. The first term is the number of ways of forming the pairs, the second term is the number of ways of filling the remaining lattice sites and the third term is the number of ways that the remaining donors may be placed in interstitial sites. Substituting 6.44 into 6.37 and 6.36, using 6.40 and setting $\frac{dF}{dP} = 0$, results in :

$$K_i = \frac{Z_i}{N_i} \exp\left(\frac{e^2}{4\pi\varepsilon_o\varepsilon_{si}akT}\right) \qquad 6.45$$

where K_i is the appropriate pairing constant. Meek and Seidel (1975) have considered the probability of this pairing reaction and conclude that Z_i, the number of equivalent "nearest neighbour" positions available to the interstitial donor when forming a pair with a substitutional acceptor, is likely to be the same as for the pairing reaction between two substitutional ions (i.e. $Z = 4$ in silicon) because of the bond symmetry. However, the number, N_i, of interstitial positions is likely to be very large indeed since the diamond lattice of silicon is very open. Andersen et al., (1971), have shown experimentally that $N_i \simeq 6 \times N$. Thus the value of K_i is probably small and pairing between an interstitial ion and a substitutional ion is likely to be a small effect - almost certainly masked by the Fermi-level effect. Calculated values of K and K_i and the resultant N_{pairs} are given in section 6.4.

6.2.3 Gold Precipitation and Impurity Complexing

The solubility enhancement effects described above result in gold remaining on substitutional lattice sites or residing in the lattice interstices. A number of workers, however, have reported precipitates and aggregates of gold in heavily doped areas or on crystal defects. Such effects can apparently contribute to the "enhancement" of the solubility of gold although they are usually indicative of an excessive rather than equilibrium amount of gold. The "apparent contribution" arises from the radio-tracer measurement technique normally employed in gold solubility experiments. Use of this method yields the total gold solubility (as a number of 'counts') but can not distinguish between electrically active, electrically inactive, substitutional, interstitial or precipitated gold. The more subtle method of Rutherford backscattering (Buck et al, 1972) may be used to identify the position of gold atoms relative to the silicon lattice matrix, although severe limitations are imposed by its sensitivity. A brief review of the precipitation effects reported in the literature is given below:

Wilcox, La Chapelle and Forbes (1964), in studies of gold in heavily doped layers, noted that the diffusion of gold through n^+ (phosphorus doped)

layers was retarded whilst the solubility after "equilibration" was increased. Formation of a gold phosphide (Au_2P_3) compound was suggested as a possible cause of these two effects.

Fairfield and Schwuttke (1966) made X-ray topography studies of device structures both before and after gold diffusions. Precipitates were observed in the n^+ phosphorus doped regions after gold diffusion. Studies as a function of depth indicated that these precipitate were confined to a region above the phosphorus junction. Beneath the junction a depletion of gold was seen when the results were compared to regions which had not been phosphorus diffused. The observed precipitate were suggested to be some kind of gold-phosphorus-silicon complex or precipitate. Joshi and Dash (1966), who worked for the same organisation as Fairfield and Schwuttke, observed similar precipitates in the T.E.M. Studies were made through gold doped, p-type, silicon wafers which were first phosphorus diffused on both sides to form shallow junctions. Precipitates were observed in the phosphorus diffused areas and at both of the wafer surfaces although the elemental gold source was only applied to one side of the wafer. The precipitates were suggested to be gold-phosphorus complexes.

Wolley and Stickler (1967) also studied and tried to identify precipitates in gold diffused wafers which were uniformly phosphorus or boron doped, or which had shallow n^+ phosphorus-diffused layers. An interesting feature of their work is that the gold source was not an elemental layer applied to the wafer surface. The diffusions were carried out by the sealed capsule technique using gold chloride ($AuCl_3$) as the source. Precipitates were once again observed in the T.E.M. In all of the samples gold was found to have precipitated in small quantities at or near dislocations, loop type defects and the silicon-silicon dioxide interface. In one phosphorus doped sample, a large precipitate was identified as gold phosphide (Au_2P_3) by comparing the electron diffraction pattern observed from the precipitates with that observed on gold phosphide samples prepared in the laboratory. The precipitates in other wafers were too small for diffraction patterns to be obtained. In the phosphorus diffused (gettered) wafers - where a solid P_2O_5 source of phosphorus was used - small precipitates were seen as well as decorated dislocation networks generated during the gettering step. It was suggested that most of the gold had precipitated onto the dislocations rather than

reacting directly with the phosphorus.

More recent studies by Meek, Seidel and Cullis (1975) and Cullis (1974), using T.E.M. and Rutherford backscattering techniques, have also yielded a little information about precipitates in gold-doped shallow diffused layers. The presence of precipitates in phosphorus diffused layers was indicated by the backscattered counts in the channelling direction. Large precipitates in boron diffused areas were identified as being "mainly gold" which was thought to have aggregated as a result of supersaturation during thecooling of the samples. Seidel, Meek and Cullis (1975) also observed gold precipitates in areas of ion-implanted damage in gold-doped silicon wafers. These results are considered later. Chou and Gibbons (1975) also carried out Rutherford backscattering studies of gold doping in phosphorus diffused layers. Backscattered counts in the channelling direction were, once again, attributed to the probable presence of precipitates.

From the lack of precise numbers it is not possible to determine whether gold precipitation contributes to solubility enhancement. However, in this work, the effect is not considered in detail since the more important features of gold solubility enhancement are shown to be explained without further reference to precipitation.

6.3 Gold Solubility Measurements from the Literature

Before making any assessment of the large number of results available in the literature, a distinction must be made between equilibrium and non-equilibrium conditions. Unfortunately, this has not always been done in the past and a number of the authors cited in 6.23 have only used the equilibrium theories of Fermi-level solubility enhancement and ion-pairing to explain their results of gold concentration in diffused layers. In general, gold solubility inferred from any of the experiments carried out on gold interaction with diffused and ion-implanted layers could be non-equilibrium and although the Fermi-level and ion-pairing effects may be linked with - and used to explain - the segregation of gold into such layers, the observed results may not safely be used as indications of the equilibrium solubility of gold. In cases where diffused layers have been saturated with gold, the gold concentration can

be much higher than the expected equilibrium solubility limit. This is probably due to excesses of other point and line defects which react with the gold atoms. Such phenomena are discussed in more detail in section 6.5 which deals with the segregation of gold into heavily doped areas.

In this section, the equilibrium gold solubility results are distinguished from the non-equilibrium ones. There are, in fact, very few measurements which may be taken as a true indication of the equilibrium solubility of gold in intrinsic or heavily doped silicon. Even these must be interpreted with some care since the results obtained may reflect poor experimental procedures. In addition, there are still unresolved problems relating to the ratio of interstitial to substitutional gold atoms, which ought to be considered in radio-tracer experiments which cannot distinguish between gold atoms of various charge states residing in various lattice locations. Amongst the few theoretical considerations of gold solubility phenomena there are inconsistencies and disagreements. In the section following this, (6.4), an attempt is made to explain how different theories can apparently each give the "right answer" for different experiments.

True "equilibrium" measurements of gold solubility have been attempted by the following authors : Collins, Carlson and Gallagher (1957); Struthers (1956 and 1957); Boltaks, Kulikov and Malkovich (1960); Sprokel and Fairfield (1965); Bullis and Streiter (1968) ; Dorward and Kirkaldy (1963 and 1969) ; Cagnina (1969); Yoshida and Saito (1970); Huntley and Willoughby (1973); O'Shaughnessey, Barber, Thompson and Heasell (1974); and Brown, Jones and Willoughby (1975). Measurements of saturated gold concentrations in shallow phosphorus and boron diffused layers have been presented by Wilcox, La Chapelle and Forbes (1964); Joshi and Dash (1966); Chou and Gibbons (1975); and Meek, Seidel and Cullis (1975). In addition there are approximately twenty papers concerning gold in diffused layers which are not listed here since the gold concentrations were not saturated - these results are reviewed in section 6.5.

The first list of papers will be considered here.

Since, as has been discussed in chapter 3, the diffusion of gold into thin wafers results in a U-shaped depth profile - indicative of some non-equilibrium effect at the surface - care must be exercised so that when

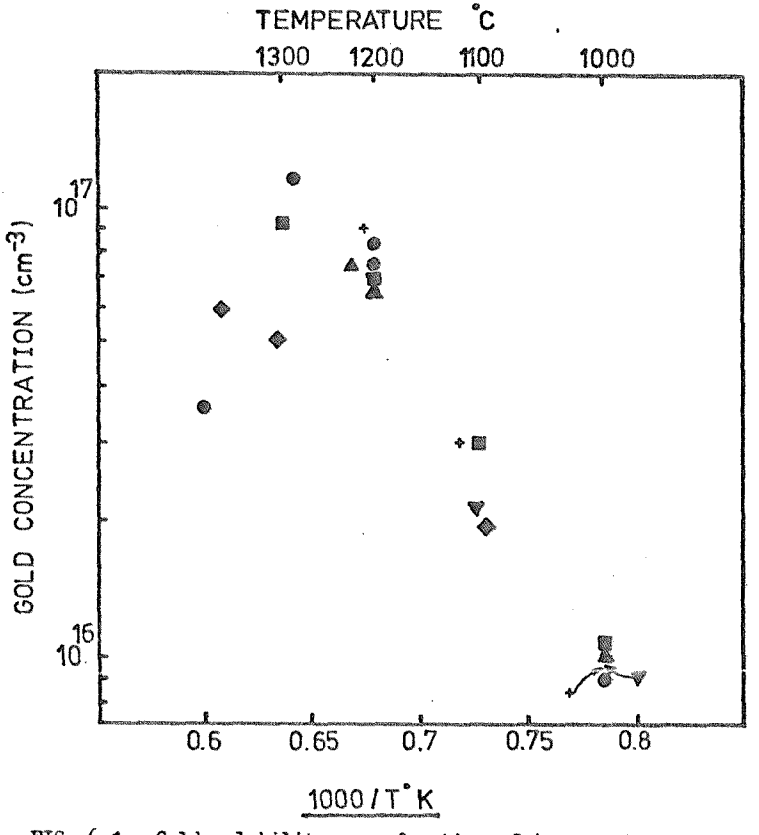


FIG. 6.1: Gold solubility as a function of inverse temperature in intrinsic silicon - questionable data:

●Collins et al., (1957) ▲Struthers (1956 & 1957)
◆Boltaks et al., (1960) ▼Sprokel et al., (1965)
■Dorward et al., (1968) + Yoshida et al., (1970).

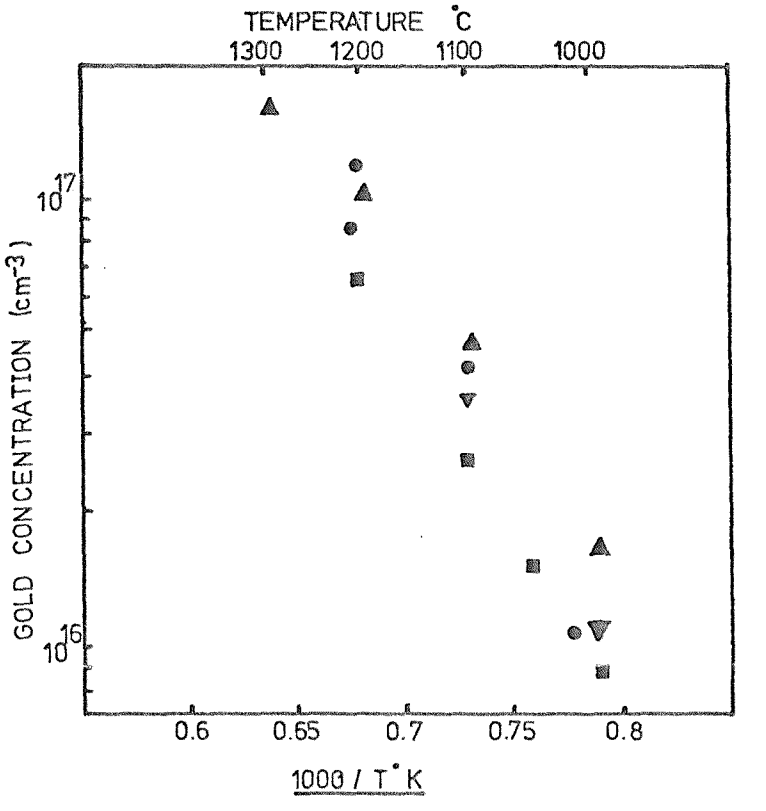


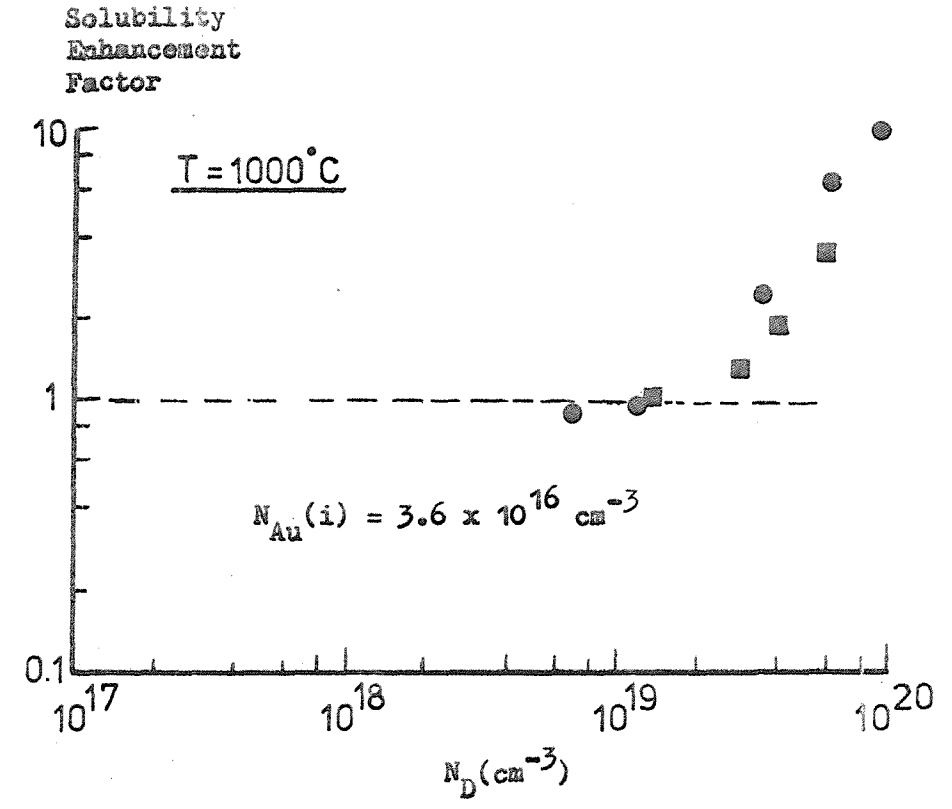
FIG. 6.2: Gold solubility as a function of inverse temperature in intrinsic silicon:

▼Cagnina (1969) ●Huntley et al., (1973
■O'Shaughnessey et al., (1974) ▲Brown et al., (1975).

carrying out gold solubility experiments by diffusion to saturation, the non-equilibrium areas near the surface are not included. In early measurements of gold solubility, it is very unlikely that this was done. The results by Struthers, Collins et al, and Boltaks et al., for intrinsic silicon, which are shown in figure 6.1, fall into this category and, therefore, must only be taken as indicative of the trend in gold solubility in intrinsic silicon. There is some doubt as to whether some authors (including those just cited) ensured that the samples had been diffused for long enough for equilibration to have occurred. Points in Figure 6.1 due to Sprokel and Fairfield (1965) and Yoshida and Saito (1970) fall into this category.

Data for gold solubility in intrinsic silicon which should be more reliable is shown in figure 6.2. These points, due to Cagnina (1969), O'Shaughnessey et al. (1974), and Brown et al. (1975), are to be discussed in detail since these authors have also carried out measurements of gold solubility in extrinsic silicon. It is reasonable to assume that these results reflect values of the solid solubility of the substitutional gold species in intrinsic silicon by inference from the results of Wilcox and La Chapelle (1964) which are discussed in chapter 3.

Adamic and McNamara (1964) showed that gold was more soluble in heavily doped n-type regions of device structures, but not in heavily doped p-type regions. Subsequently several authors presented results showing the higher solubility of gold in phosphorus diffused layers of silicon. Joshi and Dash (1966) observed higher gold concentrations in phosphorus diffused layers and invoked the "Fermi-level" and "ion-pairing" effects as probable causes. The first attempt to measure the actual increase in the equilibrium solubility of gold in low resistivity silicon was made by Bullis and Streiter (1968). Gold was diffused to saturation in arsenic doped wafers of 0.2 - 0.3Ωcm and 4 - 6 Ωcm. The diffused gold was then neutron activated and the total gold concentrations in the wafers were measured after removal of 10μm of silicon from each face. It is more than likely that removal of only 10μm of silicon was not enough to get rid of the surface tip-up regions and for this reason these results, which show a large scatter, can not be considered accurate. Bullis and Streiter did not observe any higher gold solubility in the more heavily arsenic doped wafers; however, at concentrations of $N_D = 3 \times 10^{16}$ atoms/c.c., (= 0.2Ωcm.), solubility



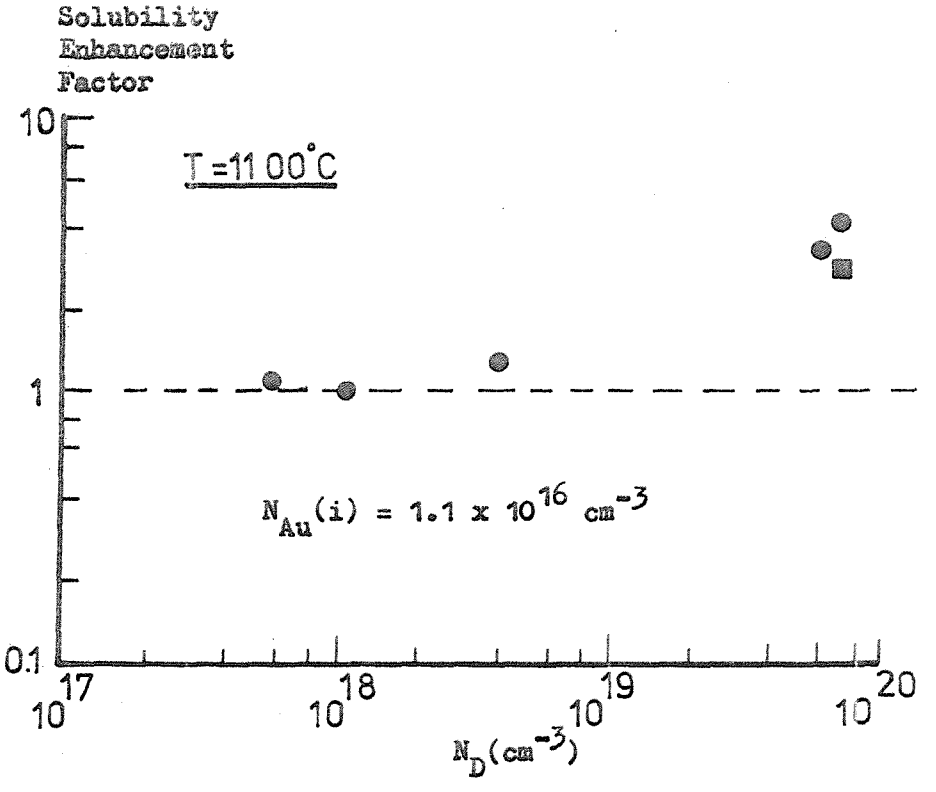


FIG. 6.3: Gold solubility enhancement as a function of shallow donor density. ● = Phosphorus, ■ = Arsenic. Cagnina (1969)

enhancement would not be expected because at the diffusion temperatures (1000°C) the samples would be intrinsic ($n_i > 10^{18}$ carriers/cc $\gg$ 10^{16} carriers/cc).

Cagnina (1969) made a more careful study of the solubility of gold by diffusion into lightly doped and heavily phosphorus and arsenic doped silicon wafers. In these experiments, after gold equilibration, 60μm of silicon was removed from each face of the wafers which were also subjected to boiling in a special solution to dissolve surface plated gold. The removal of 60μm of silicon is probably adequate since gold profiles presented in chapter 9 of this work indicate that the tip-up region is usually within 50μm of the surface (radio-tracer profiles obtained by Huntley and Willoughby - chapter 3 - show similar results). The dissolution of surface gold has proved to be a problem in radio-tracer studies of gold diffusion because a substantial amount of the gold removed into solution by chemically etching a layer of silicon can be plated back on to the wafer surface and cause a spuriously high gold concentration to be measured in the remainder of the wafer. The result of Cagnina's experiments at 1100°C and 1000°C are shown in figure 6.3. It is interesting to note that phosphorus and arsenic doping yield a different gold solubility.

Dorward and Kirkaldy (1968) measured the solubility of gold in intrinsic (2000Ωcm) silicon wafers. In the experimental details there is no evidence that the surface tip-up region was removed before the gold concentrations were measured by neutron activation. Thus, these results must be treated with some caution. The values for gold solubility in intrinsic silicon are also included in figure 6.1. A second paper by Dorward and Kirkaldy in 1969 reported a measurement of gold solubility at 1000°C in silicon doped with 1×10^{19} atoms/cc of boron. Comparison of this result with that obtained in intrinsic silicon indicated that the solubility had decreased by 40% as a result of the heavy doping. A rough calculation based on the Fermi-level effect was used to show that this result was not unreasonable - a fact disputed by Brown et al. (1975), who measured small gold solubility enhancements in heavily boron-doped silicon wafers (9×10^{19} atoms/cc boron) over a range of temperatures. It will be shown in the next

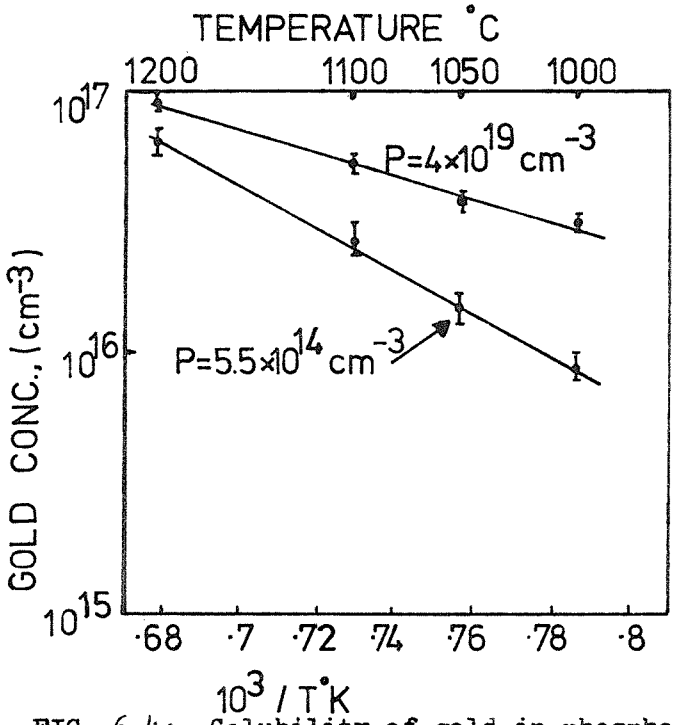


FIG. 6.4: Solubility of gold in phosphorus doped silicon
O'Shaughnessy et al., (1974)

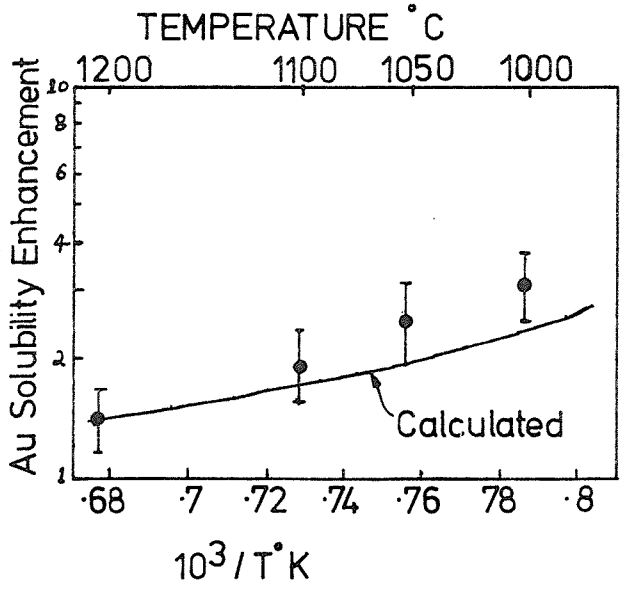


FIG. 6.5: Temperature dependence of gold solubility enhancement in silicon doped with 4 x 10^{19} phosphorus atoms/cc.
O'Shaughnessy et al., (1974).

section, that both solubility enhancement in silicon doped with 9×10^{19} atoms/cc of boron and solubility reduction in silicon doped with 1×10^{19} atoms/cc of boron are not inconsistent in terms of the Fermi-level effect, however the magnitude of Dorward and Kirkaldy's result is greater than would be expected and is probably due to incomplete equilibration. Brown et al. noted that the time taken for equilibration in heavily boron doped silicon was much longer than for equilibration in intrinsic silicon. A similar observation was made by Cagnina (1969) in his work on heavily doped n-type silicon.

O'Shaughnessey et al. (1974) and Brown et al. (1975 mentioned above) have carried out the most reliable measurements of gold solubility in heavily doped n-type and p-type silicon respectively. However, there are a number of aspects of both works which are open to some discussion. These will be dealt with here since they have a bearing on the theoretical results presented in the next section.

The results, obtained by O'Shaughnessey et al. of gold solubility in heavily doped silicon wafers, are shown in figures 6.4 and 6.5. The first of the diagrams illustrates the absolute values of solubility in lightly doped (5.5×10^{14} phosphorus atoms/cc) and heavily doped (4×10^{19} phosphorus atoms/cc) wafers, and the second diagram shows the solubility enhancement factor as defined in section 6.2.1. The line plotted in this latter diagram is the calculated value obtained by O'Shaughnessey et al. using a model of monovalent gold energy levels the positions of which vary in proportion to the variation of the silicon energy band-gap with temperature (see section 4.3.2). The effects of the temperature variation and degeneracies of the gold energy levels are discussed in the next section. The position of the Fermi-level in the heavily doped wafers at the diffusion temperature was calculated using a band model which made allowances for heavy doping effects such as impurity banding and band-edge taling - these effects are also considered briefly in the next section. It is worth noting that O'Shaughnessey et al. took care to avoid another cause of inaccuracy in gold solubility measurements - namely the loss of gold from the bulk of the silicon wafer during cooling from the equilibration temperature. As shown in figure 6.1, the gold solubility becomes lower as temperature becomes lower. This suggests that, since gold is a very fast

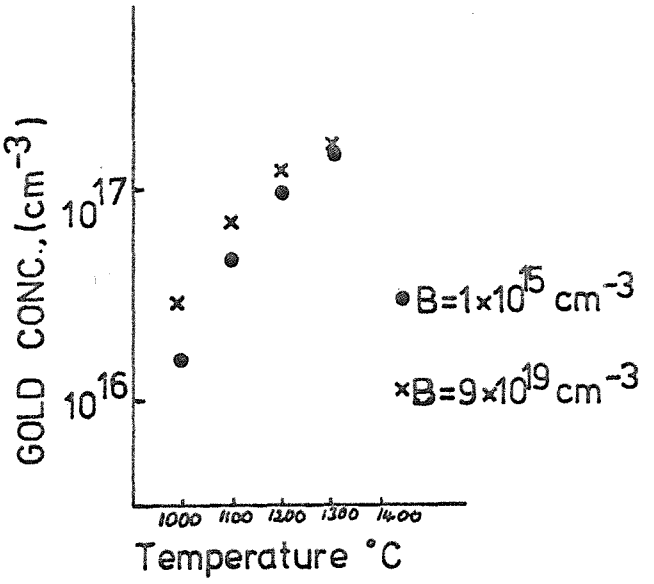


FIG. 6.6: Solubility of gold in boron doped silicon.
Brown et al., (1975).

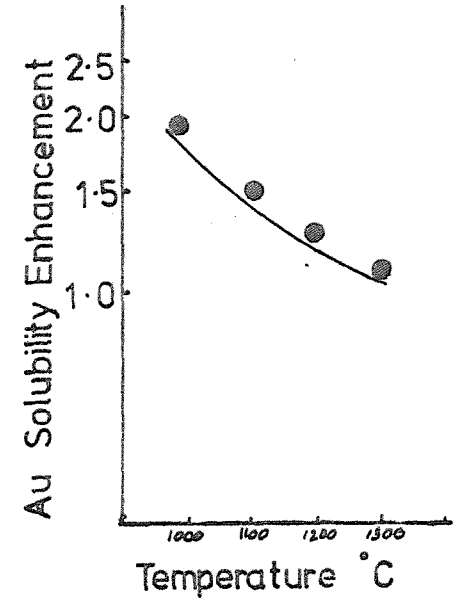


FIG. 6.7: Temperature dependence of gold solubility enhancement in silicon doped with 9 x 10^{19} boron atoms/cc.
Brown et al., (1975).

diffuser (see chapter 3), appreciable redistribution of gold between the bulk and surface of a thin silicon wafer may occur during cooling. To avoid this equilibration was carried out in a nitrogen ambient and the samples were then oxidised in-situ to form 1000 Å of silicon dioxide before removing them from the furnace. O'Shaughnessey et al. have shown in another paper (O'Shaughnessey, Barber and Heasell, 1976) that this has the effects

(a) of removing excess gold from the silicon surface and

(b) of preventing out-diffusion of gold from the silicon bulk.

Care was also taken to ensure that equilibration had indeed occurred by measuring the gold concentration as a function of diffusion time.

The main criticism of O'Shaughnessey et al's analysis of their results, is their rejection of the possible occurrence of ion-pairing. In both the data shown in figure 6.5 and in measurements of gold solubility enhancement in a shallow phosphorus diffused layer with a mean phosphorus concentration of 5×10^{20} atoms/cc, a difference between the predicted and the measured enhancement factor was noted. The measured values were consistently higher. Measurements of the effect of gold on the resistivity of silicon initially doped with 5×10^{14} and 4.5×10^{15} atoms/cc of phosphorus were shown to "agree" (although not very well!) with theoretical results of Thurber et al. (1973). This was taken as an indication that ion-pairing and compound formation could be ruled out ; however, the phosphorus concentrations in these samples were much too low for the ion-pairing effect to be significant (5 orders of magnitude too low!) and discrepancies in the work of Thurber et al. (which are pointed out in their paper) were ignored by O'Shaughnessey et al. (see chapter 7 for a detailed consideration of the effect of gold on the resistivity in silicon). In the next section it is shown that ion-pairing is, in fact, dominant in the result of the shallow phosphorus diffused layer and not insignificant at 4×10^{19} atoms/cc of phosphorus.

The results of Brown et al. (1975), for p-type silicon doped with 1×10^{15} and 9×10^{19} atoms/cc of boron are shown in figures 6.6 and 6.7. The solid line drawn through the results in figure 6.7 is taken from Brown et al's calculated values of solubility enhancement using a model of trivalent gold with energy levels remaining a fixed distance from the conducted band edge as

temperature varies. This degeneracy model, proposed by Brown et al., is the one favoured by this author in the light of a comparison of theoretical results presented in this and the next chapter. The Fermi-level in Brown et al's calculations was obtained graphically from values of the Fermi-integral (Blakemore, 1962) without making allowances for heavy doping effects, although this point is made in their paper. Once again very careful experimental procedures were used to ensure equilibration and the samples were quenched after annealing to prevent out-diffusion of gold. Measurements of the gold profiles through the wafers were made to avoid inclusion of the surface tip-up region in the final results. The theoretical solubility enhancement calculated by Brown et al., was obtained by assuming that the only interaction between gold and boron was via the Fermi-level effect. Ion-pairing was ruled out by a rough calculation based on Cagnina's values of the pairing constant between phosphorus and gold. In fact, the gold-boron pairing constant is larger than the gold phosphorus one (see section 6.4) so in this case Brown et al's result was fortuitous. The reason for gold-boron pairing being a small effect is shown in section 6.4 to be due to the small relative densities of positively ionised gold atoms to negatively charged boron atoms. The factor which was ignored but which could be dominant in the results is the Fermi-level solubility enhancement effect on interstitial gold atoms if they behave as shallow donors. The possibility of this has been discussed in section 4.4.2 where the strong evidence for positively charged interstitial gold obtained by Boltaks et al (1961) was mentioned. It is paradoxical that Brown et al. used this very possibility to explain why the rate of change of gold concentration was slower in heavily boron-doped silicon than in intrinsic silicon and yet ignored it in their solubility enhancement theory. Calculations presented in section 6.4 include this possible effect.

Comparison of O'Shaughnessey et al's, and Brown et al's results for intrinsic silicon (in figure 6.2) indicates that, despite the care taken to ensure total equilibration of the samples with gold, undersaturation in the former results may have occurred.

6.4 Calculations of Gold Solubility Enhancement

In this section, the models for gold solubility enhancement, which were

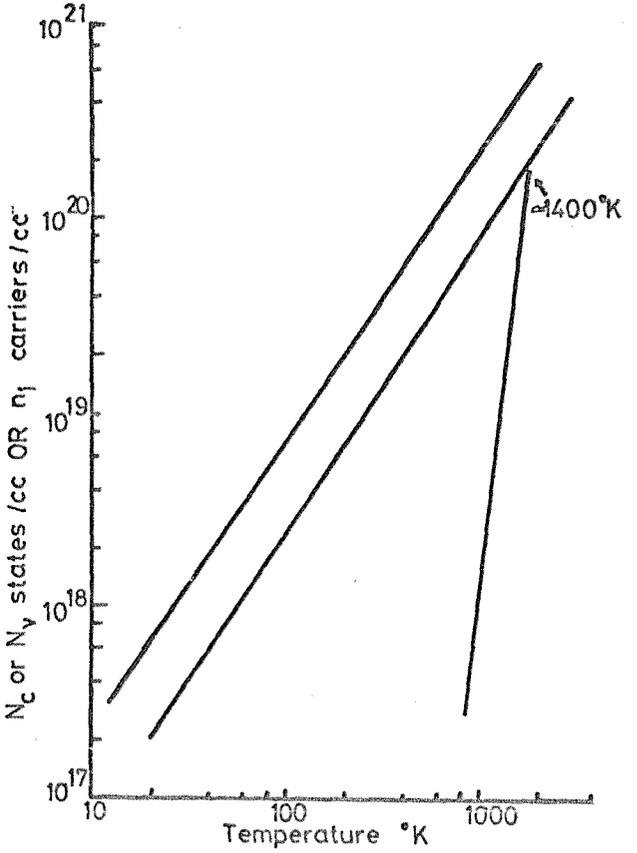


FIG 6.8: N_c, N_v and n_i versus temperature for silicon.

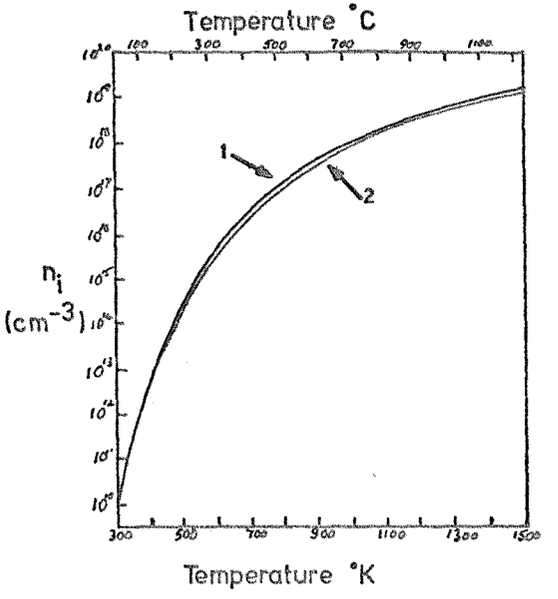


FIG. 6.9: Temperature dependence of the intrinsic carrier concentration n_i in silicon.

(1) Measured values of Morin and Maita (1954)

(2) Calculated values (see text) of Jain and Overstraeten (1974).

developed in 6.1 are evaluated for heavily doped p-type and n-type silicon. The solubility enhancement of gold in heavily phosphorus doped silicon is considered in the greatest detail since the development of a model which predicts the gettering effect of shallow phosphorus diffused layers is one of the prime objects of this work.

Solution of equations 6.6, 6.10, 6.21, 6.23, 6.43 and 6.45 requires knowledge of several parameters in silicon at the diffusion temperature. These are : the positions of the Fermi level in intrinsic and extrinsic silicon, the magnitude of the ion pairing constant between gold and the dopant of interest and the gold energy level positions and degeneracies.

6.4.1 The Fermi level in intrinsic silicon at high temperatures

In intrinsic non-degenerate silicon, the Fermi level is given by : (Blakemore,1962, p 95)

$$E_{Fi} = \frac{Eg}{2} + kT \ln \left[\frac{N_v}{N_c}\right]^{\frac{1}{2}} \qquad 6.47$$

where the symbols have the previously defined meanings. N_v and N_c are given by: (Blakemore p 79)

$$N_v = 2 \left[\frac{2\pi m_h kT}{h^2}\right]^{3/2} \qquad 6.48$$

$$N_c = 2 \left[\frac{2\pi m_e kT}{h^2}\right]^{3/2} \qquad 6.49$$

where m_e and m_h are the density of states effective masses of holes and electrons in the appropriate energy bands.

Equation 6.47 becomes :

$$E_{Fi} = \frac{Eg}{2} + kT \ln \left[\frac{m_h}{m_e}\right]^{3/4} \qquad 6.50$$

If, at the diffusion temperature $n_i < N_c$ and N_v, the position of E_{Fi} may be calculated if E_g, m_h and m_e are known. The condition $n_i < N_c$ and N_v, which ensures that the silicon is nondegenerate, is fulfilled below 1400°K (see figure 6.8).

The value of the silicon energy band gap may be obtained by extrapolation of the data of MacFarlane et al.(1958), who indicate that the variation of Eg with temperature is linear above 300°K (see section 4.3.1) and obeys the relationship :

$$Eg(T) = 1.205 - 2.8 \times 10^{-4}.T \qquad 6.51$$

The behaviour of m_e and m_h as T varies is only tabulated up to 600°K. Stradling and Zhukov (1966), Ulkhanov and Mal'stev (1964) and Barber (1967) found that m_e and m_h both increase with temperature up to this limit. No data are available for higher temperatures, however Barber (1967) suggests that for higher temperatures N_v and N_c are inversely proportional to the energy band gap Eg. Using this assumption, and the values of Eg, m_e and m_h at 300°K (1.121 eV, $1.18m_o$ and $0.81\ m_o$ respectively, where m_o = electron rest mass)

$$m_e(T) = \left[K_1 \cdot \frac{1}{Eg(T)}\right]^{2/3} .m_o = \left[\frac{1.4369}{Eg(T)}\right]^{2/3} .m_o \qquad 6.52$$

$$m_h(T) = \left[K_2 \cdot \frac{1}{Eg(T)}\right]^{2/3} .m_o = \left[\frac{0.8172}{Eg(T)}\right]^{2/3} .m_o \qquad 6.53$$

The accuracy of this approximation may be tested by calculating the intrinsic carrier concentration, n_i, and comparing the result with measured values of n_i from Morin and Maita (1954). This has been done by Jain and Overstraeten (1974) and the result, which indicates good agreement, is shown in figure 6.9. Equation 6.50 may be evaluated using 6.52 and 6.53.

6.4.2 The Fermi-level in extrinsic silicon at high temperatures

The Fermi level in extrinsic silicon may be obtained by solution of the charge balance equation for the system :

$$n + N_A^- = p + N_D^+ \qquad 6.55$$

(This method is used and discussed fully in chapter 7). For nondegenerate situations Boltzmann statistics are used so that :

$$n = N_c \exp \frac{(E_F - E_c)}{kT} \qquad 6.56$$

In the examples of interest here, solubility enhancement of gold is only observed in very heavily doped layers. In such situations the doping concentration is close to or greater than the density of available quantum states and Boltzmann statistics may not be used. Equation 6.56 assumes the form obtained when Fermi-Dirac statistics are used :

$$n = N_c \mathcal{F}_{\frac{1}{2}} \frac{(E_F - E_c)}{kT} \qquad 6.57$$

where $\mathcal{F}_{\frac{1}{2}}$ is the Fermi integral of order $\frac{1}{2}$. (When $E_F - E_c \gg kT$, the Fermi integral $\mathcal{F}_{\frac{1}{2}}(\eta)$ approximates to the Boltzmann form : $\exp(\eta)$). Bullis (1966) and Brown et al. (1975) used this equation to obtain the extrinsic Fermi level for calculations of gold solubility enhancement. Other authors (Joshi and Dash, 1966; Dorwark and Kirkaldy, 1969; Cagnina, 1969 ; Chou and Gibbons, 1975) have only used the Boltzmann approximation.

Unfortunately the situation for heavy doping is further complicated by a departure of the energy dependence of N_c or N_v, from the normal $E^{\frac{1}{2}}$ relationship (Nichols and Vernon, p 77). This makes direct use of equation 6.57 invalid. At high impurity concentrations, interactions between ionised impurity atoms (N_D^+ or N_A^-) and free carriers (n or p) can change the crystal properties. The result is a deformation of the majority carrier band-edge known as 'band-edge tailing'. This effect is illustrated schematically in figure 6.10.

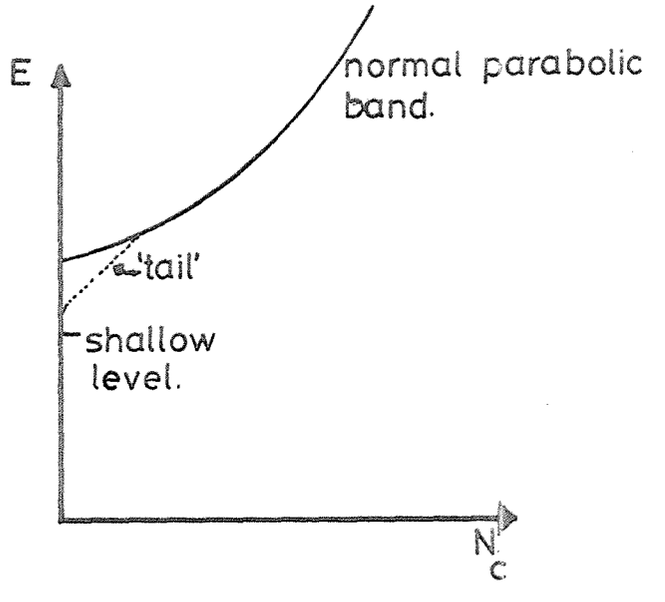


Figure 6.10. Conduction band edge tailing at high doping concentrations

In addition to this effect, coulombic interaction between the ionised dopant atoms also becomes significant, resulting in the discrete impurity energy levels splitting into a continuous impurity band. Kleppinger and Lindholm (1971) in calculations of this effect have shown that in heavily phosphorus doped silicon the impurity band merges with the conduction band. As a result the ionisation energy of the phosphorus atoms is reduced to zero.

The band edge tailing and impurity banding effects have been considered in detail by Kane (1963), Bonch-Bruyevich (1966) and Fistul (1969). Jain and Overstraeten (1974) have calculated the position of the Fermi level in phosphorus doped silicon in the temperature range 900 - 1200°C, using Kane's equation for band edge tailing and an expression due to Morgan (1965) for the density of states in the impurity band. It is particularly interesting to note that, even at extremely high phosphorus concentrations (10^{22} atoms/cc), the predicted Fermi level still lies within the silicon energy band gap whereas equation 6.57 predicts a value which lies above the conduction band edge.

In calculations of the solubility of gold in heavily phosphorus doped silicon presented here, the Fermi level values predicted by Jain and Overstraeten are used.

6.4.3 Evaluation of the ion-pairing constant K_p

The ion-pairing constants in silicon between gold and phosphorus, gold and arsenic and gold and boron have been evaluated using equation 6.43.

The nearest neighbour distance, a, was taken to be the sum of the atomic radii of the paired atoms; values given by Wolf (1969) and Van Vechten and Phillips (1970) were used. The value of N, the density of lattice sites is equal to 5×10^{22}/cc at 300°K and an allowance for thermal expansion was made using a volume expansion coefficient of 6.99×10^{-6} per degree centigrade (Wolf, 1969).

The dielectric constant, ε_{si}, presents a problem since the validity

of using the macroscopic value to describe the microscopic interaction between two ions on nearest neighbour lattice sites is questionable. Srinivasan (1969) has calculated the microscopic dielectric constant in silicon and shows that it reaches its macroscopic value within one nearest-neighbour distance. On the basis of this result, the macroscopic value ($\varepsilon_{si} = 11.7$) has been used, although the effect of reducing it, which is also illustrated, can be large since it appears inside an exponential in equation 6.43. Values of K_p obtained with $\varepsilon_{si} = 11.7$ for the gold-phosphorus pairing reaction result in a much better fit to experimental data than those obtained using a smaller ε_{si} and hence a larger K_p.

The values of K_p obtained using various parameters are summarised in table 6.1:

Pairing Ions	Atomic Radii $\overset{o}{A}$	Temperature ^oK	ε_{si}	K_p cm^{-3}
$Au^- + P^+$	Au = 1.5 P = 1.128	1273	11.7	5.762×10^{-21}
		1373	11.7	4.225×10^{-21}
		1473	11.7	3.232×10^{-21}
		1273	8.0	3.0×10^{-20}
$Au^- + A_s^+$	Au = 1.5 A_s = 1.18	1273	11.7	5.304×10^{-21}
		1373	11.7	3.669×10^{-21}
		1473	11.7	2.834×10^{-21}
$Au^+ + B^-$	Au = 1.5 B = 0.88	1273	11.7	9×10^{-21}

Table 6.1 Ion Pairing Constants in Silicon

The difference in pairing constant between gold and phosphorus and gold and arsenic goes some way to explaining the difference in Cagnina's

(1969) measurements of gold solubility in phosphorus and arsenic doped silicon (see figure 6.3). The pairing constant between positively charged gold and negatively charged boron is much larger than that between negatively charged gold and positively charged phosphorus; however, for the same boron or phosphorus concentration $N_{Au}^{+} \ll N_{Au}^{-}$ and the overall effect between gold and boron is negligible (see figure 6.26).

6.4.4 Remarks on published solubility enhancement calculations

Most theoretical treatments of the solubility enhancement of gold in heavily doped silicon have either incorrectly used non-degenerate assumptions (see 6.2.2) or have not included the band tailing and impurity banding effects in obtaining E_{Fex}. The notable exceptions are O'Shaughnessey et al. (1975) and Meek and Seidel (1975). The former authors, whose work has already been reviewed in 6.3, only calculated the gold solubility enhancement for a phosphrous concentration of 4×10^{19} atoms/cc with one gold energy level and degeneracy model : they gave no details of the values or the method used to calculate E_{Fex}. Ion-pairing was ruled out despite the theoretical values falling short of the measured values (see figure 6.5).

Meek and Seidel, using a Mass Action approach to the Fermi level and ion-pairing effects, employed the electron and hole activity coefficients predicted by Hwang and Brews (1971) which included band tailing and impurity banding effects. Meek and Seidel's model, however, is open to considerable criticism since it was assumed that all of the gold was negatively charged at the diffusion temperature - even in heavily doped p-type silicon where gold would be expected to be predominantly ionised as a donor. No details of the gold energy levels are given either. The assumption that all gold was negatively charged was made because in Meek and Seidel's calculations, when all of the charge states of gold were allowed, a considerable solubility enhancement in p-type (boron doped) silicon was predicted despite experimental evidence to the contrary.

The factor which was not appreciated by any of the authors cited, is the sensitivity of the predicted enhancements both to the position of the Fermi-level in the heavily doped material and more especially to the positions and degeneracies of the gold energy levels at the diffusion

temperature. If all of the gold charge states are included, it is still possible to predict experimentally observed gold solubility enhancements providing the "correct" gold energy levels and degeneracies are used. Valid comparison of the various gold energy level models can only be made if the Fermi level, E_{Fex}, is identical in each case. In order to compare the effects of the various gold energy level and degeneracy models without any constraint due to the actual position of E_{Fex}, the results of the Fermi-level-solubility-enhancement calculations are initially presented as a function of the Fermi level position. E_{Fex} will, of course, range from close to the valence band edge in heavily doped p-type silicon to the conduction band edge in heavily doped n-type silicon.

Ion pairing effects, which require values of E_{Fex} and the corresponding ionised shallow dopant density (equation 6.19) may also be compared provided the same values are used in each case.

The Fermi-level solubility enhancement results are presented first and only those models which do not actually contradict experimental evidence are subsequently used in the ion pairing calculations.

6.4.5 "Fermi level effect" solubility enhancement

The gold energy level and degeneracy models which are to be considered have been described in chapter 4. It is assumed here that the gold energy levels remain discrete and that the energy level temperature dependence models may be extrapolated to elevated temperatures. The positions of each gold level may be separately considered to remain fixed to E_c, fixed to E_v, or to vary in proportion to E_g as temperature varies. It is not necessary to carry out calculations for every possible combination of these models, since in heavily doped n-type material the solubility enhancement is virtually unaffected by small variations in the position of the gold donor level and in heavily doped p-type material the solubility enhancement is virtually unaffected by small variations in the position of the gold acceptor level. In heavily doped n-type material the same solubility enhancement would be obtained if both gold levels are fixed to E_c or if only the gold acceptor is fixed to E_c.

The calculations have been carried out for the following models of temperature dependence :

A : Both gold levels vary in proportion to $E_g(T)$
B : Both gold levels are fixed to E_c
C : Both gold levels are fixed to E_v.

Combinations such as one level obeying model A and the other level obeying model B may be inferred from these results.

Each of the models was additionally subdivided by considering the various degeneracy factors which could be associated with each gold energy level (chapter 4). These are :

(1) Gold monovalent : $g_A = 2/3$, $g_D = 1/4$
(2) Gold monovalent with split-off valence band included in obtaining degeneracies : $g_A = 4/3$, $g_D = 3/4$
(3) Gold trivalent : $g_A = 4$, $g_D = 3/2$
(4) Gold trivalent with split-off band included : $g_A = 6$, $g_D = 5/2$.

The gold energy levels for each model were extrapolated from the 0^oK values given by Engstrom and Grimmeiss (1975) and Wong and Penchina (1975) - see chapter 4.

In the results presented in figures 6.11 to 6.17, the models and degeneracies are designated by the identifying numbers given above (e.g. A1 = both gold levels vary in proportion to $E_g(T)$, degeneracies : $g_A = 2/3$, $g_D = 1/4$).

The figures illustrate the Fermi level solubility enhancement effects for these models as follows :

Fig	6.11	A1, 2, 3, 4.	1273^oK
Fig	6.12	B1, 2, 3, 4.	1273^oK
Fig	6.13	A1, 2, 3, 4.	1373^oK
Fig	6.14	B1, 2, 3, 4.	1373^oK
Fig	6.15	A1, 2, 3, 4.	1473^oK
Fig	6.16	B1, 2, 3, 4.	1473^oK
Fig	6.17	C1, 2, 3, 4.	1273^oK

A Levels vary $E_g(T)$
B Levels vary E_c
C Levels vary E_v

	g_A	g_D
1:	2/3	1/4
2:	4/3	3/4
3:	4	3/2
4:	6	5/2

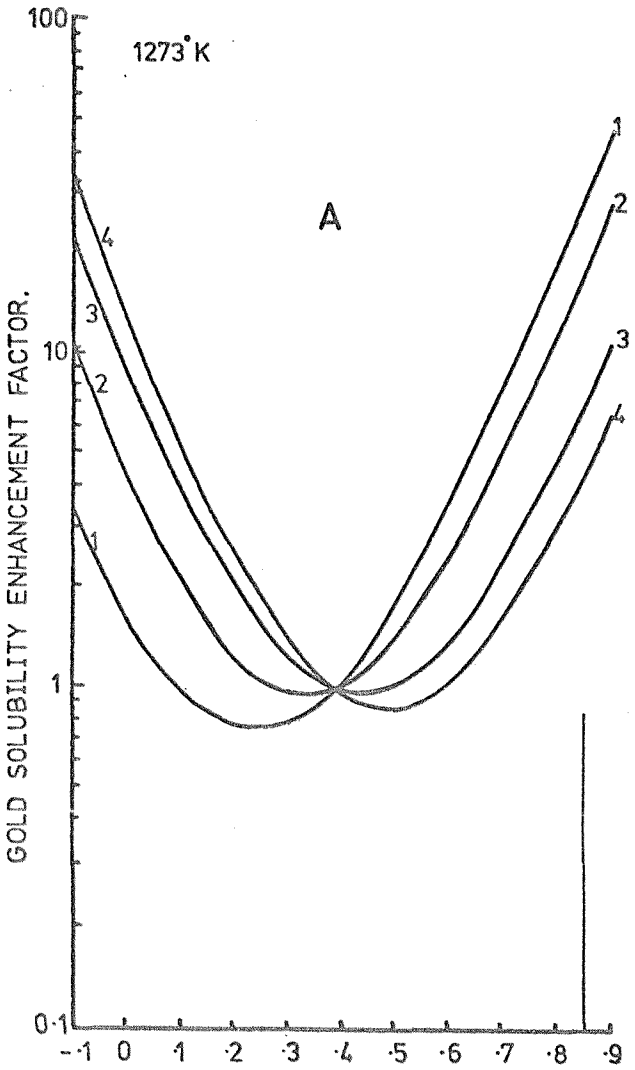


FIG. 6.11: Substitutional gold solubility enhancement as a function of the Fermi-level. 1273°K
Gold energy level positions vary in proportion to $E_g(T)$
See text for degeneracy values.

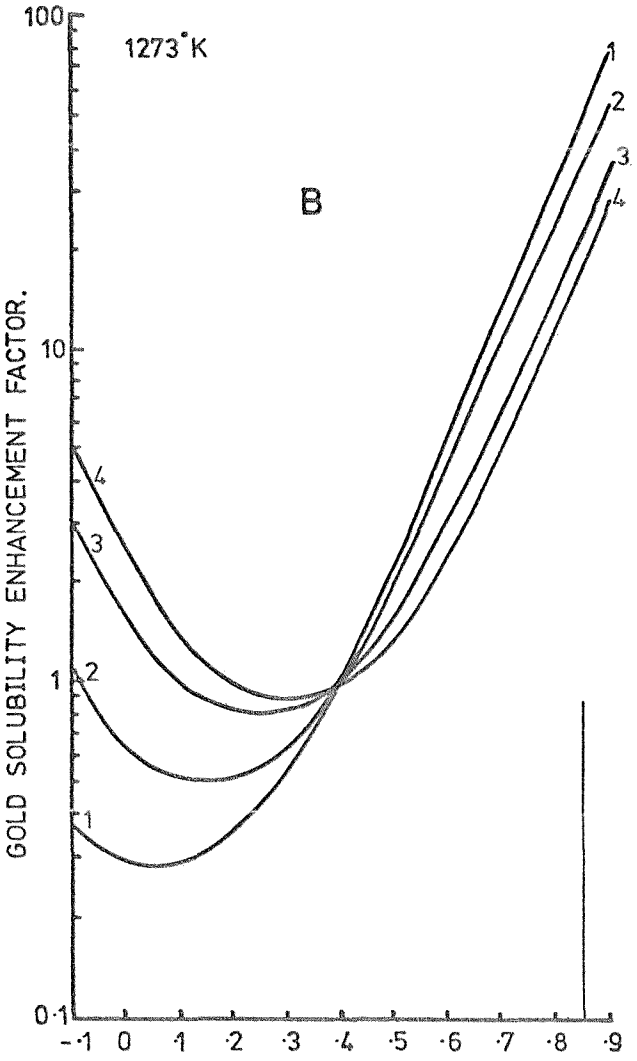
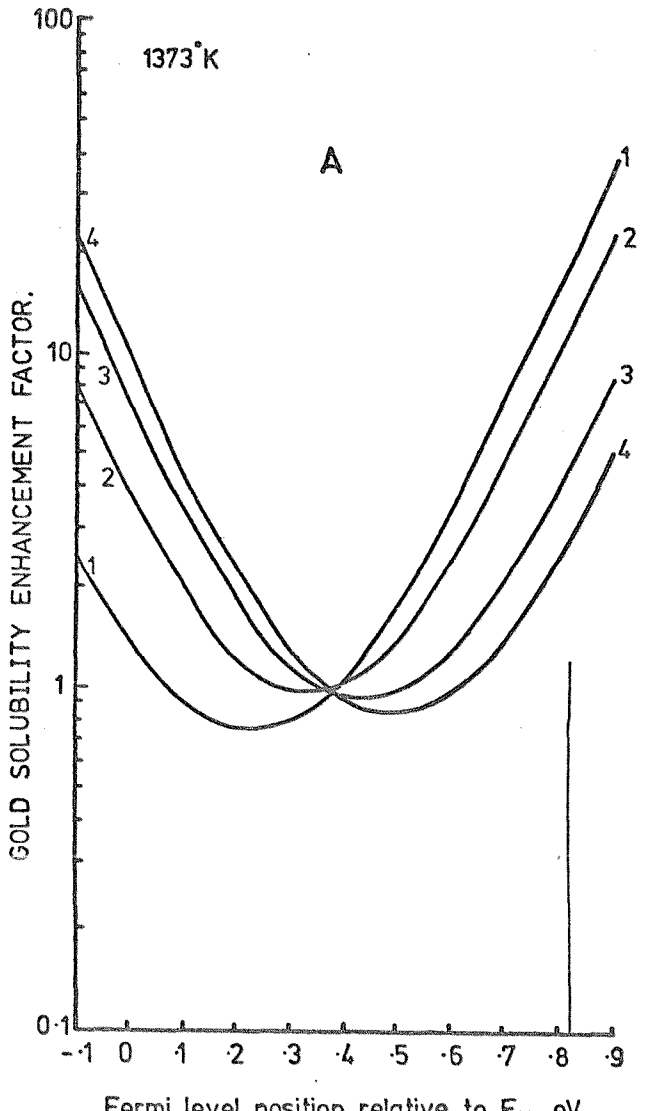


FIG. 6.12: Substitutional gold solubility enhancement as a function of the Fermi-level. 1273°K
Gold energy levels fixed with respect to E_c as temperature varies. See text for degeneracy values.



A Levels vary $E_g(T)$
B Levels vary E_g
C Levels vary E_c E_v

	g_A	g_D
1:	2/3	1/4
2:	4/3	3/4
3:	4	3/2
4:	6	5/2

FIG. 6.13: Substitutional gold solubility enhancement as a function of the Fermi-level. 1373°K
Gold energy level positions vary in proportion to $E_g(T)$
See text for degeneracy values.

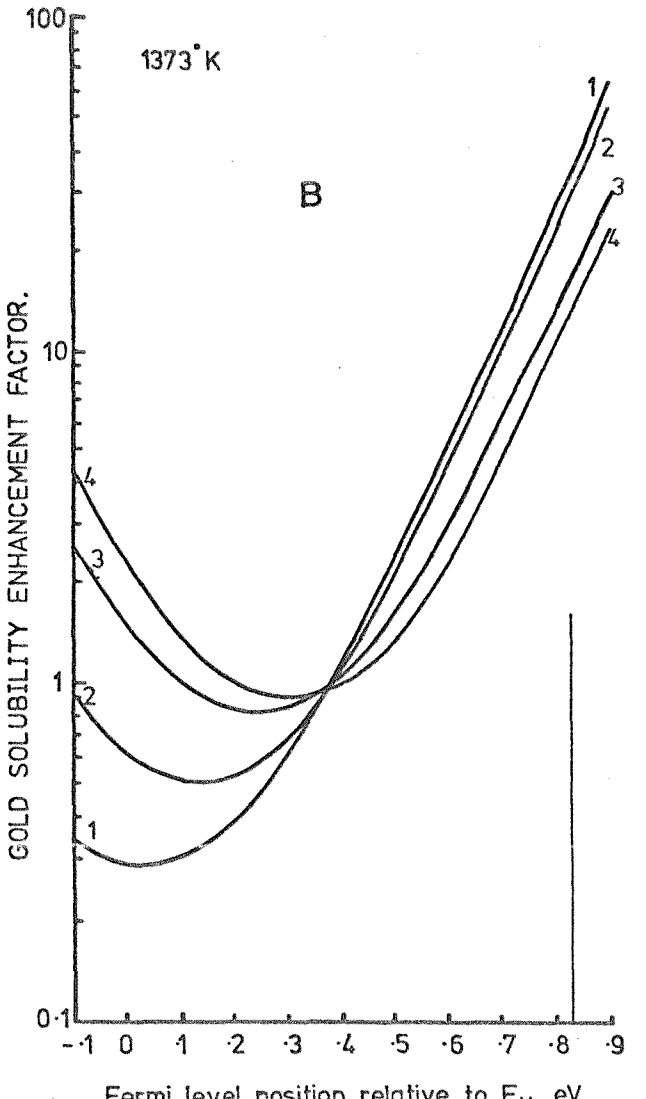
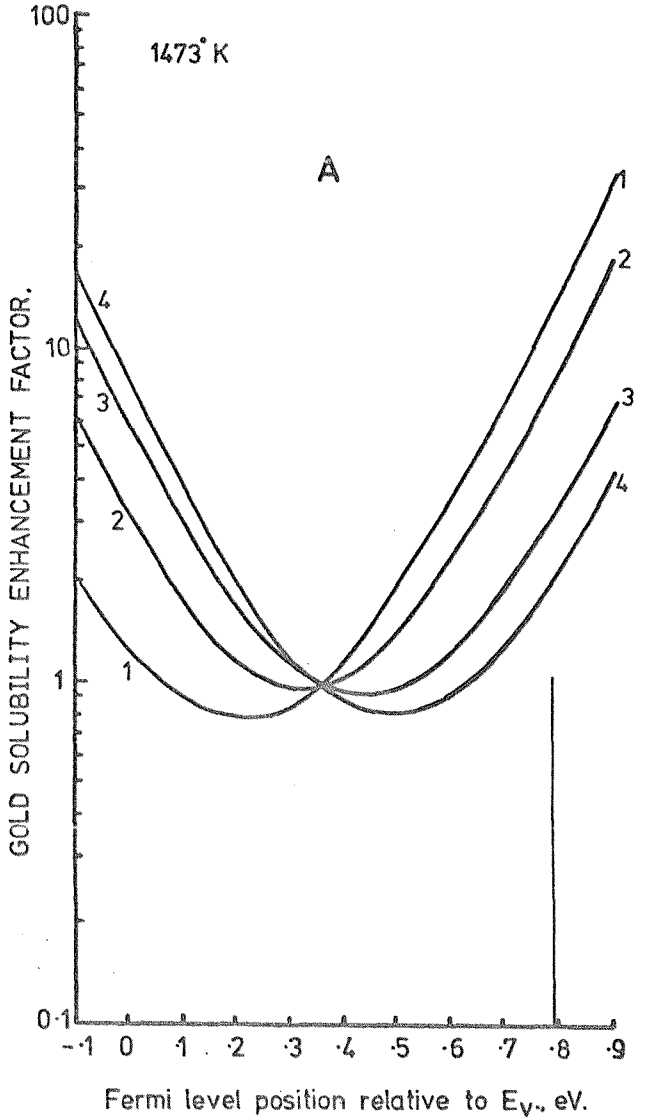


FIG. 6.14: Substitutional gold solubility enhancement as a function of the Fermi-level. 1373°K
Gold energy levels fixed with respect to E_c as temperature varues. See text for degeneracy values.



A Levels vary $E_g(T)$
B Levels vary E^g
C Levels vary E^c_v

	g_A	g_D
1:	2/3	1/4
2:	4/3	3/4
3:	4	3/2
4:	6	5/2

Fig. 6.15: Substitutional gold solubility enhancement as a function of the Fermi-level. 1473°K
Gold energy levels vary in proportion to $E_g(T)$
See text for degeneracy values.

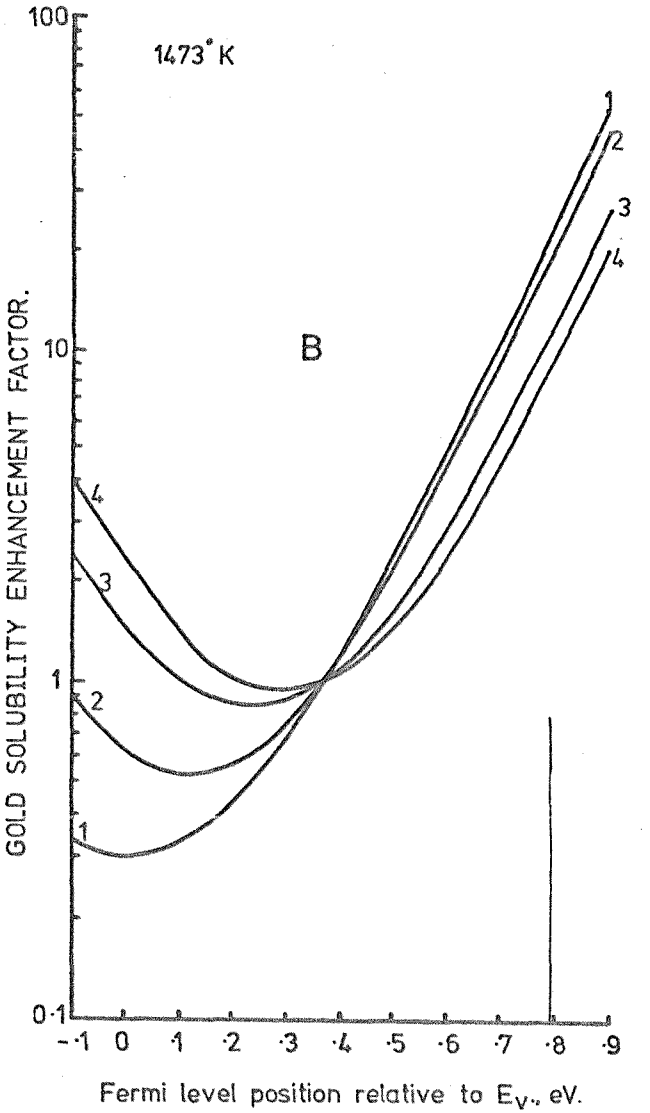


Fig. 6.16: Substitutional gold solubility enhancement as a function of the Fermi-level. 1473°K
Gold energy levels fixed with respect to E_c as temperature varies. See text for degeneracy values.

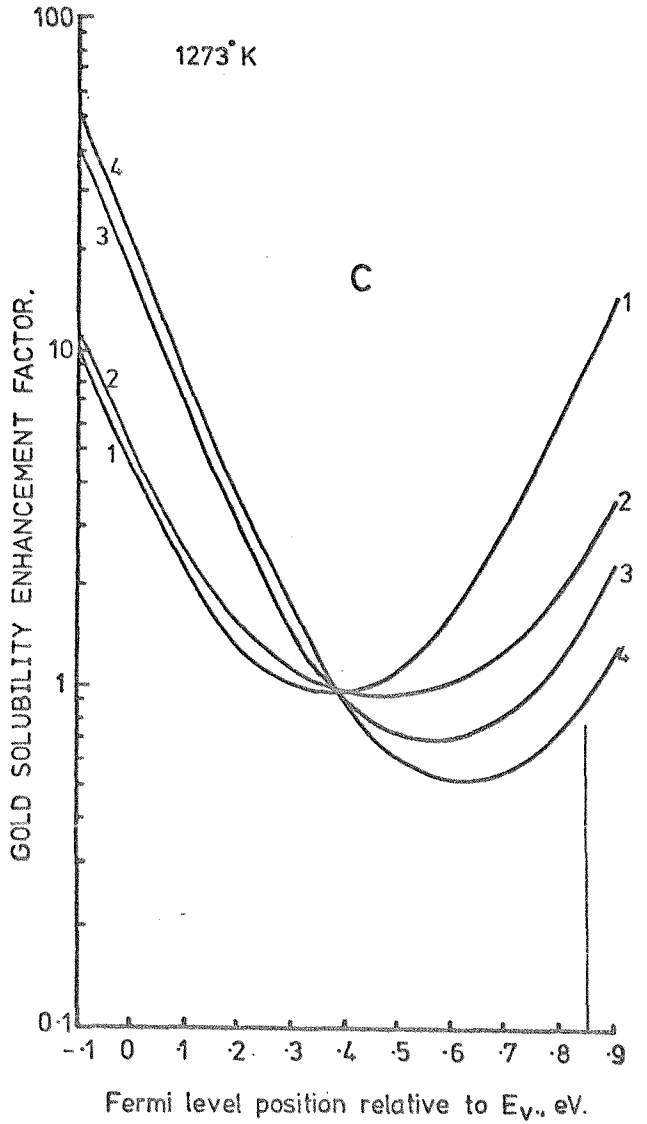


Fig. 6.17: Substitutional gold solubility enhancement as a function of the Fermi-level. 1273°K
Gold energy levels fixed with respect to E_v as temperature varies. See text for degeneracy values.

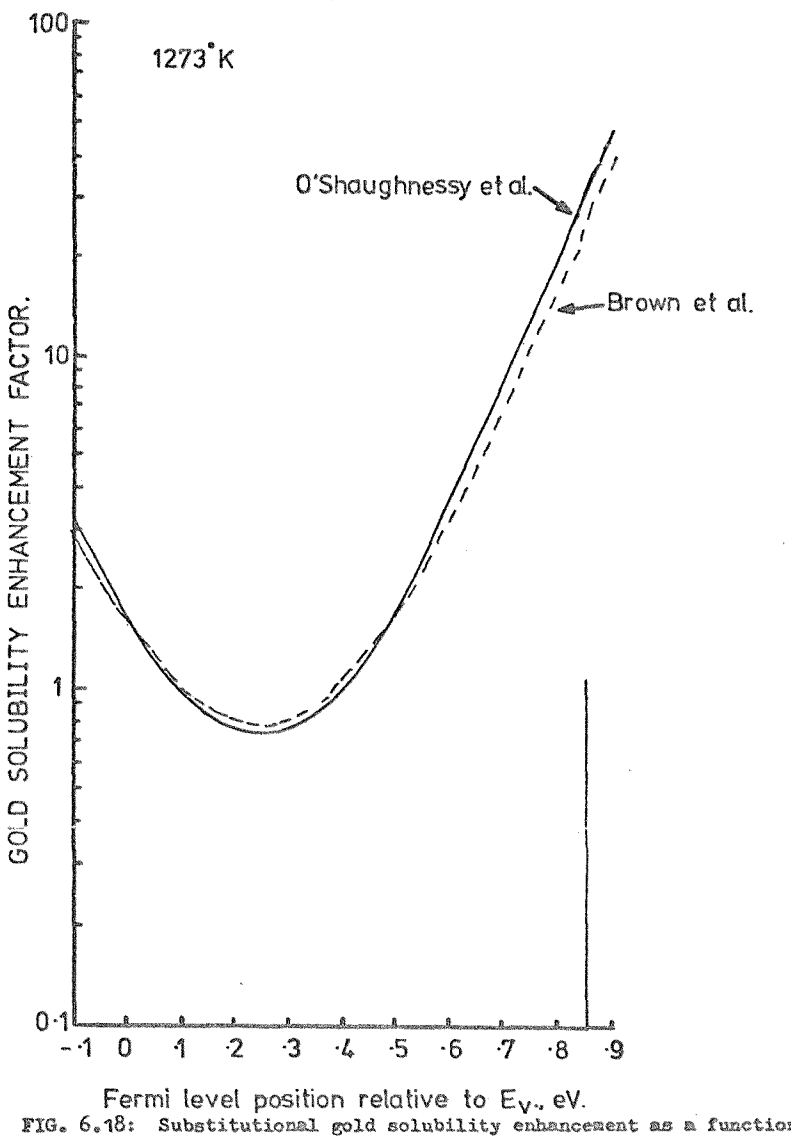


FIG. 6.18: Substitutional gold solubility enhancement as a function of the Fermi-level. Comparison of models used by O'Shaughnessy et al., (1974) and Brown et al., (1975), at 1273°K.

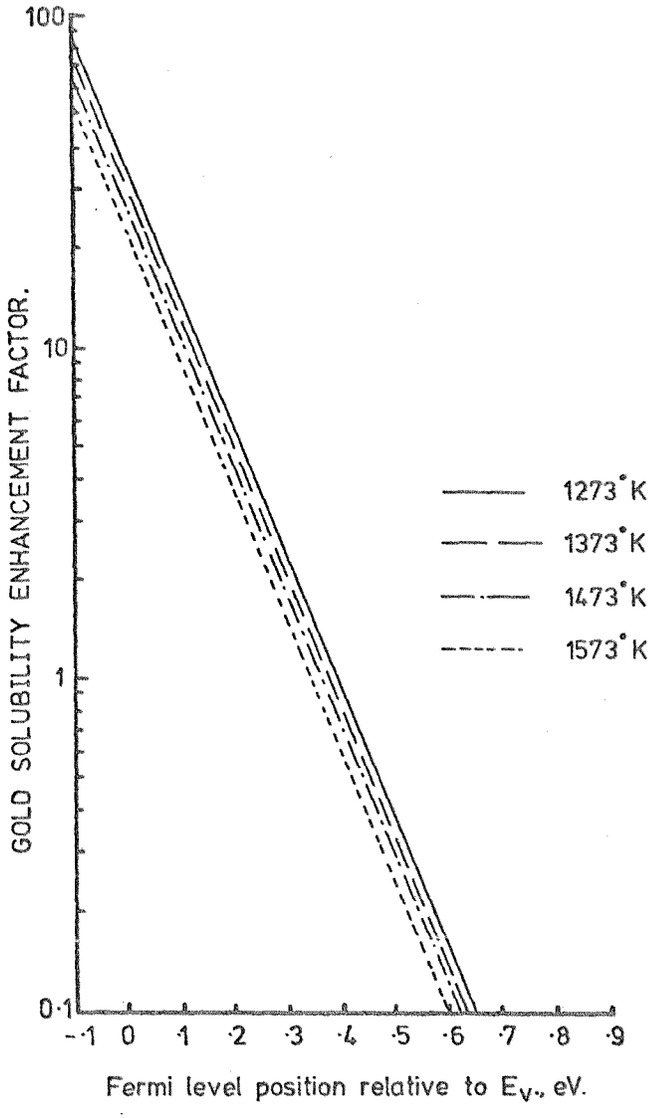


FIG. 6.19: Interstitial gold solubility enhancement as a function of the Fermi-level and temperature.

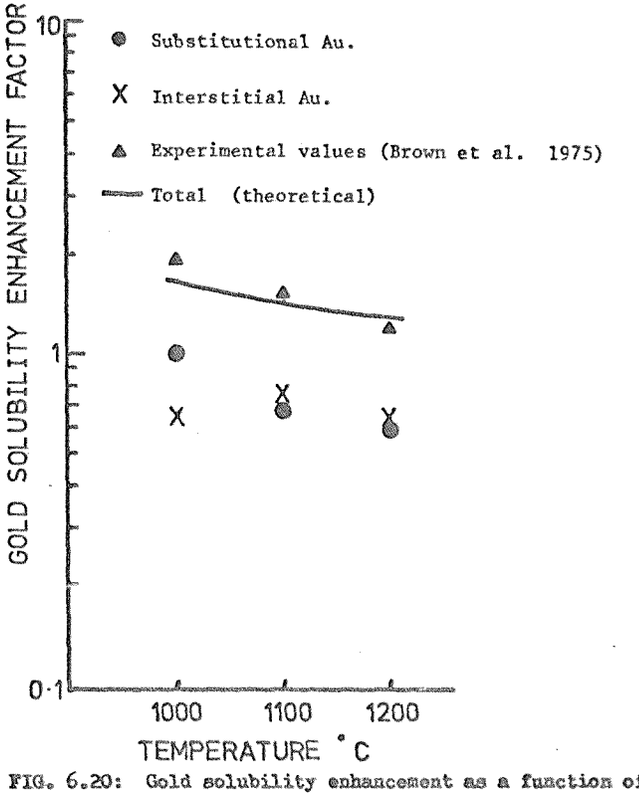


FIG. 6.20: Gold solubility enhancement as a function of temperature for silicon doped with 9×10^{19} boron atoms/cc.

The cross over point of the curves in each figure occurs at the intrinsic Fermi level position for that temperature. There is no solubility enhancement, of course, for the intrinsic case. The zero of the energy scale is at the valence band edge and the position of the conduction band edge, E_c, is marked on each diagram.

Comparison of figures 6.11, 6.12, and 6.17, in the light of the well documented experimental results, allows certain of the models to be rejected. The relevant experimental information is : significant gold solubility enhancement occurs in uniformly, heavily doped n-type silicon ; very large solubility enhancements are noted in shallow diffused, highly doped phosphorus layers ; only a very small solubility enhancement has been observed in uniformly heavily doped p-type silicon and in some experiments a solubility depression has been measured. The models which can be rejected immediately since they predict large solubility enhancements in p-type silicon are : A3 and A4, C1, C2, C3 and C4. Models B1 and B2 can also be rejected since they predict a very large solubility depression in heavily doped p-type silicon, which although in accord with the results of Dorward and Kirkaldy (1969) are contrary to the more reliable results of Brown et al.

Figure 6.18 compares models A1 and B3 since they are the ones used by O'Shaughnessey et al. and Brown et al. respectively. It is particularly interesting to note that these two models predict an almost identical solubility enhancement at every position of E_{Fex} although neither author realised this because they had obtained E_{Fex} with different models.

Figure 6.19 illustrates the Fermi level solubility enhancement of interstitial gold if it behaves as a shallow donor at the diffusion temperature. The precise position of the shallow energy level is not a critical parameter in this calculation provided it is close to E_c. Quite large enhancements are predicted when E_{Fex} is close to E_v (p-type silicon). An accurate calculation of the enhancement expected for silicon doped with 9×10^{19} boron atoms/cc (as used by Brown et al) requires knowledge of E_{Fex} at this doping concentration. No values are available in the literature, however an approximate calculation based on a value of E_{Fex} inferred from the results of Jain and Overstraeten

(1974) for phosphorus doped silicon may be made. The ratio of interstitial to substituional gold at the solubility limit was taken from the data of Wilcox et al.(1964) - see chapter 3. Figure 6.20 illustrates the overall Fermi-level solubility enhancement due to both interstitial and substitutional gold compared to the results of Brown et al. (Model B3 was used for the substitutional gold species). A surprisingly close fit between the calculated and observed results is obtained considering the way in which E_{Fex} was determined. If interstitial gold is a shallow donor, then it is almost certainly the main cause of the solubility enhancement observed by Brown et al. A boron-gold pairing effect which could add to this enhancement is shown in figure 6.26 to be negligible. An attempt to solve the problem of which gold species is responsible for the observed enhancement in boron doped silicon is described in chapter 8.

6.4.6 Ion-Pairing Calculations

Using values of the Fermi level which were calculated by Jain and Overstraeten (1974) the ion pairing effect on gold solubility has been determined. The values obtained at 1273^{o}K due to pairing between gold and phosphorus are shown in figures 6.21 and 6.22. The solubility enhancement is plotted against the ionised phosphorus concentration for models A1, 2, 3, and 4 in figure 6.21 and for models B1, 2, 3, and 4 in figure 6.22. (Models A3 and 4 and B1 and 2 have already been rejected but are included here for comparison). The trends illustrated at this temperature are applicable to other diffusion temperatures of interest. It should be noted that $N_p^+ = N_p^{[total]}$ when the impurity band merges with E_c.

If gold is monovalent with its energy levels varying in proportion to $E_g(T)$, the enhancement due to ion-pairing is only slightly higher than if the trivalent model with the gold energy levels fixed to E_c is used. As there is only a small difference between these models, in both the Fermi-level and ion-pairing effects, model B3 only is used in the results which follow.

Figure 6.23 illustrates the effects of varying the pairing constant K_p

	g_A	g_D
1:	2/3	1/4
2:	4/3	3/4
3:	4	3/2
4:	6	5/2

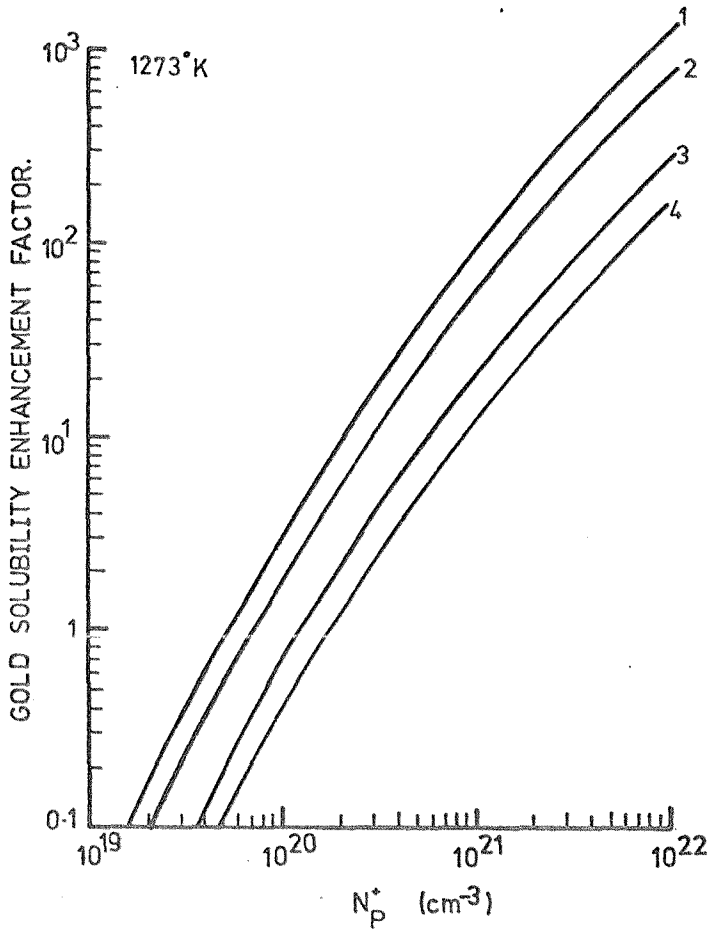


FIG. 6.21: Gold solubility enhancement due to Ion-Pairing. Energy levels vary in proportion to $E_g(T)$.

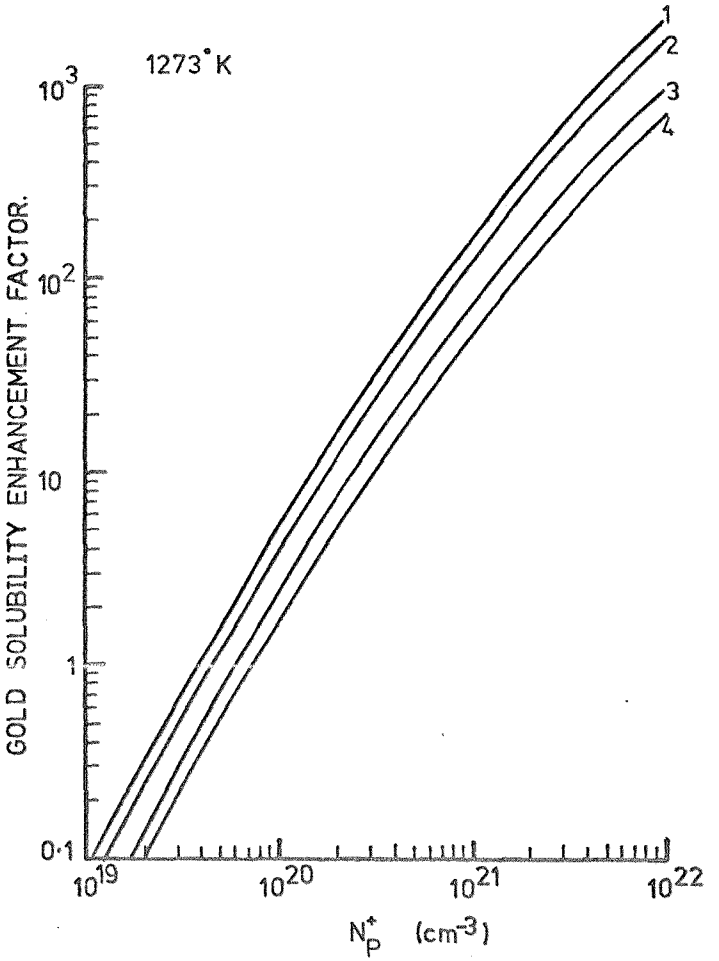


FIG. 6.22: Gold solubility enhancement due to Ion-Pairing. Energy levels fixed to E_c as temperature varies.

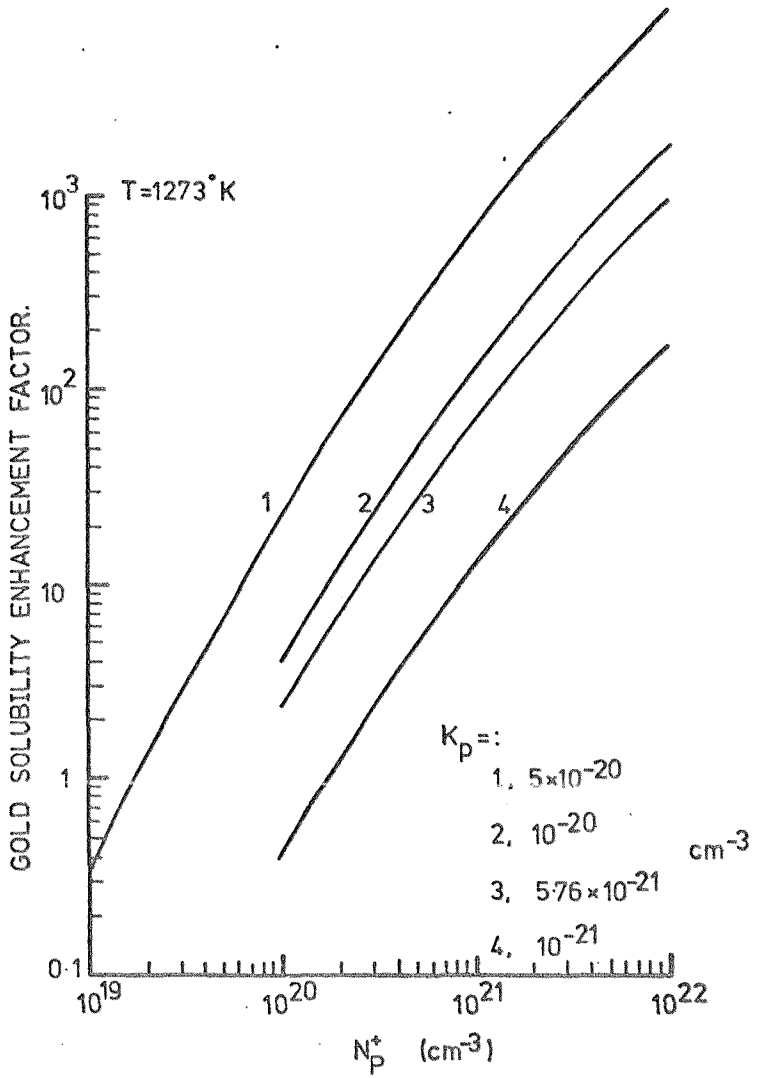


FIG. 6.23: Effect of varying K_p on Ion-Pairing Effect.

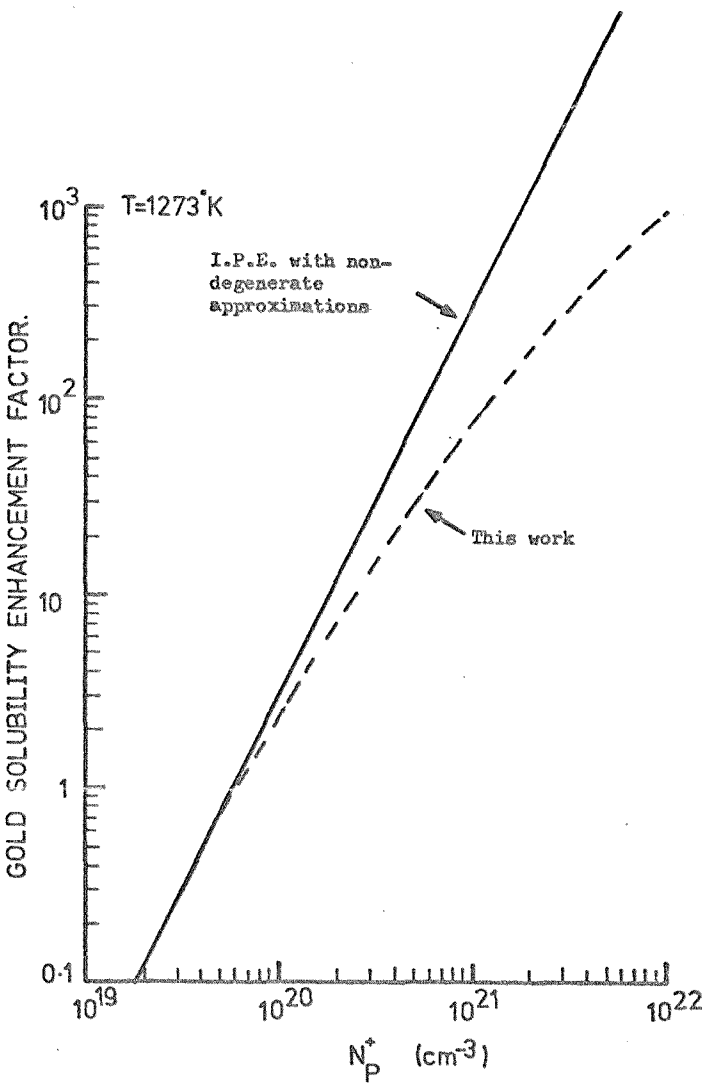


FIG. 6.24: Ion-Pairing effect using non-degenerate approximations

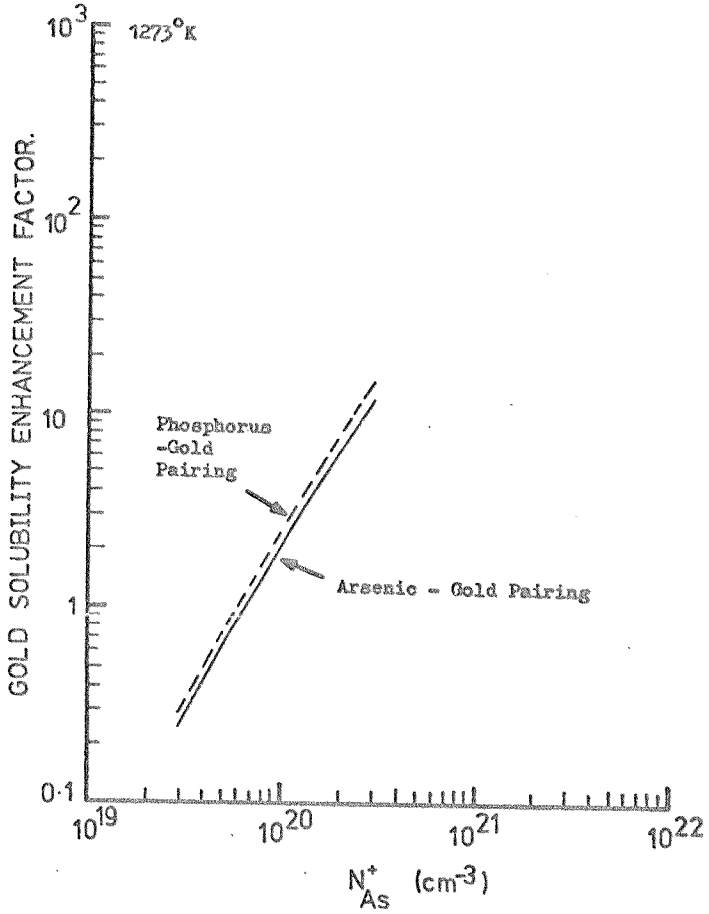


FIG. 6.25: Ion-Pairing effect between arsenic and gold.

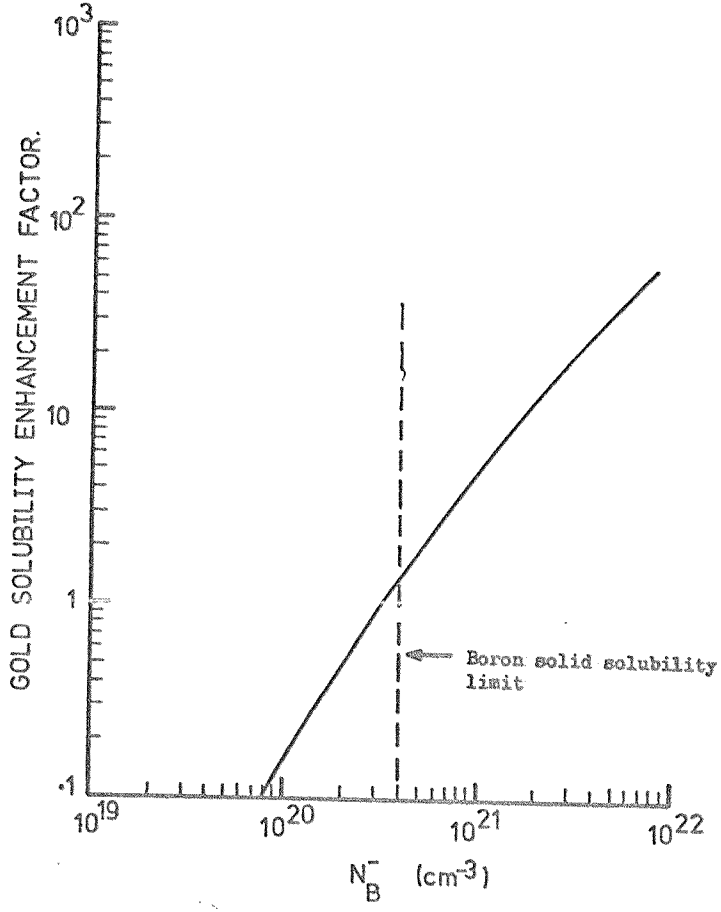


FIG. 6.26: Ion-Pairing effect between boron and gold.

at 1273^{o}K using model B3. Curve number 3 on this figure is for the same pairing constant used in figures 6.21 and 6.22.

Figure 6.24 contrasts the predicted solubility enhancement obtained in this work with that obtained using equation 6.33 where non-degenerate assumptions were made ; the predicted enhancement, at high phosphorus concentrations is much higher in this case.

Figure 6.25 compares the ion pairing enhancement in phosphorus doped and arsenic doped silicon. The predicted enhancement due to arsenic is 80% of that due to phosphorus whereas Cagnina (1969) measured an enhancement 60% of that due to phosphorus; a small error in K_p could easily explain the difference between the value calculated here and the value measured by Cagnina.

Figure 6.26 illustrates the predicted ion-pairing effect in boron doped silicon. Despite the higher value of the gold-boron pairing constant the predicted effect is insignificant even at the solid solubility limit of boron in silicon (4 x 10^{20} atoms/cc, Trumbore 1960). The gold solubility enhancement in boron doped silicon which is shown in figure 6.20 will not be affected by ion pairing.

6.4.7 Total solubility enhancement in phosphorus doped silicon

The total gold solubility enhancement factor in heavily phosphorus doped silicon is illustrated for three different temperatures in figures 6.27, 6.28 and 6.29. In all cases the model used is that of trivalent gold with its energy levels fixed to E_c (Model B3). Results obtained with models A1, A2 and B4 are not substantially different. Experimental results obtained by Cagnina (1969) and O'Shaughnessey et al. (1975) in uniformly phosphorus doped wafers in the range 1 x 10^{19} to 10^{20} phosphorus atoms/cc are also plotted on these diagrams. A very close correspondence between the measured and predicted values is apparent. The other experimental values, which were obtained in shallow diffused phosphorus regions, will be discussed later.

These figures also indicate the solubility enhancement due to the Fermi level effect alone. Ion pairing can be seen to be the dominant reaction,

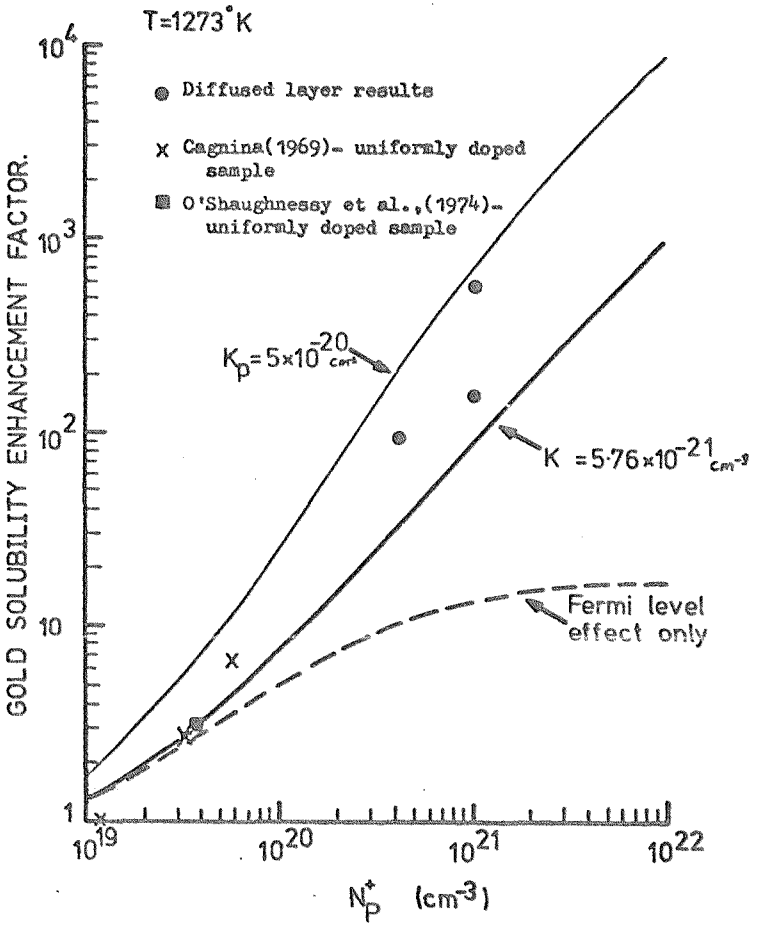


FIG. 6.27: Total gold solubility enhancement versus phosphorus concentration. 1273°K

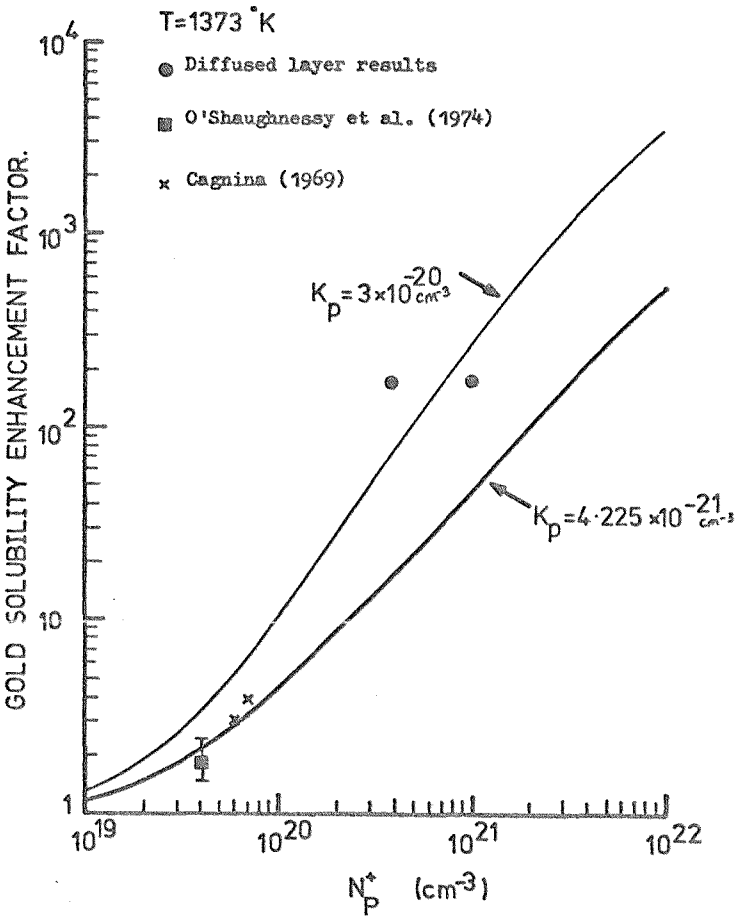


FIG. 6.28: Total gold solubility enhancement versus phosphorus concentration 1373°K

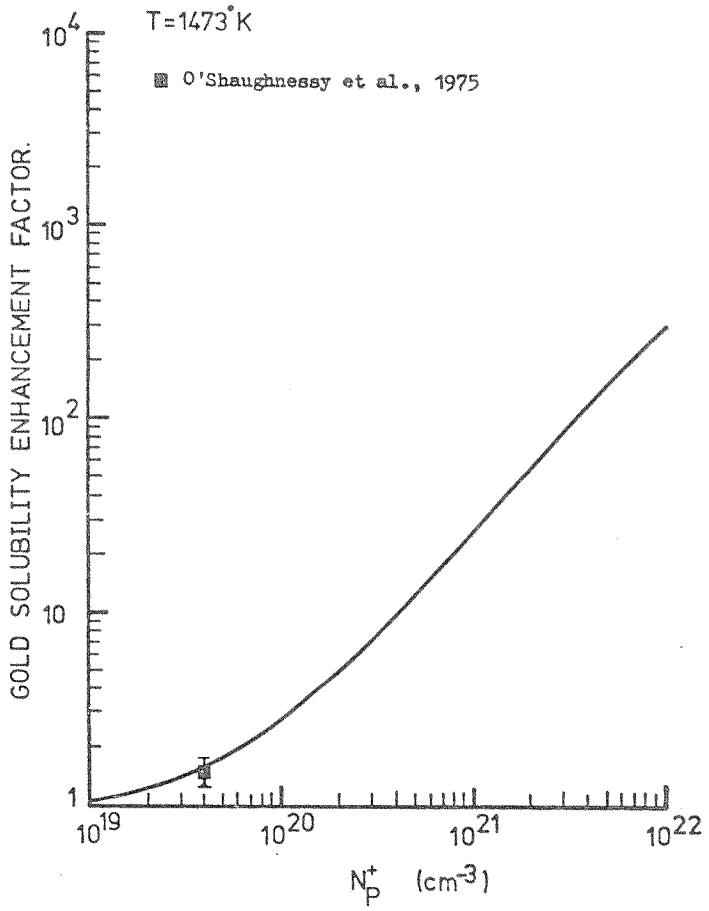


FIG. 6.29: Total gold solubility enhancement versus phosphorus concentration. 1473°K

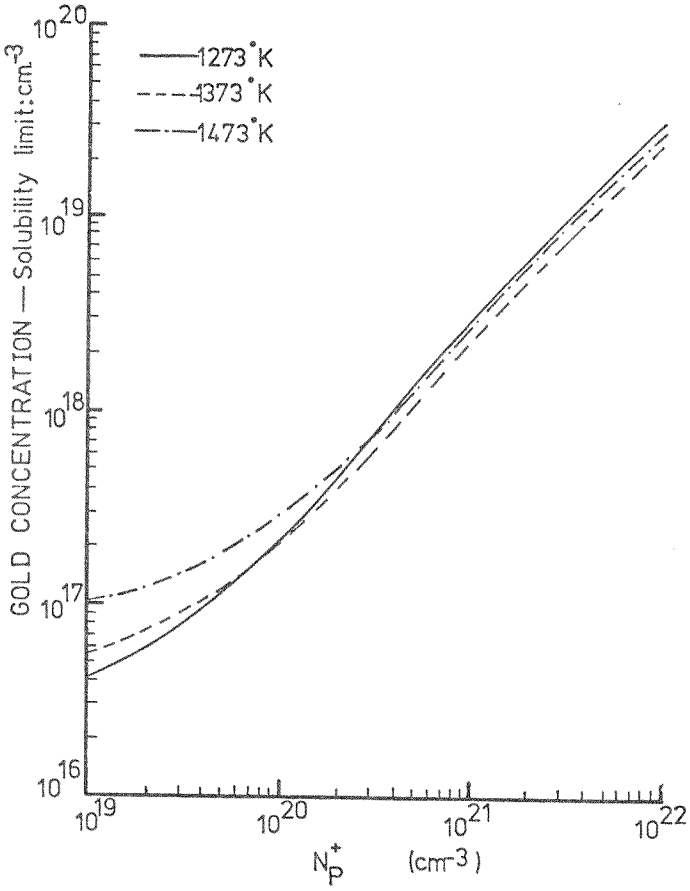


FIG. 6.30: Gold concentration at the solubility limit versus phosphorus concentration. (Obtained using intrinsic gold solubility data of Brown et al., 1975).

even at fairly low solubility enhancements. The total enhancement that would be predicted if the pairing constant between phosphorus and gold were $\simeq 5 \times 10^{-20}$ cm^{-3} is shown on figure 4.27. A much higher overall solubility enhancement, which does not fit the measured results on uniformly doped silicon, is obtained. The results from phosphorus diffused layers do lie close to this line, however it will be shown in a following section that these enhancements are probably dominated by a further effect, (see 6.5.2. ff).

Comparison of figures 6.27, 6.28 and 6.29 shows that the solubility enhancement for a given phosphorus concentration is much greater at the lower temperatures. Figure 6.30 illustrates the actual gold concentrations at the solubility limit as a function of phosphorus concentration at various temperatures. The gold solubilities in intrinsic silicon were taken from the data of Brown et al.(1975). Despite the greater enhancement at lower temperatures the overall solubility is similar because of the lower intrinsic solubility at lower temperatures.

It should be pointed out that by careful choice of the positions of the gold energy levels, their degeneracies and the ion pairing constant, almost any model could be made to fit the experimental results. The fact that there is good experimental evidence at lower temperatures for the gold energy levels being fixed to E_c, coupled with the fact that the trivalent degeneracies also provide a good fit to the resistivity data presented in chapter 7, must lend strong weight to the model used here.

6.5 Gettering of Gold

The greatly increased solubility of gold in phosphorus doped or crystallographically damaged areas of silicon results in a redistribution of gold present in a wafer when such a region is introduced. The phenomenon, whereby an impurity such as gold may become trapped in a localised area of a wafer, is known as gettering. As gold is a very effective recombination centre in both n-type and p-type silicon, keeping its concentration to a minimum is very important for devices where high

minority carrier lifetime is desired. In this application, gettering techniques applied on the opposite side of a silicon wafer to the active devices are of the greatest technological importance. Paradoxically, the gettering effect presents serious problems to the manufacturers of fast switching and power devices where gold, deliberately added as a minority carrier lifetime "killer", is undesirably segregated into heavily doped n-type regions of the devices. A better understanding of the gettering effect would clearly be of benefit in both cases.

The three techniques which have been reported for gettering are : phosphorus diffusion, ion-implantation damage or mechanical damage all applied to the back surface of the silicon wafer. It is proposed to consider the phosphrous diffusion gettering effect in some detail here since it is felt that this process offers the most reproducible and effective removal of accidental gold contamination. In addition,it is the unwanted effect of phosphorus gettering that generally results in difficulties when deliberate gold doping of devices is required.

Gold is not the only undesirable lifetime killer found in silicon wafers ; copper, iron, nickel, cobalt, and platinum are also a problem. Gold, however, is by far the most ubiquitous contaminant, especially in high temperature furnace tubes where, unlike other metals, it does not oxidise and become 'harmless'. Any gettering process must be aimed particularly at the removal of gold.

There is a large amount of data on gettering available in the literature. It is not proposed to review these papers in very great detail, rather an attempt is made to elucidate the phosphorus gettering process by considering aspects of the mechanism which have been ignored in previous works. Experimental results which are presented in chapter 8 of this thesis confirm the more important aspects of the dynamics of the model described here. It will become clear from this model, that by understanding what happens during the gettering process - rather than just obtaining the desired result empirically - more efficient use may be made of it.

6.5.1 Brief review of literature on gettering

Reports of gold gettering effects may be divided into two categories. Those which measure gettering of deliberately contaminated wafers which contain a large amount of gold and those which only measure gettering of very small quantities of gold more typical of unwanted contamination levels. The latter type of measurement presents problems of experimental technique because detection of gold concentrations of the order of 10^{12} atoms/cc is necessary. Neutron activation analysis offers the only reliable method for doing this provided large enough sample volumes can be obtained. Experiments which require detection of deliberately added gold may be carried out with radio-tracer gold isotopes or, at gold concentrations above 10^{17} atoms/cc, Rutherford backscattering may be used. The effect of gold on the resistivity of lightly doped silicon may also be employed to detect its presence ; this has been done by Brotherton and Rogers (1972) and in this work using the spreading resistance technique (see Appendix A).

The relevant references, together with brief details of the order of magnitude of the gold concentration, the gettering techniques and the gold measurement technique are given in table 6.2. Brief comments indicating the main results are also made.

Comparison of some of the early results on gettering due to the deposition of phosphorus and boron glass layers on the silicon surface indicates a certain amount of disagreement over what is actually responsible for the gettering effect. For example Adamic and McNamara (1964) showed that most of the gettered gold resided in the phosphorus doped layer in the silicon wafer, whereas Lambert and Reese (1968) found significant quantities of gold in the phosphorus doped glass. Similarly Nakamura et al.(1968) found gold in the phosphorus glass and very large amounts of gold in boron glass layers when gettering with boron was attempted. More recent neutron activation results presented by Murarka (1976) show that the phosphorus glass has no effect and that gettering is solely carried out in the phosphorus diffused layer. Meek and Seidel (1973) and Meek, Seidel and Cullis (1975) also showed that the phosphorus glass was not responsible for gettering.

Authors	Amount of gold contamination	Gettering Techniques	Measurement Technique	Comments
Adamic et al. (1963)	Deliberate gold doping	Oxidation of silicon surface	Radiotracer Au.	Au found piled up at Si - SiO_2 interface. SiO_2 found to be a good mask against Au.
Adamic et al. (1964)	Deliberate gold doping	Boron glass. Phosphorus glass.	Radiotracer Au.	Boron glass gettering: Au found at glass-Si interface. Phosphorus glass gettering: Most Au found in P-doped layer in Si.
Joshi et al. (1966)	Deliberate gold doping	Phosphorus Diffusion	Radiotracer Au.	High concentrations of Au found in P-diffused layers.
Lambert et al. (1968)	Low level of gold contamination (10^{13} a/cc)	Phosphorus diffusion	Neutron activation	Au found in P-glass as well P-diffused layer. Amount of Au in glass very temperature dependent. Samples cooled extremely slowly (see text).
Nakamura et al. (1968)	Deliberate gold doping	Phosphorus glass. Boron glass. Damage (lapping).	Radiotracer Au.	Phosphorus glass gettering: some Au in glass, most in P-diffused layer. Boron glass gettering: 80% Au in B-glass. Damage gettering: Au gettered into damaged layer.
Brotherton et al. (1972)	Deliberate gold doping	Phosphorus emitter diffusion in n.p.n. structure.	Spreading resistance measurements.	Large gettering effect due to phosphorus diffusion noted - not well quantified.
Buck et al. (1972 & 1973)	Deliberate gold doping	Ion implantation damage	Rutherford back-scattering (RBS)	Ion implantation damage found to getter efficiently Gettering rate dependent on Au_i diffusion rate at 900°C.
Seidel et al. (1973)	Deliberate gold doping	Ion implantation damage	RBS	High concentrations of Au (~10^{19} a/cc) found in damaged layers.
Mielke (1975)	Deliberate gold doping	Phosphorus diffusion	Diode reverse recovery 4 pt. probe resistivity Neutron activation	Gold accumulated in phosphorus layer - suggested to be caused by Au-P compound formation. Results difficult to interpret.
Chou et al. (1975)	Deliberate gold doping	Phosphorus diffusion	RBS	High concentrations of Au found in P-diffused layers. 90% on substitutional lattice sites.
Meek et al. (1975)	Deliberate gold doping	Phosphorus and Boron diffusions	RBS	Boron - very little gettering. Phosphorus - high concentrations of substitutional Au in layers; no Au detected in P-glass.
Seidel et al. (1975)	Deliberate gold doping	Ion implantation damage vs phosphorus diffusion	RBS	Comparison of efficiency of I.I.damage gettering (caused by various ions) with phosphorus diffusion gettering. Phosphorus diffusion gettering equal to or better than damage from most ions.
Lisiak et al. (1976)	Deliberate gold doping	"Process induced" gettering, phosphorus and boron from spin-on films.	Neutron activation	Some Au in spun-on P-glass, most in P-diffused layer. No B gettering but 'significant' accumulation of Au at glass-Si interface.
O,Shaughnessey et al. (1976)	Au depositied on surface	Redistribution of surface Au during oxidation +P-glass	Radiotracer Au	Growing oxide rejects surface Au to the oxide surface. ∴ oxide makes good mask. Phosphorus glass does not getter Au but P-diffused layer does
Sigmon et al. (1976)	Deliberate gold doping	Ion implantation damage	RBS	Gettering depends on residual ion implantation damage. Effectiveness varies on wafers of different orientation and can be lost during post-gettering heat treatments.
Murarka (1976)	Moderate gold contamination (10^{13}-10^{14} a/cc)	Phosphorus diffusion	Neutron activation	No Au in P glass. Effective gettering of Au into P-diffused layer. Very significant furnace contamination.
Brown (1976)	Deliberate gold doping	Phosphorus diffusion	Radiotracer Au	Some Au found at glass-Si interface, large amounts of Au in P-diffused layers.

Table 6.2. Review of literature on gettering

O'Shaughnessey et al.(1976) found no significant increase in gold solubility in phosphorus glass compared with silicon dioxide which they showed to be a very effective mask against gold diffusion. It is likely that this sort of effect, and the measurements of significant quantities of gold at the Si - SiO_2 interface (Willoughby and Brown, 1973), are mainly due to redistribution of gold during cooling of the samples from the diffusion temperature - Lambert and Reese used a very slow cooling cycle indeed ($60^{o}C$ per hour). The discovery of extremely large (81% of the total) amounts of gold in the glass of "boron glass gettered" wafers, by Nakamura et al, may reflect the fact that, since any solubility enhancement in the boron diffused layers would be interstitial gold which is very mobile, large scale rejection to the surface could easily occur in a very short period during cooling.

From these results the clearest conclusion is that a reproducible gettering effect due to the introduction of a shallow phosphorus diffused layer does occur. This reproducibility, and the ability of a phosphorus layer to retain its gold solubility enhancement - and therefore gettering - properties during subsequent processing are the most important aspects of the phenomenon. In their comparison of phosphorus diffusion gettering and ion-implantation gettering, Seidel et al.(1975) showed that ion implantation damage can, in some instances, be a more effective process than phosphorus gettering. The gettering efficiency of the implanted damage is closely related to the amount of residual damage after the post-implantation anneal. It is this 'annealing out' of implantation damage that presents the most serious drawback to the technique. Sigmon et al.(1976) show that the residual implantation damage can vary considerably on silicon wafers of different orientations. The residual disorder also depends strongly on the thermal history of a given sample during and after the implantation.

For the conventional application of a gettering cycle at the very end of the device process, such limitations are slight, however it is argued here that the correct time in the processing schedule for gettering is as early as possible. Ideally a high concentration phosphorus gettering layer should be the very first process applied to the back of a silicon wafer while the front surface is protected by a low temperature deposited oxide.

6.5.2 Phosphorus Diffusion Gettering Model

In almost all of the publications listed in table 6.2, the observed phosphorus gettering effects have only been explained - if at all - by invoking enhanced solubility due to precipitation, compound formation, the Fermi level effect or ion pairing. Although these effects undoubtedly may and do contribute to the enhanced solubility of gold in the phosphorus rich layers (although the first two effects are minimal - see 6.2.3.), the gettering phenomenon requires several processes, each of which may be the rate limiting step, to occur. These are :

(a) Dissolution of gold atoms from the sites occupied before gettering.

(b) Transport of gold atoms to the gettering layer.

(c) Reaction of the gold atoms with the gettering agent.

If each of these processes is fully understood, it may be possible to apply the phosphorus gettering process to better advantage.

Process (a) will not be considered in detail here since it depends very much on what the site occupied by the gold before gettering is, and the activation energy required to remove a gold atom from that site. For gold atoms residing on substitutional lattice sites, process (a) is shown in experiments described in chapter 8 not to be rate limiting ; in the case of gold precipitates on lattice defects, the kinetics of dissolution are likely to be complex. In this discussion, the reasonable assumption that process (a) is very rapid will be made. (If gettering is applied in the manner suggested in this chapter, the likelihood of process induced defects becoming decorated with gold atoms should be greatly reduced).

Process (b), transport of gold to the gettering layer, may be considered in terms of the gold diffusion process already described in chapter 3. The main flux of diffusing gold is carried by the interstitial species ; to all intents and purposes substitutional gold may be considered to be immobile. Therefore process (b) will only be efficient if the gold diffusing species is interstitial.

Process (c), the reaction of gold atoms with the gettering agent must also be fast. Indeed for very efficient gettering this process should be fast enough for (b) to be the rate limiting step. That is, gold should be gettered into the layer as fast as it can arrive and the layer should appear to be an infinite sink for gold.

The only reported attempt to identify the rate of a gettering process was by Buck et al. (1972 and 1973) who identified the gold interstitial diffusion rate as the rate limiting process for ion implantation damage gettering. The reaction which gave rise to this rapid gettering was not considered.

In this thesis, experiments reported in chapter 8 show that phosphorus diffusion gettering is also rate limited by the diffusion of interstitial gold. This implies that reaction (c) is very rapid inside the phosphorus diffused layer.

It is a common opinion that the gettering process in the phosphorus layer is similar to that in an ion implanted layer - namely : precipitation of gold onto lattice damage induced by the phosphorus diffusion. If this were the case, one would not expect to find gold atoms within the gettering layer residing on substitutional lattice sites and similarly one would not expect to observe significant gettering in a phosphorus diffused layer which had not given rise to diffusion induced damage. Seidel and Meek (1973), Chou and Gibbons (1975) and this work, chapter 8, show that most of the gold within a phosphorus gettered layer is on substitutional lattice sites (identified using channelled Rutherford backscattering). In addition the gettering experiments described in this thesis are carried out on phosphorus diffused layers which exhibit little damage disorder.

Although lattice damage may contribute a small effect to phosphorus diffusion gettering, the main mechanism is, as shown in previous sections, ion pairing between phosphorus and gold ions. Such a reaction requires a negatively charged gold ion to reside on an adjacent substitutional lattice site to a positively charged phosphorus ion. As the flux of gold arriving in the phosphorus layer is interstitial, two reactions must occur ;

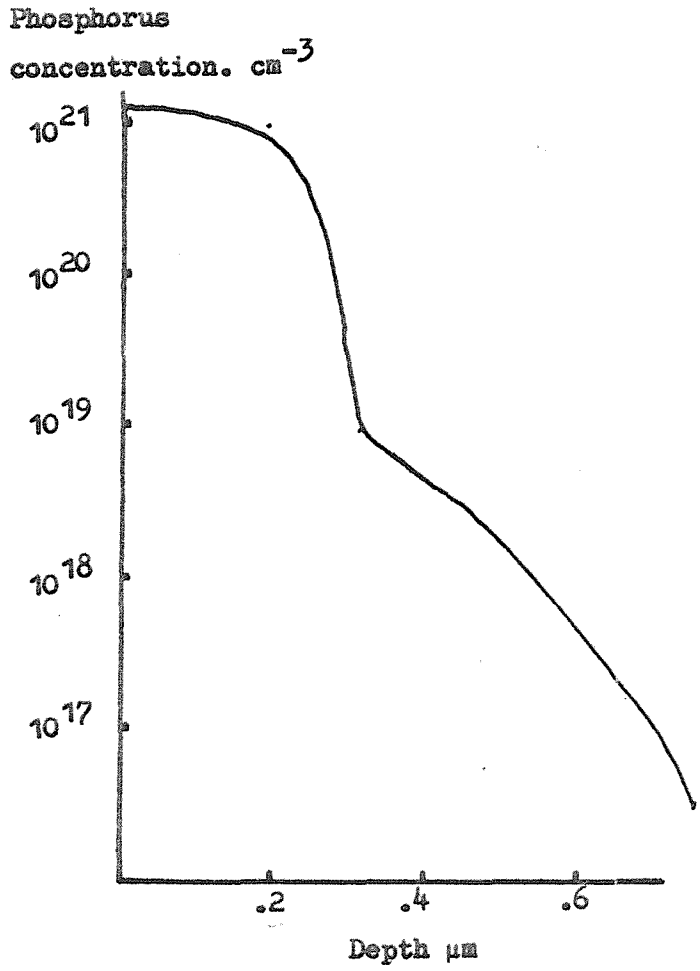


FIG. 6.31(a): Typical high concentration shallow phosphorus diffusion showing kink and fast diffusing tail.

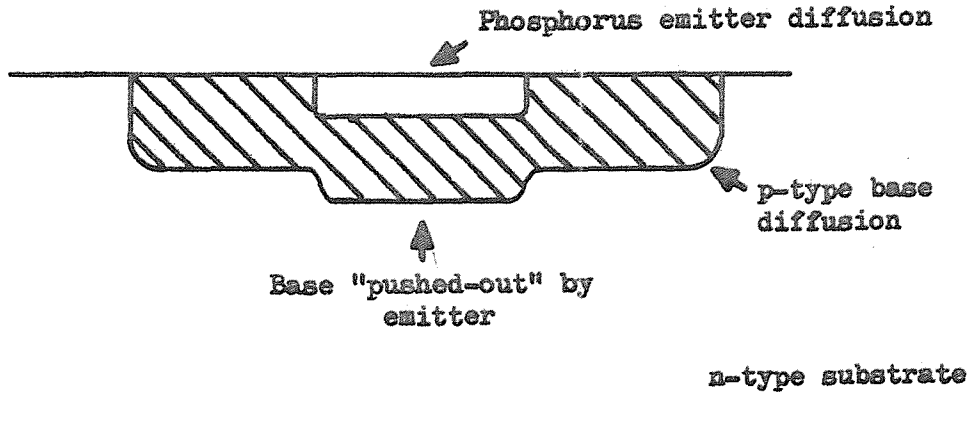


FIG. 6.31(b): Schematic section through an n-p-n transistor with a phosphorus emitter showing pushed out base.

firstly the interstitial gold atom must react with a vacancy in the lattice to become substitutional and secondly the substitutional gold atom (which will probably be negatively ionised) must diffuse to a substitutional site adjacent to a phosphorus ion. The first reaction ($Au_i + V \rightarrow Au_s$) is shown by Huntley (1972) and in chapter 8 of this work to be very fast indeed providing there are vacancies available - close to the wafer surface this should be the case. The second reaction, however, can not be expected to occur rapidly at all because of the relative immobility of the substitutional gold species.

Overall, therefore, the phosphorus-gold pairing reaction which is the dominant cause of enhanced gold solubility, would not appear to be the fast reaction required for efficient gettering because of the slow substitutional gold diffusion rate. However, the phosphorus layer undoubtedly getters efficiently!

The mechanism proposed here to explain the rapidity of the phosphorus gettering process, is that the phosphorus atoms are generally on nearest neighbour sites to lattice vacancies. This means that, when an interstitial gold atom which enters the gettering layer reacts with a lattice vacancy, the condition for ion-pairing is immediately satisfied. If the phosphorus-vacancy pair does exist in large concentrations, then the gettering process will be limited simply by the rate at which gold can diffuse to the gettering layer.

The phosphorus-vacancy pair is by no means a new concept, indeed it has received much attention in the literature. The E-centre, as it is generally known, was shown to exist by Watkins and Corbett (1964) and is thought to be responsible for anomalies associated with high concentration phosphorus diffusions. Figure 6.31 (a) illustrates the shape of a typical high concentration phosphorus diffusion profile in which a kink followed by a rapidly diffusing tail is observed. Figure 6.31 (b) illustrates the phenomenon of "base push-out" or "emitter dip" in which enhanced diffusion of the p-type base dopant under a phosphorus emitter in an n-p-n transistor structure occurs.

Schwettman and Kendall (1972) have proposed that phosphorus diffusion

occurs mainly in the form of E-centres and that the kink in the diffusion profile is due to a difference in diffusivity between the two charge states of the E-centre as it passes the point where the Fermi-level corresponds to the E-centre energy level. (The E-centre is thought to be an acceptor 0.4 eV below E_c).

Peart and Newman (1972) suggest that the phosphorus diffusion tail is caused by the break-up of E-centres. This is attributed, once again, to diffusing E-centres crossing the point at which the Fermi-level corresponds to the E-centre level ; when this occurs, the change in charge state makes the E-centre unstable.

Jones (1974) in a study of the push-out effect, has considered the E-centre break-up model in some detail since the excess vacancies so produced may be the cause of enhanced diffusivity of the base dopant under the phosphorus emitter. Using the assumption that the base dopant (boron or gallium in Jones' study) diffuses by a simple vacancy mechanism, push-out is considered in terms of an enhanced diffusivity due to a vacancy excess. A diffusion enhancement factor, χ, is estimated from experimental results of push-out. For the simple vacancy diffusion mechanism, $\chi = V/V^*$, where V is the enhanced vacancy concentration in the base region and V* is the normal equilibrium vacancy concentration. χ is calculated on the basis of E-centre break-up within the emitter region, giving rise to an enhanced vacancy concentration V in the base region. The conclusions reached are that the E-centre model may indeed explain the push-out effect if most phosphorus diffuses in E-centre form and most of the E-centres break-up beyond the position of the kink in the phosphorus profile.

Lee (1974) compares values of χ obtained from push-out experiments with values obtained when analysing kinked phosphorus profiles. A close link between the two phenomena is established since the values of χ are almost equal in each case.

If E-centres give rise to an excess of vacancies upon breaking up, the number of vacancies within the part of the phosphorus profile which contains the E-centres must be above the equilibrium vacancy concentration

(Normal equilibrium number of vacancies + E centres). This suggests that, if the E-centre is the cause of the rapid gettering of gold on to substitutional, phosphorus - nearest - neighbour sites, a further enhancement of the gold solubility due to the vacancy excess may occur. The experimentally determined values of $\chi = V/V^*$, may be used to obtain a measure of this additional gold solubility enhancement. The extra effect on gold solubility will be incorporated into the previously developed model of solubility enhancement ; firstly however, some possible consequences of attributing the phosphorus-gold interaction to the same mechanism as the cause of the phosphorus kinked profile and base push-out effects will be considered.

For a situation in which gold is diffused to saturation in the phosphorus layer (i.e. in a gettering situation with an infinite source of gold) the E-centres would be largely replaced by phosphorus - gold ion-pairs, thus preventing formation of the kinked part of the phosphorus profile and also preventing the vacancy excess which leads to push-out. In chapter 8 results of an investigation of the effect of gold on these two phenomena are reported. It is shown that both push-out and the enhanced part of the phosphorus diffusion profile are inhibited when gold is simultaneously diffused with the phosphorus. This provides strong evidence in support of the model proposed here.

Having identified the diffusion of interstitial gold to be the rate limiting step in the gettering process, further implications of the very high rate of the reaction Au_i + E-centre $\rightarrow (Au\ P)_{Pair}$ within the phosphorus layer may be considered. Since substitutional gold is virtually immobile (especially when ion-paired as most of it will be) once gold has been gettered into the phosphorus layer it should remain there during any subsequent processing unless it is at a higher temperature than the gettering treatment. (In this case it is possible, though not certain, that thermal energy may be enough to break the Au-P pairs). Also, any gold interstitial atom which diffuses into the gettering layer will tend to be 'frozen-in' because of the rapidity of the gettering action and the subsequent immobility of the gettered gold. A phosphorus layer diffused into the back of a silicon wafer at the start of processing should have a two-fold effect, firstly it will getter any gold already present in the

wafer or any gold subsequently entering and secondly it will provide an effective barrier against contamination entering through the back of the wafer. This latter effect could conceivably be a serious source of contamination when, during normal photo-mechanical-etching stages, the oxide mask on the back of a wafer is inevitably removed when etching diffusion windows in the front. The ability of a phosphorus diffused layer to act as a barrier against gold diffusion has also been investigated. No penetration of the layer was detected until it had been saturated with gold (see chapter 8).

6.5.3 Prediction of gold solubility enhancement in a shallow diffused phosphorus layer in silicon

In order to predict the total gettering effect of a phosphorus diffused layer, the additional gold solubility enhancement due to the vacancy excess must be considered. The preceding theory considered an equilibrium situation only. Addition of a vacancy enhancement factor, it must be emphasised, involves non-equilibrium effects which would be very difficult to model accurately since little reliable data concerning point defects close to the silicon surface is available. The solubility enhancement already predicted (see figures 6.27 - 6.29) may be used as a measure of the minimum gettering effect of the phosphorus layer. The additional enhancement factor which is to be estimated here should indicate a more realistic gettering effect, although precipitation on lattice defects at very high phosphorus concentrations could well enhance the apparent gold solubility still further.

The additional solubility enhancement factor due to excess vacancies will be assumed to apply equally to all substitutional gold atoms whether they are ion paired or not. (This point is further discussed in appendix G). The total solubility enhancement at any point in a phosphorus diffused layer, with a given vacancy excess will be given by :

$$\text{Total solubility Enhancement} = \text{Equilibrium solubility enhancement} \times \frac{V}{V^*}$$

$$= \text{Equilibrium solubility enhancement} \times \chi \qquad 6.58$$

Vacancies exhibit acceptor behaviour in silicon (Kendall et al. (1969)) and it is arguable that if interstitial gold atoms are positively charged they will react preferentially with negatively charged vacancies (i.e. ionised acceptors). Provided, however, the excess vacancy charge states are in the same ratio as in equilibrium, which is a reasonable assumption, the enhancement effect will be the same. That is, if $V = V^- + V^x$ and $V^* = V^{*-} + V^{*x}$:

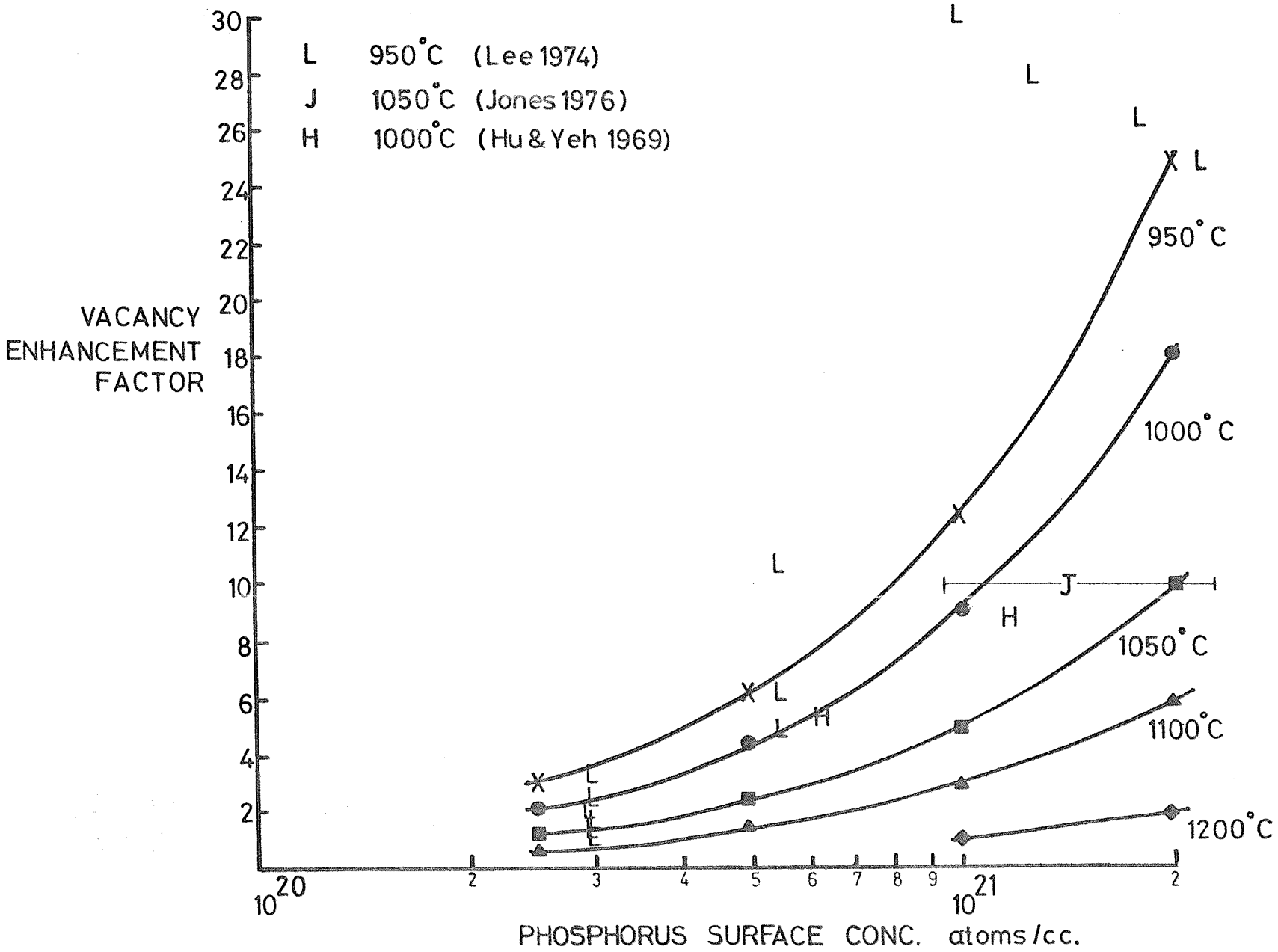
$$\chi = \frac{V^-}{V^{*-}} = \frac{V^x}{V^{*x}} = \frac{V}{V^*} \qquad 6.59$$

Values of χ are taken from data available in the literature as follows : Jones (1974 and 1976), push-out data ; Lee (1974), kinked profiles and push-out data ; Hu and Yeh (1969), push out data. (The values presented by Lee, from analysis of kinked phosphorus diffusion profiles, are the result of a review of a large amount of data available in the literature before 1974). Lee also measured the dependence of push-out on the phosphorus emitter surface concentration C_s. The result was a marked dependence of χ on C_s at 950°C. Results from Hu and Yeh for two difference surface concentrations at 1000°C indicate a similar dependence. Jones' push-out data were all obtained for phosphorus emitter surface concentrations which were at the solubility limit for phosphorus in silicon ($> 10^{21}$ atoms/cc), as were the values of χ calculated by Lee from kinked phosphorus profiles.

In order to infer values of χ over the concentration range of major interest here, (10^{20} — 2 x 10^{21} atoms/cc phorphorus) it is assumed that varies linearly with phosphorus surface concentration. Such a relationship was proposed by Hu and Yeh in an "approximate theory of emitter push" (1969) where it was assumed that the excess vacancies were generated by the climb of dislocations induced by the phosphorus diffusion. Although E-centre break-up is favoured here as the cause of the vacancy excess, Jones has shown that a similar dependence of χ on C_s could result from it.

Assuming that, at all temperatures in the range 950°C to 1200°C, the phosphorus solubility limit is 2 x 10^{21} atoms/cc (see profiles in Lee,

FIG. 6.32: Vacancy enhancement factor vs. phosphorus surface concentration.



Authors	Diffusion Conditions	Surface Concentrations after gettering/cc	Measured Au Solubility Enhancement	Required Value of χ	Actual Value of χ
Meek et al. (1975)	Au equilibrated 1100°C P diffused 1000°C	Au = 1.5×10^{19} P = 1×10^{21}	600	7	9
Chou et al. (1975)	P diffused 1100°C Au diffused 1100°C	P = 3.5×10^{20} Au = 7.1×10^{18}	150	10	2
Joshi et al. (1966)	P diffused 1150°C Au diffused 1000°C	P ≃ 10^{21} Au = 2×10^{18}	80	1	2
Brown (1976)	P diffused 1050°C	P ≃ 10^{21} Au 5×10^{18}-10^{19}	200-400	3-6	5
Adamic et al. (1964) 1.	P diffused 1100°C Au diffused (15 mins only) 1100°C	Au not saturated	-	-	-
" 2.	Au equilibrated 1000°C P diffused 1070°C	Au ≃ 10^{18} P ≃ 10^{21}	150	3	4
This work	P diffused 1050°C Au diffused 1000°C	P = 4×10^{20} Au = 2.9×10^{18}	95	3	2

Table 6.3. Gold solubility enhancement in shallow, high concentration phosphorus diffused layers.

1974), and using the linear relationship between C_s and χ, the vacancy enhancement factor as a function of phosphorus surface concentration has been plotted in figure 6.32 for temperatures of 950°C, 1000°C, 1050°C, 1100°C and 1200°C. Also shown are the measured values from push-out data, due to Lee (1974) at 950°C, Hu and Yeh (1969) at 1000°C and Jones (1976) at 1050°C. The values of χ predicted in this figure are well within a factor of two of the measured ones.

A brief comparison with gold solubility data available in the literature is given in table 6.3. These are all experiments in which a large amount of gold was diffused into a silicon wafer which either already had a phosphorus gettering layer at one surface or which was subjected to a gettering treatment after gold diffusion. The gold concentrations are those measured at the surface. The predicted, equilibrium enhancement effects are obtained from figures 6.27, 6.28 and 6.29. The predicted value of χ (figure 6.32) for the phosphorus surface concentration given in each reference is compared to the measured 'extra' enhancement . In most cases the agreement is well within a factor of two, which is better than the uncertainty in the values of intrinsic gold solubility from which the actual enhancement was calculated (see figure 6.2). The predicted enhancement for the data of Chou and Gibbons (1975) is somewhat low compared to their experimental values, however for the phosphorus diffusion conditions described, a surface concentration of only 3.5×10^{20} atoms/cc seems rather low compared to values obtained, under similar conditions, in gettering experiments carried out by this author (Unter, 1977). If the phosphorus surface concentration had been close to 10^{21} atoms/cc, the predicted and required values of χ would both $\simeq 3$.

6.5.4 Prediction of the gettering effect for removal of contamination

Using the predictions of the model described above, it is possible to predict the gettering effect of a given phosphorus diffused layer with reasonable accuracy. Important factors of the model which determine this gettering effect are :

(1) Gold diffuses into the gettering layer as an interstitial species.

(2) The reaction rate of interstitial gold with the gettering agent is very high - high enough to make D the rate limiting step.

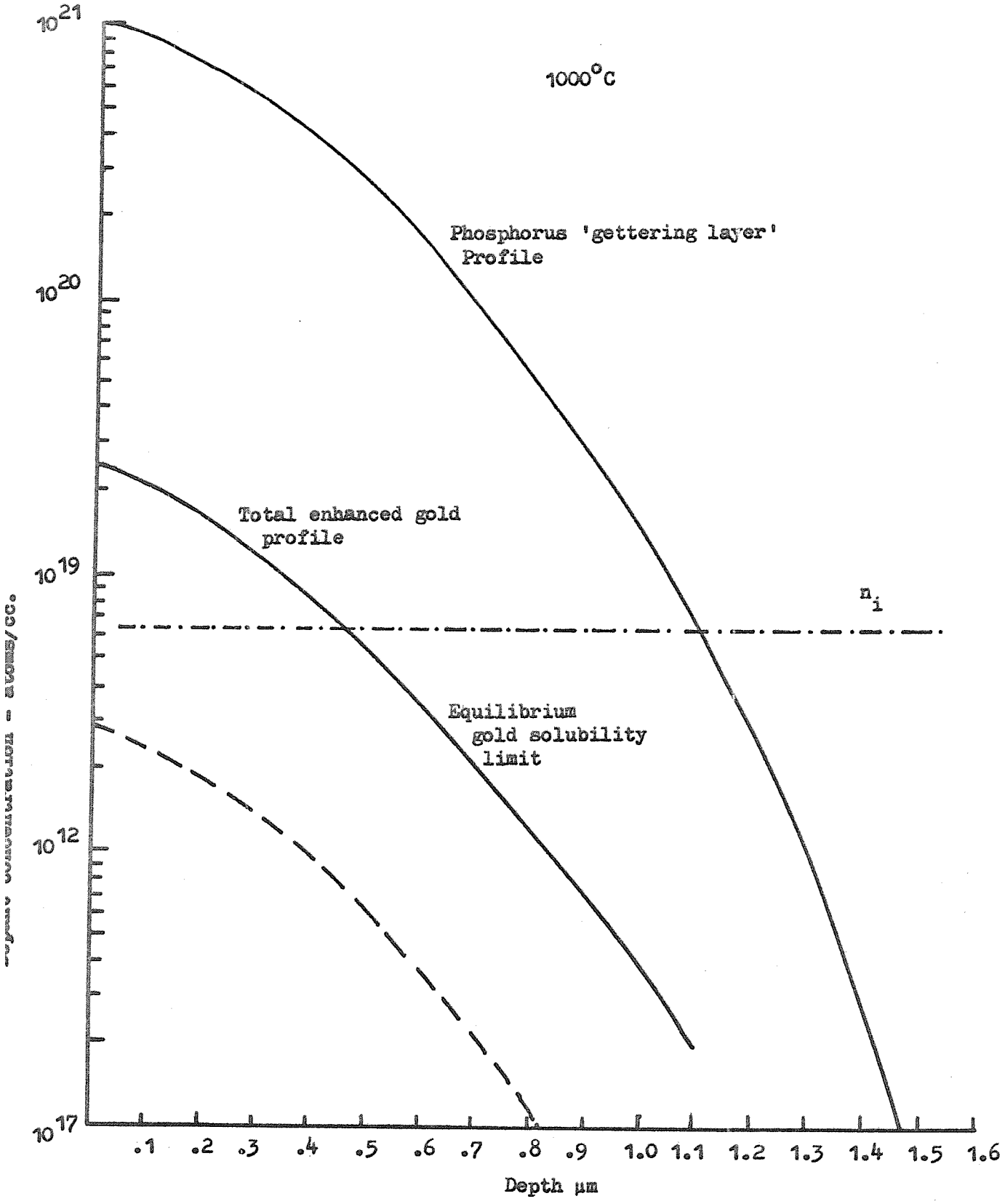


FIG. 6.33: Method for estimation of total gettering capacity of a phosphorus diffused layer (see text).

(3) Essentially all of the gold within the gettering layer is on substitutional lattice sites.

(4) The gettering effect in the phosphorus diffused layer is attributed to the reaction of gold atoms with E-centres which are present in significantly larger quantities than the normal equilibrium vacancy concentration.

From these effects and experimental data presented in chapter 8, it appears that the gettering layer behaves as an infinite sink for gold diffusing from the bulk of the wafer as an interstitial species. Data available in the literature (Lambert and Reese, 1968; Murarka, 1976) also indicates that this is so. If the actual capacity of the gettering layer for gold is estimated, the expected reduction in a known 'bulk' gold concentration can be calculated.

Estimation of the capacity of a phosphorus layer is described below and illustrated in figure 6.33. First, the phosphorus diffusion profile resulting from the gettering step is plotted. It may either be measured by one of the conventional profiling techniques (spreading resistance, see appendix A ; incremental sheet resistance, see appendix B ; radio-tracers, see for example Chou et al. 1975) or it may be assumed that the profile obeys the complementary error function (erfc) shape predicted by standard diffusion theory (Grove, 1967, p 41) for diffusion from an infinite source. In this case the sheet resistance, R_{SH}, of the layer (see appendix B) is measured, the depth of the profile, x_j, is then measured (normally by junction staining - see chapter 8) the product $R_{SH} \cdot x_j$ yields the average resistivity of the layer, $\bar{\rho}$. This value may be related to the surface concentration, C_s, of the phosphorus diffused layer using Irvins curves of C_s vs $\bar{\rho}$ for erfc distributions (Irvin, J.C., 1962). Once C_s is known, the depth profile may be plotted using the relationship

$$C(x,t) = C_s \operatorname{erfc} \frac{x}{2\sqrt{Dt}} \qquad 6.60$$

where x is the depth, t the diffusion time and D the diffusion coefficient of phosphorus at the chosen temperature. The saturation gold profile corresponding to this phosphorus diffusion may now be plotted using data

from figure 6.30, but with each gold concentration multiplied by the value of χ corresponding to the phosphorus surface concentration. (Only the gold profile in the region $N_p > n_i$ is of interest since below n_i no gold solubility enhancement is seen). In the example in figure 6.32, the phosphorus diffusion has a surface concentration of $C_s = 1 \times 10^{21}$ atoms/cc which, at the chosen temperature of 1000°C gives $\chi = 9$. A quick graphical integration of the saturation gold profile yields the gold capacity of the phosphorus layer, which in this example is 8.5×10^{14} atoms/cm^2 or, an average concentration of 6.5×10^{18} gold atoms/cc within the layer at saturation.

It is highly unlikely that the gettering layer remains an infinite sink for gold at all times until it is saturated, however experiments (see chapter 8) indicate that it certainly appears to be an infinite sink until gold has reached half of the saturation level. To make a conservative estimate of the gettering layer capacity, it may be assumed that it will act as an infinite sink until filled to 10%of its capacity. This means an average concentration of about 7×10^{17} atoms/cc or 8.5×10^{13} atoms/cm^2. This amount of gold distributed through a wafer of typical thickness 350 microns would give an average concentration of 3×10^{15} atoms/cc. Thus it appears that the phosphorus layer depicted in figure 6.32 is capable of completely removing all of the gold from a wafer which is grossly contaminated - even when its capacity is underestimated by a factor of ten. It should be pointed out that the example shown in figure 6.32 is not a typical "quick getter" profile, it is in fact quite deep and would have required >60 mins diffusion at 1000°C, however its apparent large capacity for gold should not be ignored.

In addition to estimating the capacity of the gettering layer, one must also consider the dynamics of the situation since the rate limiting step is the diffusion of interstitial gold. A simple model will serve to give a first order idea of the time required for gettering.

Fick's first law of diffusion states that the flux, F, of a diffusing species is given by :

$$F = -D \frac{dC}{dx} \qquad 6.61$$

where D is the diffusion coefficient and dC/dx is the concentration gradient. Consideration of the relationship between the flux and the rate of change of concentration with time (Grove, 1967, p 42) yields the well known 'diffusion equation' :

$$\frac{dC}{dt} = \frac{Dd^2C}{dx^2} \qquad 6.62$$

The solution to this equation for diffusion from an infinite source is the complementary error function profile expressed in equation 6.60. This dopant distribution will change from the erfc shape to a linear distribution when the diffusion length, $L = 2\sqrt{Dt}$, is greater than the thickness of the sample. At 1000°C, the diffusion coefficient, D_i, of interstitial gold is - 8×10^{-6} cm^2/second (Wilcox et al. 1964). In a 350μm thick silicon wafer, the time required for L to be greater than the wafer thickness is $\simeq$ 40 seconds. Thus, if a wafer originally contaminated with N_o gold atoms/cc is being gettered at 1000°C, after one minute the gold distribution will approximately be a linear fall from N_o to zero atoms/cc across the wafer thickness (assuming that the gettering layer is an infinite sink). This situation is shown schematically below :

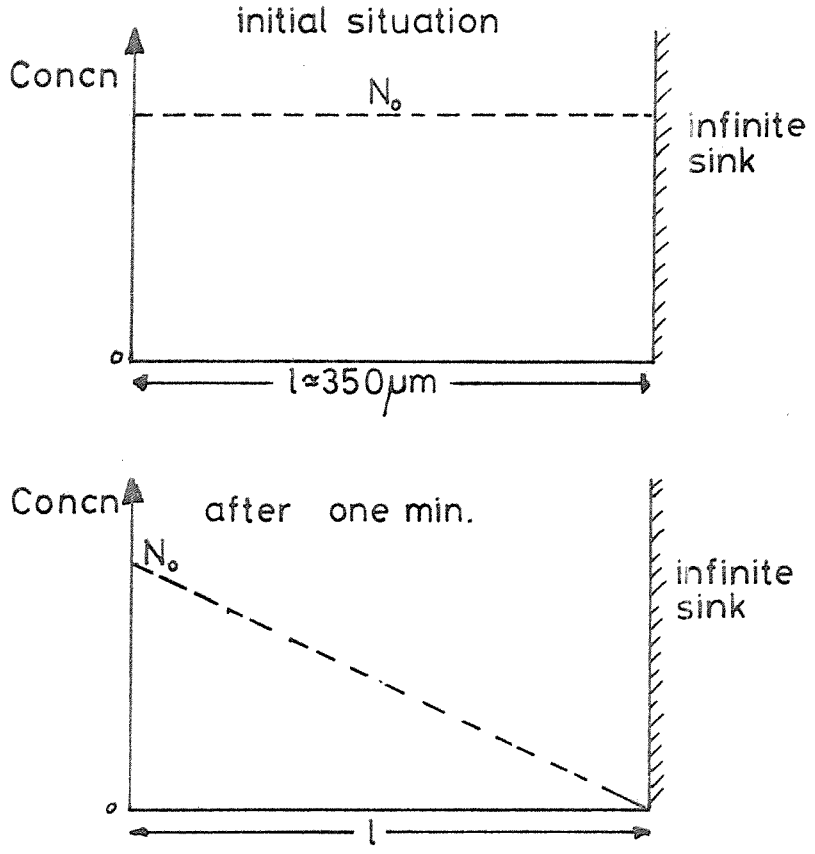


fig 6.34

The situation, as time progresses, may be modelled as shown below, with the time taken to reach the distribution shown in figure 6.34 neglected:

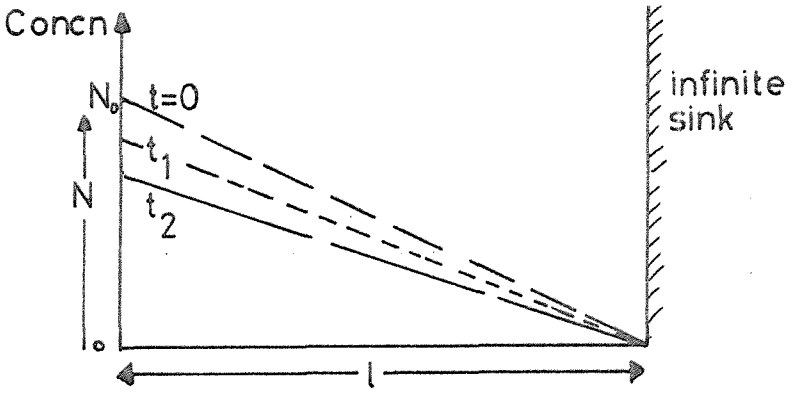


fig 6.35

The concentration gradient, dC/dx, at any time t is given by N/l. The flux into the infinite sink is given by the rate at which the area under the concentration distribution line shrinks. Equation 6.61 may be expressed :

$$F = \frac{-DdC}{dx} = \frac{d}{dt}\left[\tfrac{1}{2}\,l.N\right]$$

$$= \frac{d}{dt}\left[\frac{l^2}{2}.\frac{dC}{dx}\right] \qquad 6.63$$

this differential equation may be solved by letting y = dC/dx; then equation 6.63 becomes :

$$-Dy = \frac{l^2}{2}\frac{d}{dt}y \qquad 6.64$$

rearranging and integrating :

$$\int\frac{dy}{y} = -\int\frac{2Ddt}{l^2} \qquad 6.65$$

$$y = y_o \exp\left[\frac{-2Dt}{l^2}\right] = \frac{dC}{dx} \qquad 6.66$$

y_o is given by dC/dx at t = 0 (boundary condition). Therefore the flux F flowing into the infinite sink is :

$$F = -D \frac{N_o}{1} \exp\left[\frac{-2Dt}{1^2}\right] \qquad 6.67$$

The flux has a time constant of ($\frac{1^2}{2D}$) (for $F = k \exp -\frac{t}{\tau}$).

In the example being considered here, at 1000°C in a 350μm thick wafer the time constant is 77 seconds. If the gettering layer is applied to the back surface of the wafer and a reduction in gold contamination of 10^3 atoms/cc at the front surface is required (e.g. from 10^{15} to 10^{20} atoms/cc), then dC/dx (= N/1) must also decrease by 10^3. Therefore, from equation 6.66, the gettering time is approximately ten minutes if a couple of minutes are allowed for the 'transient' during which the distribution shown in figure 6.34 is achieved.

The above analysis does make one assumption which, if grossly incorrect, could modify the results considerably. It was assumed that the gettering layer provided an infinite sink at the very outset of the gettering process whereas in reality at t = 0 there is no phosphorus present! If the flux of phosphorus atoms entering the layer is comparable to or higher than the rate of arrival of gold, the analysis will still give a close idea of the time required to achieve the desired effect. The higher the original gold contamination level the more likely it is that the gettering effect will be limited by the phosphorus diffusion because, from equation 6.67, it can be seen that the gold flux entering the infinite sink is proportional to $N_o/1$. A simple calculation shows that the phosphorus layer quickly becomes an infinite sink for the highest gold contamination likely :

Assuming that the phosphorus diffusion profile is a complementary error function, integration of equation 6.60 yields the total number of atoms per unit area in the phosphorus layer after time t. (Grove, 1967).

$$Q(t) = \frac{2}{\sqrt{\pi}} \sqrt{Dt}\ C_s$$

$$\simeq \sqrt{Dt}\ C_s \qquad 6.68$$

The distribution may be closely approximated as a triangular diffusion profile of height C_s and base (= diffusion depth) $= 2\sqrt{Dt}$. (Note : Grove

gives as the approximation $Q(t) = \frac{1}{2}\sqrt{Dt}.C_s$, however $2/\sqrt{\pi} = 1.13$ so the approximation used here is better). The slope of the triangular profile may be used in equation 6.61 to give the flux of phosphorus atoms entering the silicon after time t :

$$F = \frac{-D\ C_s}{2\sqrt{Dt}} \qquad 6.69$$

If the phosphorus surface concentration is 10^{21} atoms/cc, then at 1000°C the flux is : $3.5 \times 10^{14}/\sqrt{t}$ atoms per second. After 5 minutes (300s) the phosphorus flux is 2×10^{13} atoms per second. Solution of equation 6.67 after the same time gives the flux of gold atoms into the gettering layer as : $\left[4.5 \times 10^{-6}.\ N_o\right]$ atoms/second for the 350μm thick wafer. Clearly even if the original gold concentration is at the intrinsic solubility limit ($\sim 10^{16}$ atoms/cc), the flux of phosphorus is much greater than that of gold after only five minutes. Thus, at the most, five minutes must be added to the original estimate of the gettering time. It is stressed that this situation may not be the same at other temperatures (see section 9.3).

6.5.5 External contamination during gettering

Contamination of the sample with gold from external sources during the gettering process must also be considered. This effect is particularly well demonstrated in measurements of gettering made by Murarka (1976). Neutron activation analysis was used to measure the initial gold contamination in wafers cut from ten different silicon crystals. In each case a wafer cut from an adjacent position was subjected to a short phosphorus gettering treatment at 1000°C. The total gold concentrations and the gold concentrations after removal of the phosphorus diffused layers were then measured. In nine of the ten samples, the total gold concentration had risen noticeably. The worst case was a rise from an original level of 1.4×10^{14} to a final level of 6.6×10^{14} gold atoms/cc, implying contamination up to 5×10^{14} atoms per cc in only thirty minutes. The extra contamination had, however, been confined to the phorphorus diffused layer. An average reduction in the bulk concentration by a factor of fifty was noted.

Three interesting points come out of these measurements :

(1) The proposal made earlier, (see section 6.52) that a phosphorus layer should make an effective barrier against external contamination (due to the high gold interstitial→gold-phosphorus pair reaction rate) is supported.

(2) The phosphorus gettering layer does behave as an infinite sink for gold at concentrations well above the solubility limit of gold in intrinsic silicon. (Murarka does not indicate what the thickness of his samples is. If it is assumed that they are more than average thickness, say 500μm, the gold concentration in the gettering layer of four of his samples is well over 10^{17} atoms/cc).

(3) Perhaps the most disturbing feature of these results is that they indicate gross contamination from the furnace in which the gettering was carried out.

It has been shown recently in the silicon device fabrication facility at Southampton University, and by other workers (Declerck et al. 1976), that cleaning furnaces with HCl gas provides an excellent means of reducing heavy metal contamination from them. Results briefly reported in chapter 8 of this thesis show that wafers which are thermally oxidised in an HCl-cleaned furnace are actually degraded by a subsequent phosphorus getter in a non-HCl cleaned furnace tube.

6.5.6 Prediction of gettering effect during deliberate gold doping

The situation where a device is to be deliberately gold doped but suffers from unwanted "gettering" due to the presence of high concentration phosphorus diffusions is much simpler to model. A gold source (infinite) is applied to one face of a wafer, the opposite face of which has a high concentration phosphorus diffusion in it. Assuming, once again, that the gettering process is gold interstitial diffusion rate limited and that the phosphorus layer acts as an infinite sink until it is almost saturated, equation 6.61 may be expressed quite simply as :

$$F = \frac{-DdC}{dx} = \frac{-D_i(C_i)}{1} \qquad 6.70$$

for $2\sqrt{Dt} \gg 1$ (as described in 6.5.4). C_i is the interstitial gold solubility

limit. (If an infinite gold source is applied to one face of the wafer then the concentration of the diffusing species will be at its solubility limit at that point).

G, the total amount of gold gettered in time t, may be obtained by integrating the flux over the time interval :

$$G = \int_0^t F\,dt = \frac{D_i C_i}{1} .t \qquad 6.71$$

One may therefore obtain the gettering rate as :

$$\frac{G}{t} = \frac{D_i C_i}{1} \quad \text{atoms/cm}^2\text{/second} \qquad 6.72$$

Using the method previously described for calculating the capacity of the phosphorus diffused layer, the time taken to saturate it may easily be calculated.

For example, at 1000°C, $D_i = 8 \times 10^{-6}$ cm^2/second and C_i (Wilcox et al., 1964) = 5×10^{14} atoms/cc. For a 350 micron thick wafer, the gettering rate is : 1.14×10^{11} atoms per cm^2 per second. The capacity of the example used in figure 6.33 is 8.5×10^{14} atoms/cm^2, which would take $\simeq$ 125 minutes to saturate.

6.5.7 Additional Comments on the "E-Centre" gettering model

The extra solubility enhancement of gold in phosphorus diffused layers has been explained here in terms of a model in which gold atoms react with excess vacancies while they are in E-centre form and thus become ion-paired with phosphorus atoms. This reaction has also been used, qualitatively, to explain why the gettering reactions : $Au_i + V \rightarrow Au_s$ and $Au_s^- + P^+ \rightarrow (AuP)$ are so fast. This latter effect is a new proposal, indeed an interaction between gold atoms and E-centres has never been suggested in the literature to the best of this author's knowledge. Meek and Seidel (1975) have suggested that a vacancy excess - attributed to the climb of phosphorus diffusion induced dislocations - may explain the additional gold solubility enhancement. No convincing qualitative results are given. This model, which is the only reasonable alternative proposal,

is rejected here for the following reasons :

(1) Extra gold solubility enhancement is observed, in this work, for phosphorus layers which do not have dislocations associated with them.

(2) The phosphorus kinked profile and base push-out effect, which may both be attributed to excess vacancies are also seen in samples which do not contain phosphorus induced dislocations. The only viable alternative model is one involving the diffusion of phosphorus as an E-centre (Jones 1974). In this work, the agent which causes kinked phosphorus profiles and the push-out effect is strongly linked with the agent responsible for gold interaction with the phosphorus layer (see chapter 8).

(3) The simple excess vacancy model (whether it be caused by dislocations or not) does not explain the rapidity of the gettering reaction. Indeed, it is generally considered that a vacancy in silicon has a very high diffusion coefficient, in which case one would not necessarily expect the excess vacancy effects to be confined to the phosphorus diffused region. On the other hand, while in the form of an E-centre, the excess vacancies are confined to the phosphorus diffused layer as is the gettering effect.

6.6 Conclusion

It has been shown, in this chapter, that the equilibrium solubility of gold in heavily doped silicon can be explained in terms of two effects: the Fermi level effect and ion-pairing. It has been demonstrated that measured values of enhanced gold solubility in p-type and n-type silicon can be accurately predicted by these theories using gold energy levels, energy level temperature dependences and degeneracy factors which were chosen as the most reliable from a reconsideration of a large amount of published data in chapter 4. No special assumptions involving a change in the basic properties of gold as an amphoteric, deep level impurity at high temperatures have been necessary to make the theory 'fit' the results.

The interaction of gold with gettering layers has been considered. The gettering reaction between gold and shallow phosphorus diffused layers has been considered in some detail. A new model for the gold-

phosphorus interaction has been proposed and has been shown to explain data available in the literature and in this work with very good accuracy considering the lack of experimental values for the vacancy enhancement factor upon which the predictions depend.

A detailed consideration of the dynamics of the phosphorus gettering process indicates that it is rate limited by the diffusion of interstitial gold at 1000°C since the phosphorus diffused layer acts as an infinite sink. Simple models enabling the prediction of gettering effects in situations where gold removal is desired and in situations where deliberate gold doping is desired have been developed. These should be applicable at all diffusion temperatures.

Proposals for the better use of phosphorus gettering have been made. Of particular importance is the ability of the phosphorus layer to halt contamination from external sources which enter through the back of the silicon wafer. It is recommended that phosphorus gettering should be carried out as early as possible in a device processing schedule. A sink for unwanted gold contamination entering the front of the wafer will be provided at all times as well as a barrier at the back of the wafer. The gettering model developed in 6.5 should enable a given gettering schedule to be optimised and to be predictable within reasonable limits. In connection with the problem of external contamination of wafers from furnace tubes, results reviewed here as well as those presented in chapter 8 indicate that some attention should be given to decontaminating furnace tubes - perhaps with the use of HCl gas.

7. THE EFFECT OF GOLD ON THE RESISTIVITY OF DOPED SILICON AT ROOM TEMPERATURE

7.1 Introduction

The effect of gold on the room temperature (298°K) resistivity of silicon already doped with a shallow donor or acceptor is to be considered. It will be seen that for gold to affect the resistivity of either type of silicon appreciably, its concentration must be comparable with, or greater than, the concentration of shallow dopant. The solubility limit of gold in silicon is generally between 10^{16} and 10^{17} atoms/cc and so concentration of shallow dopants up to 10^{17} atoms/cc will be considered. At room temperature silicon doped with 10^{17} atoms/cc of shallow dopant is non-degenerate and, therefore, the non-degenerate approximations to the Fermi function may be used for all of the cases to be considered (see section 6.4).

In the n-type silicon the Fermi-level is well above mid-gap (the actual position depending on the concentration of shallow donors N_d). If a small amount of gold is added ($N_{Au} < N_d$) the gold acceptor levels will be ionized immediately and the number of free majority carriers (electrons) in the conduction band reduced accordingly. When the gold concentration becomes comparable with the shallow donor concentration ($N_{Au} \simeq N_d$) there is a sharp increase in the resistivity because of compensation by the gold acceptor energy level. The Fermi-level moves towards mid-gap and further addition of gold results in holes becoming the majority carriers as more electrons taken from the valence band fill the gold acceptor levels ; this makes the silicon high-resistivity p-type. When gold is the dominant impurity ($N_{Au} \gg N_d$) the position of the Fermi-level is determined by the gold concentration and the relative occupancy of the gold levels. The gold-coupled shallow acceptor level, reported by Bruckner (1971) and discussed in section 4.4, is only significant when the gold concentration is greater than 4×10^{16} atoms/cc, and will have the effect of making the silicon lower resistivity p-type at gold concentrations greater than this value.

In p-type silicon, the process is similar to that described above,

due to interaction between gold atoms ionized as donors and the shallow acceptor levels. The increase in resistivity, however, is monotomic and more gradual than in n-type silicon. Conversion to n-type silicon does not occur because the gold donor is further from the middle of the silicon energy band gap than the gold acceptor and the Fermi-level stays between the valence band edge and the intrinsic position. Once again, the effect of the gold coupled shallow acceptor level, at high gold concentrations, is to reduce the resistivity and make the silicon more strongly p-type.

7.2 Measurements of resistivity as a function of gold concentration reported in the literature

There is a lack of reliable data on the effect of gold on the resistivity of silicon although several sets of measurements have been reported by : Boltaks, Kulikov, and Malkovich, 1960; Wilcox, La Chapelle and Forbes, 1964 ; Bullis and Streiter, 1968 ; Thurber, Lewis and Bullis, 1973. These authors have also carried out theoretical analyses but only Thurber et al. have had any success in predicting the experimental data. None of the theoretical treatments, with the possible exception of that due to Thurber et al., are adequate and it is for this reason that a reanalysis is carried out here using models for gold energy levels and degeneracies which have already been developed in chapter 4.

Experimental measurements of the effect of gold on resistivity in silicon are subject to similar problems to those already described for gold solubility measurements. The experiments basically involve measurement of the initial resistivity of a silicon wafer. This is followed by diffusion of gold - usually from an evaporated elemental source - at a high temperature after which the final resistivity and gold concentration are measured. The concentration of gold is obtained either by neutron activation analysis (Thurber et al. and Bullis and Streiter) or by use of radio-tracer gold (Boltaks, et al., and Wilcox et al). In all cases the surface 'tip-up' region must be excluded from the gold 'count' to ensure that the gold distribution throughout the sample is uniform. There is no evidence that Boltaks et al. removed any of the surface layers before measuring the resistivity or gold

concentration in their samples. Wilcox et al. and Bullis and Streiter only removed thin ($\sim$10μm) layers from their samples which, as shown in experimental results in the literature and in this work, is not enough to avoid the 'tip-up' region. Thurber et al. in a comprehensive set of experiments, removed 125μm thick layers from each face of their silicon wafers before measuring gold concentrations and resistivities (samples thus prepared were also used for Hall effect measurements described in chapter 4). These results are probably the most reliable and have been used, in this chapter, as the data to which the theoretical curves are to be fitted. As will be seen, there are still some problems in fitting the theoretical calculations to the measured values. The gold concentration data does, of course, include all electrically-active and electrically-inactive gold atoms. There still remains a need for a technique which yields, unambiguously, the electrically-active gold concentration

7.3 Theoretical calculations of resistivity as a function of gold concentration at room temperature

Theoretical calculations of resistivity as a function of gold concentration have been carried out, as mentioned above, by several authors. As the approximations and data used in each of the references cited are open to some doubt the calculations have been carried out in this work using more recent data for gold energy levels etc., and for various models of gold energy level variations with temperature and gold degeneracies. (See chapter 4).

The position of the Fermi-level is first calculated by solving the charge balance equation (at room temperature) and from this the number of free carriers and hence the resistivity are obtained.

The charge balance equation for this problem is :

$$n + N_a^- + N_A^- + N_G^- = p + N_d^+ + N_D^+ \qquad 7.1$$

where :

n = electron density

N_a^- = ionized shallow acceptor density

N_A^- = density of negatively ionized gold (filled acceptors)

N_G^- = density of negatively ionized gold-coupled shallow acceptors.

p = hole density

N_d^+ = ionized shallow donor density

N_D^+ = density of positively ionized gold (empty donors).

The electron density, n, is obtained from : (Nichols and Vernon 1966).

$$n = N_c . \exp\left[\frac{E_F - E_c}{kT}\right] \qquad 7.2$$

where the symbols have the same meanings as in previous chapters. If E_{Fi} is the Fermi energy in intrinsic material :

$$n = N_c . \exp\left[\frac{E_{Fi} - E_c}{kT}\right] \quad \exp\left[\frac{E_F - E_{Fi}}{kT}\right] \qquad 7.3$$

now, in intrinsic material, where n_i is the intrinsic electron (or hole) concentration :

$$n_i = N_c \exp\left[\frac{E_{Fi} - E_c}{kT}\right] \qquad 7.4$$

Therefore,

$$n = n_i \exp\left[\frac{E_F - E_{Fi}}{kT}\right] \qquad 7.5$$

Similarly, the hole concentration, p, is given by :

$$p = n_i \exp\left[\frac{E_{Fi} - E_F}{kT}\right] \qquad 7.6$$

In intrinsic material, $n = p = n_i$, thus E_{Fi} may be obtained by equating equation 7.2 with the corresponding expression for the hole concentration, p, in intrinsic material. This results in :

$$E_{Fi} = \left[\frac{E_c + E_v}{2}\right] + \frac{kT}{2} \ln \left(\frac{N_v}{N_c}\right) \qquad 7.7$$

N_v and N_c have been defined in equations 6.48 and 6.49, and $E_c + E_v = E_g$ the silicon energy band gap. Equation 7.7, may be rewritten in terms of the density of states effective masses for holes and electrons m_h and m_e :

$$E_{Fi} = \frac{E_g}{2} + \frac{3kT}{4} \ln \left(\frac{m_h}{m_e}\right) \qquad 7.8$$

The ionized shallow acceptor and donor densities respectively are :

$$N_a^- = \frac{N_a}{1 + g_a \exp\left[\frac{E_a - E_F}{kT}\right]} \qquad 7.9$$

$$N_d^+ = \frac{N_d}{1 + \frac{1}{g_d} \exp\left[\frac{E_F - E_d}{kT}\right]} \qquad 7.10$$

where : N_a = total number of shallow acceptors, g_a = degeneracy factor for shallow acceptors, N_d = total number of shallow donors and g_d = degeneracy factor for shallow donors (see equations 4.11, 4.12).

Similarly, the ionized gold-coupled shallow acceptor density is :

$$N_G = \frac{N_G}{1 + g_G \exp\left[\frac{E_G - E_F}{kT}\right]} \qquad 7.11$$

The densities of ionized gold acceptors and donors may not be obtained from the simple Fermi probability function in the way that N_a^- and N_d^+ have. A gold atom may exist in one of three charge states but its existence as, say an ionized donor, precludes it from being in either of the other states at the same time, viz., as a neutral atom or as an ionized acceptor. Expressions for the densities of ionized gold acceptors and donors have been derived in this work, from the Shockley-Last (1957) statistics of the charge distribution on a localized flaw in a semiconductor (see appendix E) and are :

$$N_A^- = \frac{N_{Au}}{1 + g_A \exp\left[\left[\frac{E_A - E_F}{kT}\right]\right]\left[1 + g_D \exp\left[\frac{E_D - E_F}{kT}\right]\right]} \qquad 7.12$$

and

$$N_D^+ = \frac{N_{Au}}{1 + \left[g_D^{-1} \exp\left[\frac{E_F - E_D}{kT}\right]\right]\left[1 + g_A^{-1} \exp\left[\frac{E_F - E_A}{kT}\right]\right]} \qquad 7.13$$

where : N_{Au} = total gold concentration, E_A = gold acceptor energy level, g_A= gold acceptor degeneracy, E_D = gold donor energy level and g_D = gold donor degeneracy.

The charge balance equation (7.1) is solved by iteration with a digital computer to obtain a self consistent value for the Fermi-level, E_F, to a specified accuracy. The values of n and p (7.5 and 7.6) are obtained and used to calculate the resistivity (ρ) from :

$$\rho = \left(q.n.\mu_n + q.p.\mu_p\right)^{-1} \qquad 7.14$$

where : q = the electronic charge, μ_n = electron mobility and μ_p = hole mobility.

The mobilities μ_n and μ_p are calculated within the computer program by evaluating the contributions of the two dominant scattering mechanisms which are acting on the carriers. These are lattice scattering due to phonons (thermal vibrations) which interrupt the periodicity of the lattice and thereby impede the motion of carriers, and ionized impurity scattering whereby an ionized donor or acceptor atom may deflect a moving electron or hole with an electrostatic interaction. There is a third scattering mechanism relating to the interactions of carriers with neutral impurity atoms which is neglected here since most of the impurities are ionized. The scattering contributions are summed reciprocally to give :

$$\frac{1}{\mu_{total}} = \frac{1}{\mu_{lattice}} + \frac{1}{\mu_{\substack{Ionized \\ impurity}}} \qquad 7.15$$

The values used in this equation are discussed in section 7.32.

The calculations are carried out for a given n-type or p-type doping concentration and a series of gold concentrations ranging from 10^{13} atoms/cc to 10^{17} atoms/cc. The resulting curves of resistivity as a function of

gold concentration for various starting materials are plotted out by the computer "Calcomp" peripheral.

The program (see appendix H) used to perform this calculation was written in such a way as to allow variation of almost all of the parameters involved, e.g. energy levels, degeneracies, silicon energy band gap, temperature, etc.

The value of n_i, the intrinsic carrier concentration, is important in determining the value of the Fermi-level during the iterative part of the program and subsequently the number of free carriers used to obtain the resistivity (see equations 7.5, 7.6 and 7.14). The value of E_{Fi} obtained from equation 7.8 also makes a significant contribution to the final value of the Fermi level and depends on the energy band gap of silicon and the density of states effective meases of electrons and holes (m_e and m_h). These parameters and the methods used to calculate mobility will be discussed in more detail since the values used have an important bearing on the final results.

7.3.1 The intrinsic concentration, n_i

Using equation 7.2 and the corresponding equation relating n_i to N_v and the position of E_{Fi} relative to E_v, one may obtain the relationship :

$$n_i = (N_v N_c)^{\frac{1}{2}} \exp\left[\frac{-E_G}{2kT}\right] \qquad 7.16$$

At room temperature (298°K), with an energy gap of 1.24eV (Bludau et al, 1974), this expression gives $n_i = 5.4 \times 10^9$ cm^{-3}. The measured value at 298°K reported by Morin and Maita (1954) is 1.18×10^{10} cm^{-3} and is the generally accepted value (Grove, 1967). However, as pointed out by Barber (1966) this value is probably unreliable as the measurements were made on silicon which was not very pure by today's standards, in fact it was suspected to contain enough oxygen to form p-n junctions. Barber (1966) considers more reliable data for n_i above 350°K attributed to Putley and Mitchell (1958) and extrapolates their measurements to lower temperatures by considering the temperature dependence of electron and hole density of states effective masses. The uncertainty in Barber's

values is estimated to be less than 15% and he gives $n_i = 9.6 \times 10^9$ cm^{-3} at 298°K. This value, which is used here, is justified by the calculated values of maximum resistivity in n-type gold doped silicon which depend strongly on it and which agree well with the measurements of Thurber et al.(1973).

7.3.2 Mobility

The mobility of a carrier (electron or hole) is derived from the mean time between collisions which scatter the carrier. This may be visualised with a simple explanation : under the influence of an electric field , ε, the acceleration of, say, an electron, is given by Newton's second law of motion :

$$a = \frac{q\varepsilon}{m^*} \qquad 7.17$$

where m* is the conductivity effective mass of the electron. If the mean time interval between collisions is $\langle\tau\rangle$, then the average drift velocity of the electrons is :

$$v_{drift} = \frac{q\varepsilon}{2m^*}\langle\tau\rangle \qquad 7.18$$

and the quantity $(q\langle\tau\rangle/2m^*)$ is defined as the mobility of electrons, μ_n .†

The time interval between collisions is determined by the mechanisms already mentioned, the probability of a collision taking place in unit time, $\frac{1}{\tau}$, being the sum of the probabilities due to the scattering mechanisms; this leads to equation 7.15 :

$$\frac{1}{\tau} = \frac{1}{\tau_{lattice}} + \frac{1}{\tau_{\substack{ionized\\impurity}}} \qquad \frac{1}{\mu} = \frac{1}{\mu_L} + \frac{1}{\mu_I}$$

Footnote. A more rigorous analysis, such as that given by Nichols and Vernon (1966)p112, yields the result that $v_{drift} = qE/2m\left(\frac{\langle\tau^2\rangle}{\langle\tau\rangle}\right)$ where $\langle\tau$ and $\langle\tau\rangle$ are the mean square time and mean time of flight between collisions. It is shown that $\langle\tau^2\rangle = 2\langle\tau\rangle$, which results in $\mu_n = q\langle\tau\rangle/m^*$.

The values of μ_L for electrons and holes used in this work are taken from well accepted measurements of drift mobilities in very low-doped silicon (where μ_I is negligible). These measurements, made by Ludwig and Watters (1956), yield the following room temperature values :

$$\mu_{nL} = 1350 \pm 100 \text{ cm}^2/\text{volt.sec.}$$

and $$\mu_{pL} = 480 \pm 15 \text{ cm}^2/\text{volt.sec.}$$

The ionized impurity scattering contribution to mobility must be calculated for each solution of the charge balance equation however, since the number of ionized impurities, N_I, is the sum :

$$N_I = N_a^- + N_A^- + N_G^- + N_d^+ + N_D^+ \qquad 7.19$$

which is different for each value of N_d, N_a or N_{Au},

The problem of the scattering of carriers in the Coulomb field of an ionized impurity atom has been discussed in terms of classical mechanics and wave mechanics - however both methods give the same result (R.A.Smith, 1964, p148). A simple model of the Coulomb field of the ionized impurity does not adequately describe the scattering of carriers in a real crystal. This is due to the modification of the Coulomb field of one ionized impurity centre by neighbouring ones and, in addition, the presence of large numbers of free carriers which may "screen" the ionized impurity. Two approaches to the problem of modifying the Coulomb field have been made ; these are due to Conwell and Weisskopf (1950) and Brooks (1955) and Herring (1954). These latter two authors arrived, independently, at the same result. The Conwell-Weisskopf formula is obtained by considering that each ionized impurity scatters independently, with its "scattering cross-section" arbitrarily cut off at a distance from the ion which is equal to half of the distance between neighbouring impurities, i.e. if r_m = the effective radius beyond which the ion ceases to be effective and N_I is the density of impurities:

$$(2r_m)^{-3} = N_I \qquad 7.20$$

The approach made by Brooks and Herring was a quantum mechanical one which takes account both of the screening of the Coulomb field of the ionized impurity centres by the free holes and electrons and the modification of the average carrier density in the neighbourhood of a charged centre. The Brooks-Herring formula yields similar results to the

Conwell-Weisskopf formula when the density of conduction electrons in n-type silicon (or holes in p-type silicon) is similar to the number of ionized impurities. However, the two formulae give quite different results when the conduction electron (or hole) density is appreciably less than the ionized impurity density, $n << N_I$. In this case, the mobility given by the Brooks-Herring formula is less than that given by the Conwell-Weisskopf formula for the same value of N_I - this is because n does not appear explicitly in the Conwell-Weisskopf formula as it is assumed to be equal to N_I. The reason for the decrease in mobility lies in the fact that the "screening" effect of the free carriers is reduced when their density, n, is reduced.

In the problem to be solved in this work, when the gold concentration becomes larger than the shallow dopant concentration, the total number of ionized impurities is much greater than the electron (or hole) concentration since the electrons (or holes) distribute themselves between the gold and shallow levels rather than occupying the conduction (or valence) bands. This fulfils the condition in the Brooks-Herring formula where $N_I >> n$ and for this reason mobilities have been calculated using the Brooks-Herring expression :

$$\mu_I = \frac{2^{7/2} (4\pi\varepsilon)^2 (kT)^{3/2}}{\pi^{3/2} (m^*)^{\frac{1}{2}} q^3 N_I \left[\ln\left[1 + b\right] - \frac{b}{1+b}\right]} \qquad 7.21$$

where:

$$b = \frac{6}{\pi} \cdot \frac{\varepsilon m^* (kT)^2}{n \hbar^2 q^2}$$

ε = dielectric constant of silicon

k = Boltzmann's constant

T = temperature, ^oK

m^* = conductivity effective mass of carriers (electrons or holes)

N_I = number of ionized impurities per unit volume (eqn. 7.19)

$\hbar$ = Planck's constant $\div 2\pi$

q = electronic charge

n = number of free carriers per unit volume (electrons or holes)

7.3.3 Effective Mass

The effective mass of an electron (or hole) arises from the quantum theory which describes the motion of electrons in a periodic lattice. Its definition comes from calculation of the rate of change of crystal momentum of an electron in a periodic lattice ; it is the proportionality factor linking force and acceleration (as in Newton's law of motion : force = mass x acceleration). For the 'ideal' one dimensional case, the effective mass, m^*, is given by the well known expression (see for example, McKelvey, 1966 p219) :

$$m^* = \hbar^2 \left[\frac{d^2E}{dk^2}\right]^{-1} \qquad 7.22$$

where E is the energy of the electron.

In a real situation this equation may be extended to three dimensions quite easily (McKelvey, p 237) and the effective mass becomes a tensor which takes account of the fact that the E(k) relationship (k = wave number, which is closely linked to crystal momentum since h.k has dimensions of momentum) may not be the same in all directions. A different effective mass is required to describe motion in each direction.

A detailed consideration of the effective mass tensor in calculating the conductivity of a semiconductor (see equation 7.18) is given by McKelvey, p 302 and a 'conductivity effective mass' for isotropic conductivity is defined. When occupancies of the valence and conduction bands (equation 7.2) are considered, however, summation of the directional components of effective mass to give a value which describes the density of states of the band (equations 6.48 and 6.49) yields the 'density of states equivalent effective mass' which is not the same as the conductivity effective mass.

This difference has not been appreciated by many authors. In the calculations carried out here, density of states effective masses of $m_e = 1.18m.$ and $m_h = 0.81\ m_o$ (Barber, 1968) are used to calculate the position of E_{Fi} in equation 7.8 ; conductivity effective masses of $m_e = 0.27\ m_o$ and $m_h = 0.38\ m.$ (Bullis, 1966 and Barber, 1968) are employed

in the solution of the Brooks-Herring expression - equation 7.21.

7.4 Results of the Calculations

Solutions of the charge balance equation, with the parameters described above, were obtained for a large number of different gold energy level positions. For clarity (and brevity) only a small number of examples which illustrate the important aspects of the calculations are included here. In all cases the calculated values are compared with the experimental data kindly provided by Dr. W. R. Thurber (Thurber et al. 1973). The silicon starting resistivities in the calculations are the same as those used in the experimental data so that a direct comparison between them may be made.

The precise positions of the shallow energy levels and their degeneracies have little effect on the calculated curves since the Fermi level is almost always far enough away from them to ensure that they are essentially all ionized. The values used were : shallow acceptor level 0.045 eV above E_v, $g_a = 4.0$; shallow donor level 0.044 eV below E_c, $g_d = 0.5$. The gold-coupled shallow acceptor was assigned an energy level of 0.033 eV above E_v and a degeneracy of 4.0 ; its density was varied as $(N_{Au})^3$ with a concentration of 4.5×10^{15}/cc at a gold concentration of 1×10^{17} atoms/cc (Thurber et al. 1973).

The calculations were carried out for the five energy level temperature dependence models suggested in chapter 4, viz:

both energy levels fixed to E_c, Model C
both energy levels fixed to E_v, Model V
each energy level fixed to opposite band edge, Model O
each energy level fixed to nearest band edge, Model N
both energy levels vary in proportion to $E_g(T)$, Model P

Models N and O gave poor results overall and are not considered further. Each of the other models was assessed using a number of sets of degeneracy factors for the gold levels including those suggested by various authors in the literature :

Degeneracy Set	g_{A-}	g_{D-}	Comment
1	1	1	No degeneracy
2	0.125	16	Bruckner (1971)
3	0.125	8	Thurber et al. (1973)
4	4	1.5	Trivalent:Brown et al(1975)
5	6	2.5	Trivalent:including split off band
6	0.666	0.25	Monovalent:Bullis (1966)
7	1.333	0.75	Monovalent including split off band

Three pairs of gold energy levels were used for each model. These were chosen to represent the range of values reported in the literature with the centre value that is given by the measurements of Engstrom and Grimmeiss (1975) for the gold acceptor level and Wong and Penchina (1975) for the gold donor level (asterisked in the list). For each model the absolute zero values were corrected to 300°K values using the appropriate temperature variation (see chapter 4) :

Model	E_c-E_A(eV)	E_D-E_v(eV)
C1	0.53	0.275
C2*	0.553	0.284
C3	0.56	0.3
V1	0.45	0.343
V2*	0.48	0.35
V3	0.49	0.365
P1	0.49	0.32
P2*	0.52	0.33
P3	0.53	0.35

In the calculations the starting material resistivities listed below were employed; the symbols correspond to the measured data of Thurber et al.

Conductivity Type	Resistivity Ωcm	Symbol
n	0.3	○
n	1.0	△
n	5.3	+
n	75	×
n	380	◇
n	2300	⊕
p	0.076	⧖
p	0.53	⧗
p	1.1	Y
p	11	⊠
p	20	✳
p	93	⧖
p	280	|
p	1100	✡

7.4.1 N-type silicon starting material results

In all of the n-type theoretical curves the sharp rise in resistivity is predicted at $N_{Au} = N_d$, however in every case the measured value of N_{Au} at the sharp rise is $N_{Au} \simeq 2N_d$. Variation in energy level position or degeneracy has no effect on this and it remains the biggest uncertainty in achieving a satisfactory fit between the predicted and measured data. There are three possible explanations of this anomaly : a systematic error has been made in the neutron activation measurements of gold concentration ; the measured gold concentration contains a significant proportion of non electrically active gold atoms - such as gold interstitials; or, as suggested by Kendall and De Vries (1969), two substitutional gold atoms on adjacent lattice sites are required to provide one electrically active centre. The first possibility is unlikely especially as the 'factor of two' anomaly has been observed by other workers (Bullis and Strieter, 1968). The second possibility is the most likely in view of the lack of data on the relative concentrations of Au_s and Au_i; although the Kendall et al. suggestion cannot be wholly discounted. The success of gold diffusion and solubility theories which assume a single substitutional gold atom per active site must support the interstitial gold explanation. For this reason the best fit calculations are taken to be those which parallel the measured

values rather than those which do fit at high resistivities and not at low ones (as in the calculations of Thurber et al.).

The effect of varying the degeneracy factor for given gold energy level positions is illustrated in figures 7.1, 7.2 and 7.3 for 1 cm n-type silicon and models C2, V2 and P2. Clearly variation of degeneracy or energy level position does not affect the position of the sharp rise in resistivity or the maximum value. This latter point is not surprising since examination of equation 7.1 shows that the maximum resistivity occurs when the silicon is effectively intrinsic, i.e. when $E_F = E_{Fi}$, and depends on n_i and E_{Fi} only. Bullis (1966) incorrectly states that the maximum value of resistivity (as the n-type silicon→ p-type) is affected by the degeneracies and positions of the gold levels.

The best overall fit between theoretical and measured data - allowing a factor of two at all gold concentrations - occurs for model C2 with degeneracy set 5 (model C2-5), i.e. both gold energy levels fixed to E_c (as measured by Engstrom and Grimeiss, 1975; and Wong and Penchina, 1975) coupled with degeneracy factors obtained in this work where it is assumed that gold is trivalent in silicon with the 'split-off' valance band included in assessing the degeneracies of the various occupancy conditions (g_A = 6, g_D 2.5). The next best fit occurs for model C2 with degeneracy set 4 (model C2-4) (Brown et al. 1975). A reasonable but not as good fit occurs for the models in which gold is monovalent with energy levels fixed to E_v, i.e. model V2-6. Model C1-4 gives an almost identical result to the best fit curves obtained with C2-5, however from the discussions of chapter 4, the energy level positions in model C2 are considered to be more accurate and hence C2-5 is the favoured model. Calculations for the monovalent model in which the gold energy levels vary $\alpha E_g(T)$ do not give good fits (i.e. models P1, P2 and P3-6).

It is apparent, from figures 7.1 - 7.3, that many of the sets of degeneracies give completely wrongly shaped curves for some of the models. It is particularly interesting that sets 2 and 3 which were used by Thurber et al. (1973) to give good predictions in their calculations do not give correctly shaped curves for any of the models here. The reason for this is not known since it has proved impossible to reproduce Thurber

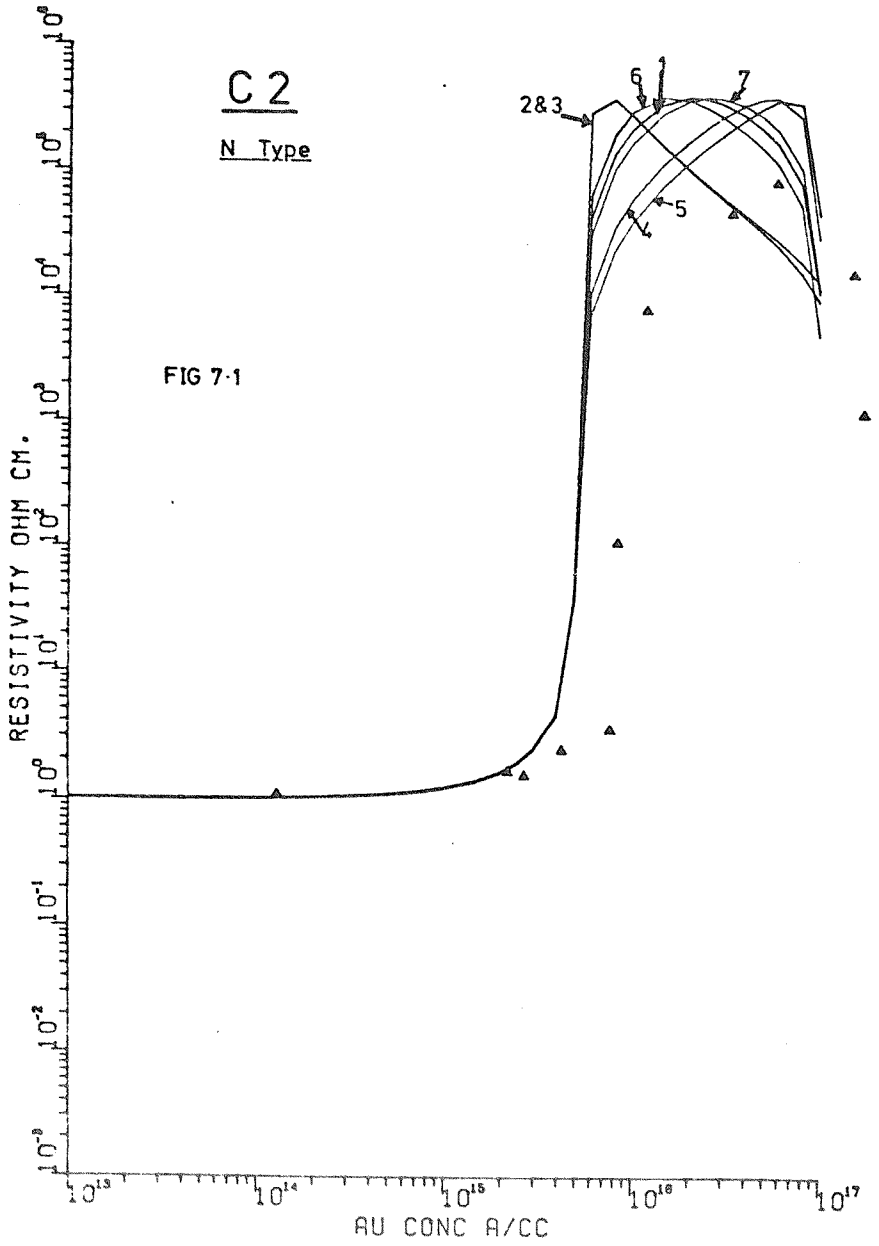

C 2
N Type
2&3
6
1
7
5
4
RESISTIVITY OHM CM.
AU CONC A/CC

FIG 7·1

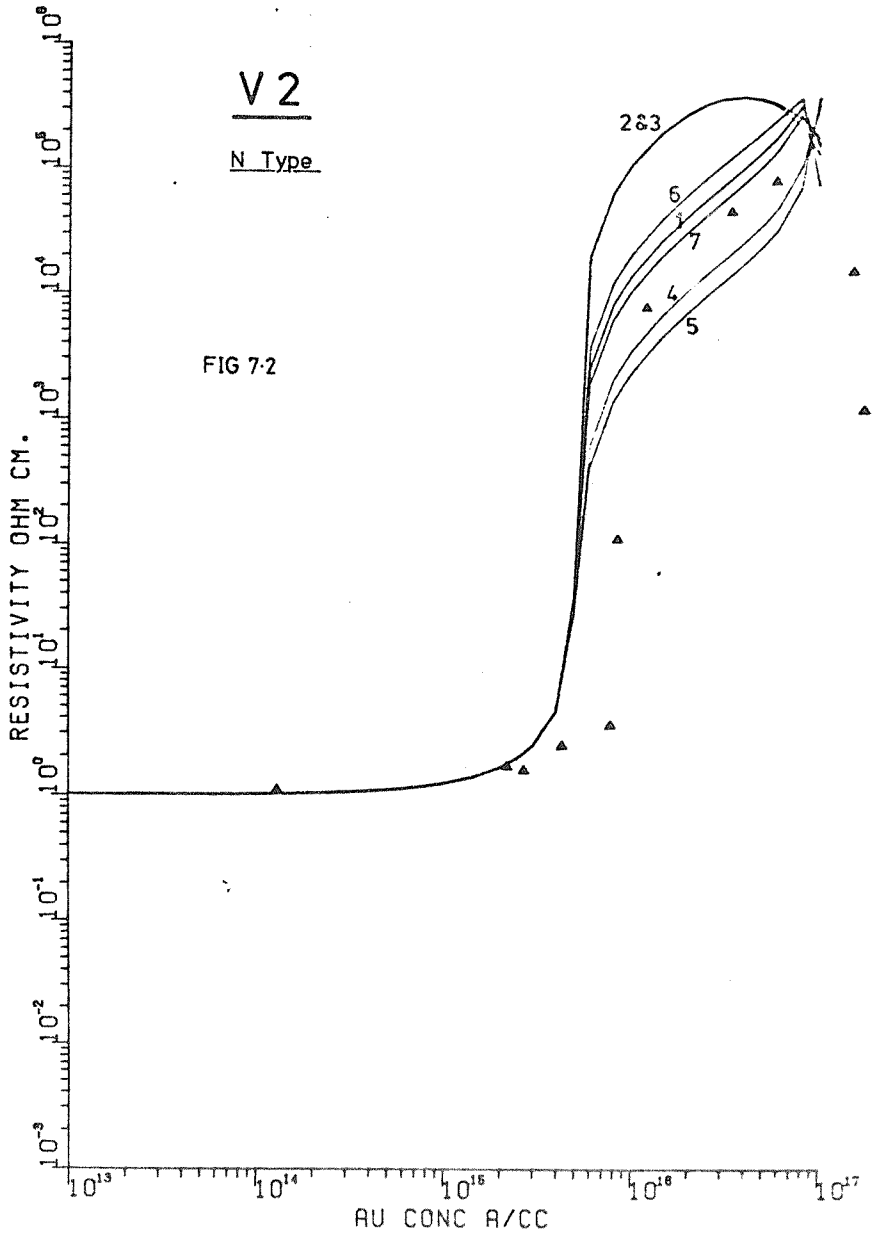

V 2
N Type
2&3
6
1
7
4
5
RESISTIVITY OHM CM.
AU CONC A/CC

FIG 7·2

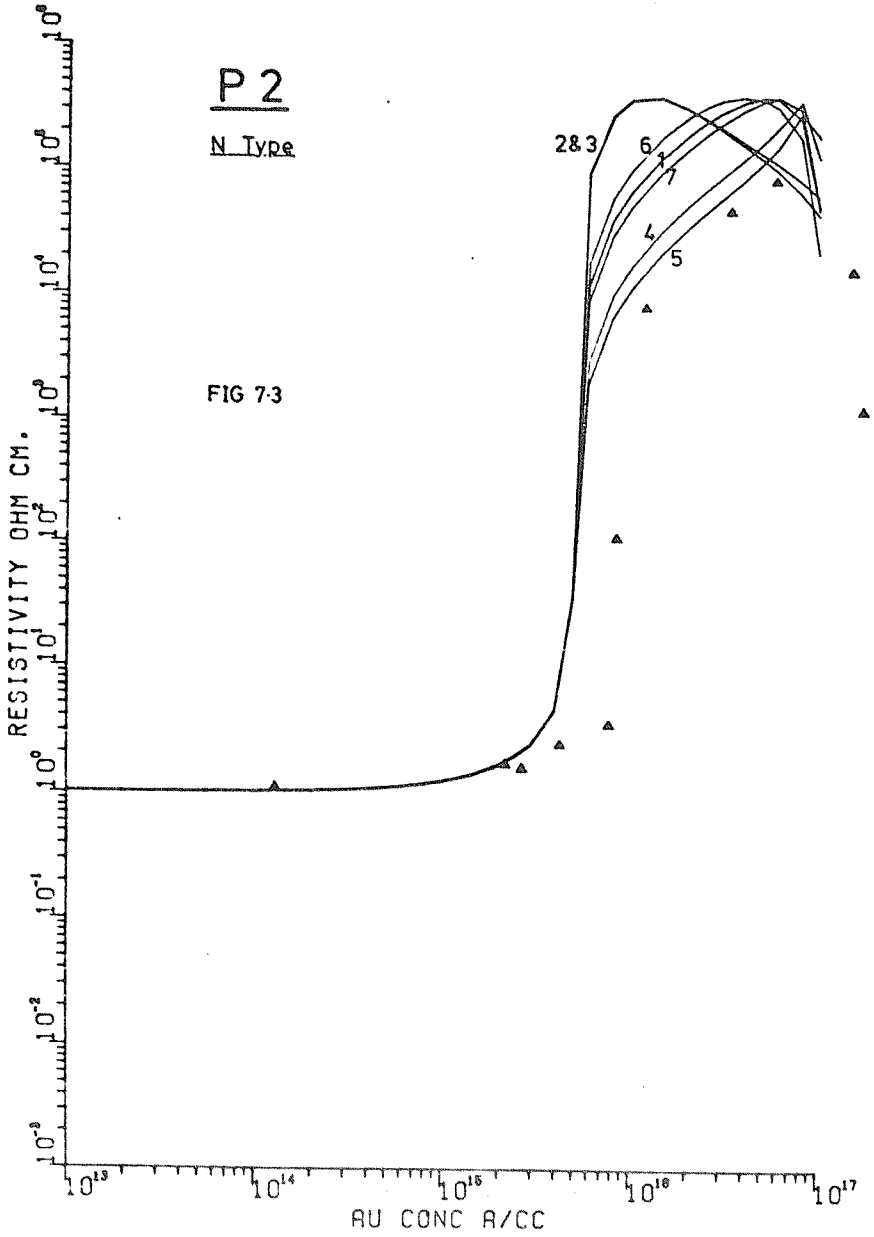
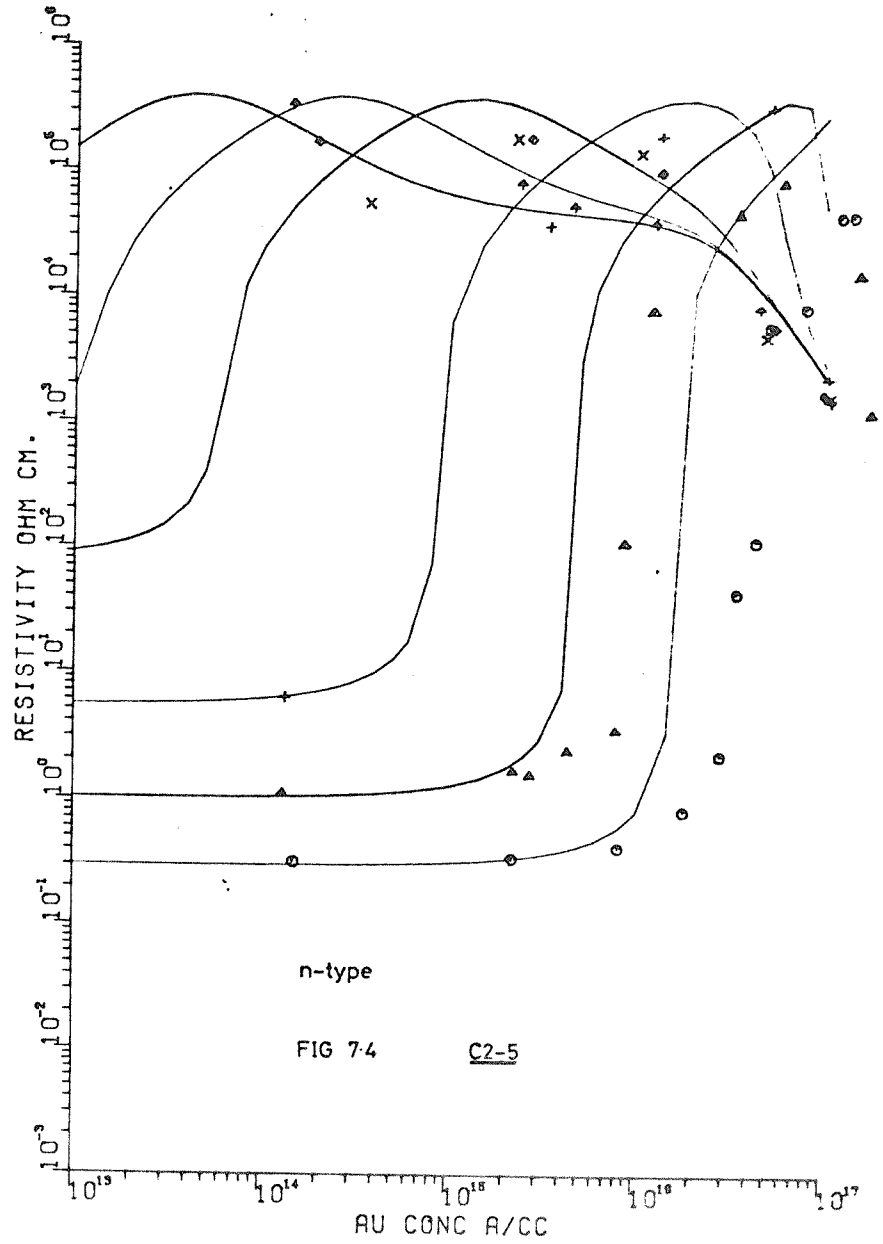
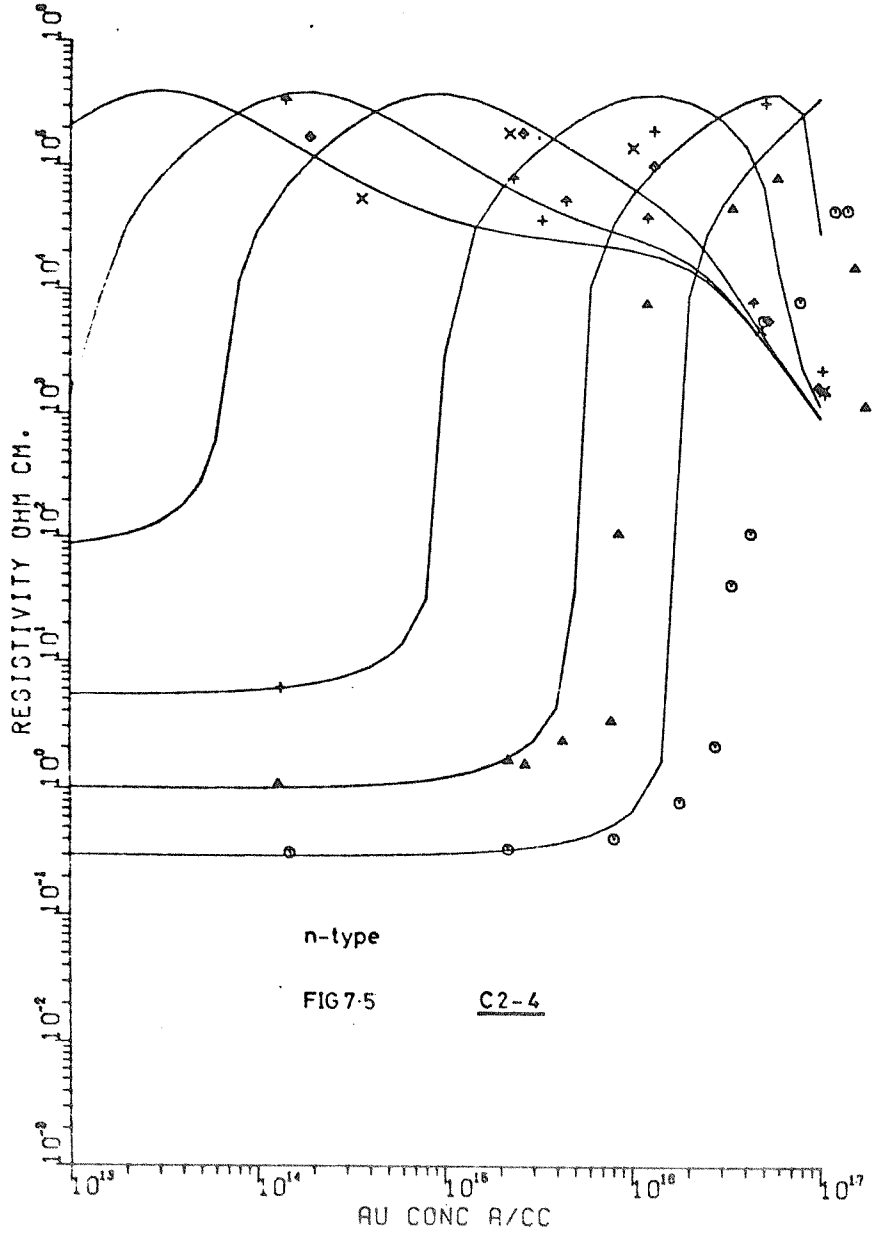


FIG 7·3



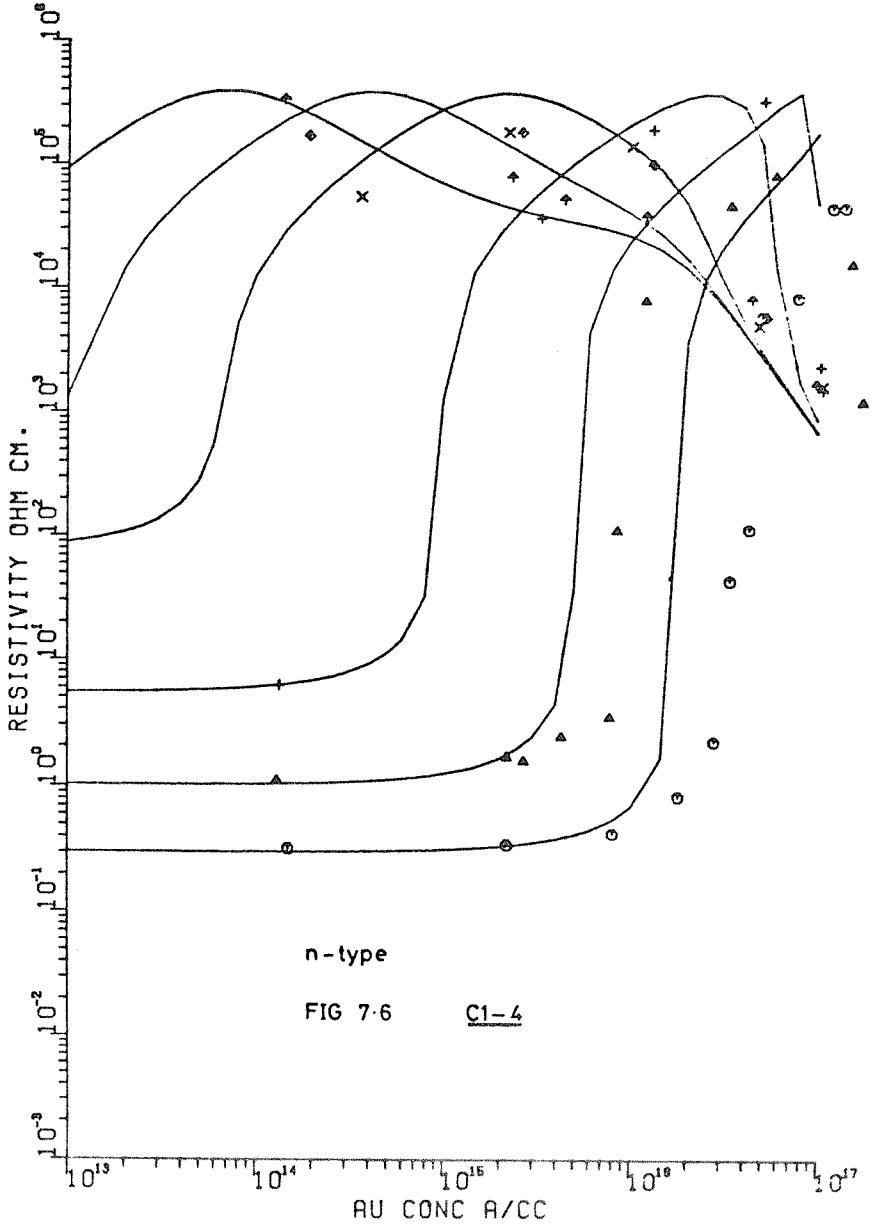
n-type

FIG 7·4 C2-5



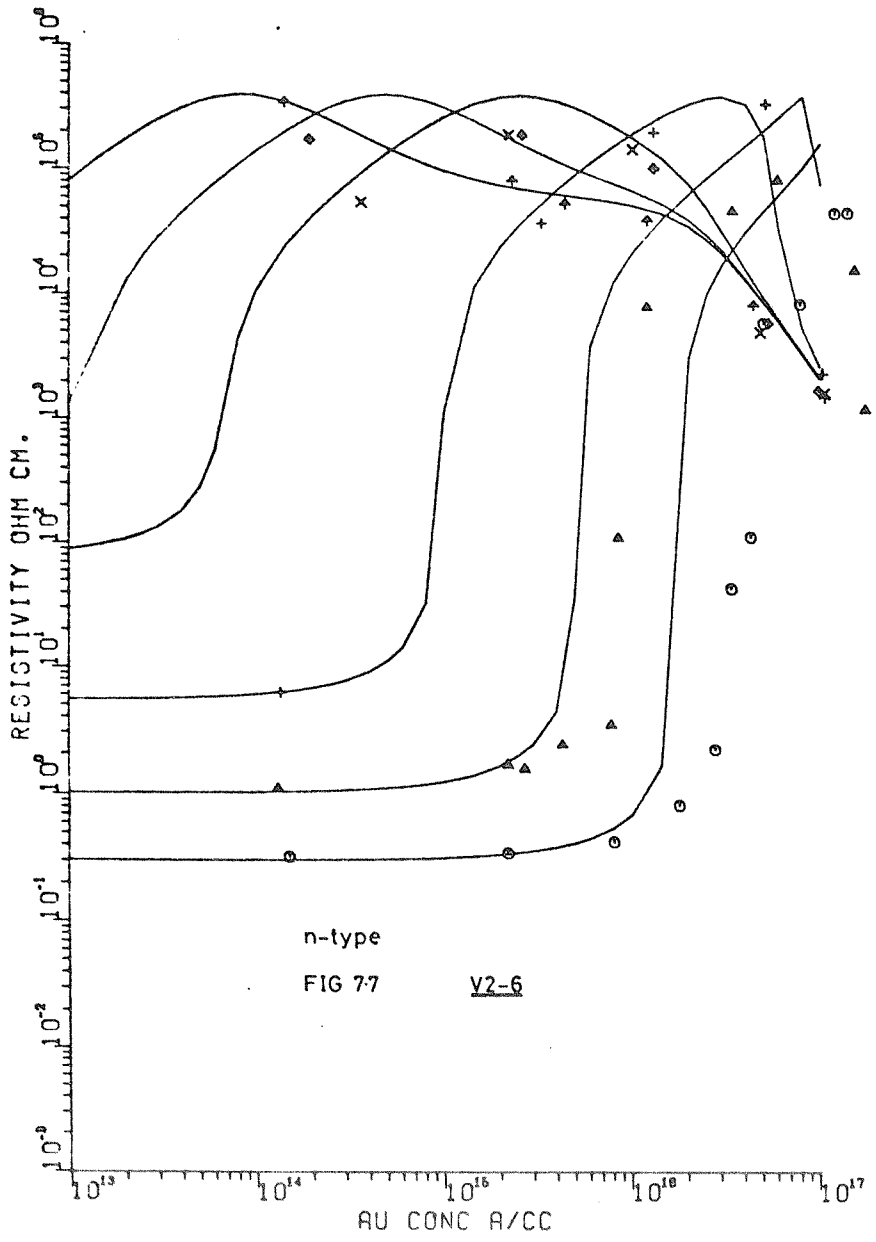
n-type

FIG 7·5 C2-4



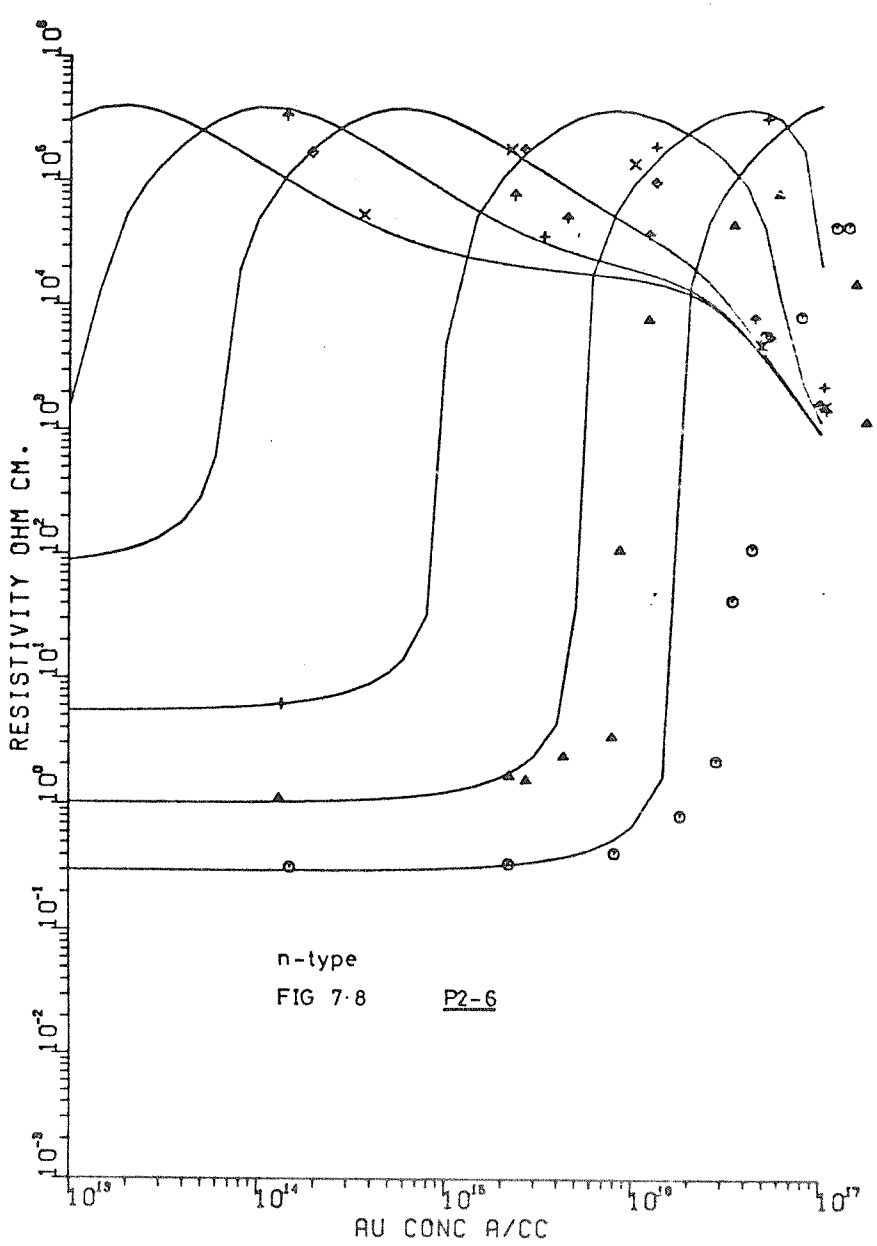
n-type

FIG 7·6 C1-4



n-type

FIG 7·7 V2-6



n-type

FIG 7·8 P2-6

et al's. results in this work. Figures 7.4 to 7.8 illustrate some of the results obtained using all of the resistivity data. At gold concentrations above 4×10^{16} the predicted values tend to be dominated by the gold-coupled shallow acceptor and are, therefore, subject to some uncertainty since its density relative to the total gold concentration has not really been measured satisfactorily. It should also be noted that the shapes of the predicted curves beyond the maximum resistivity are more dependent on the gold donor level (since the material is then p-type) whereas on the left of the peak the curve is almost solely dependent on the gold acceptor energy level. The models illustrated are as follows: figure 7.4 : C2-5 ; figure 7.5 : C2-4 ; figure 7.6 : C1-4 ; figure 7.7 : V2-6 and figure 7.8 : P2-6. (That model C2-5 gives the best fit is more easily seen in figure 7.17 where the assumption that only half of the gold is electrically active has been made in the calculation).

7.4.2 P-type silicon starting material

The p-type calculations are more difficult to fit to the measured data. This is apparent from figures 7.9, 7.10 and 7.11 where the effect of degeneracy on the predicted values for 1.1Ωcm p-type, gold doped, silicon starting material is shown. Varying the energy level or degeneracy factor is seen to affect the whole range of the curve. A second problem is that Thurber et al. (1973) expressed some doubt as to the validity of the measurements on higher resistivity material in which it was suspected that compensation from some donor-type dopant, other than gold, had occurred.

Including the factor-of-two correction for probable inactive gold, the best fits to the p-type gold doped silicon resistivity measurements occur for models : C2-5, C3-4 and V2-6. Of these, model C2-5 is also the best fit to the n-type gold doped silicon resistivity data and V2-6 is also a good fit; C3-4 is a poor fit. The monovalent models with energy levels varying in proportion to $E_g(T)$ do not give good fits to the p-type data. Figure 7.12 - 7. illustrates some of these p-type gold doped silicon resistivity results as follows: figure 7.12: C2-5 ; figure 7.13: C3-4; figure 7.14 : V2-6; figure 7.15: P2-6; figure 7.16: V1-4 (example of a poor-fit calculation).

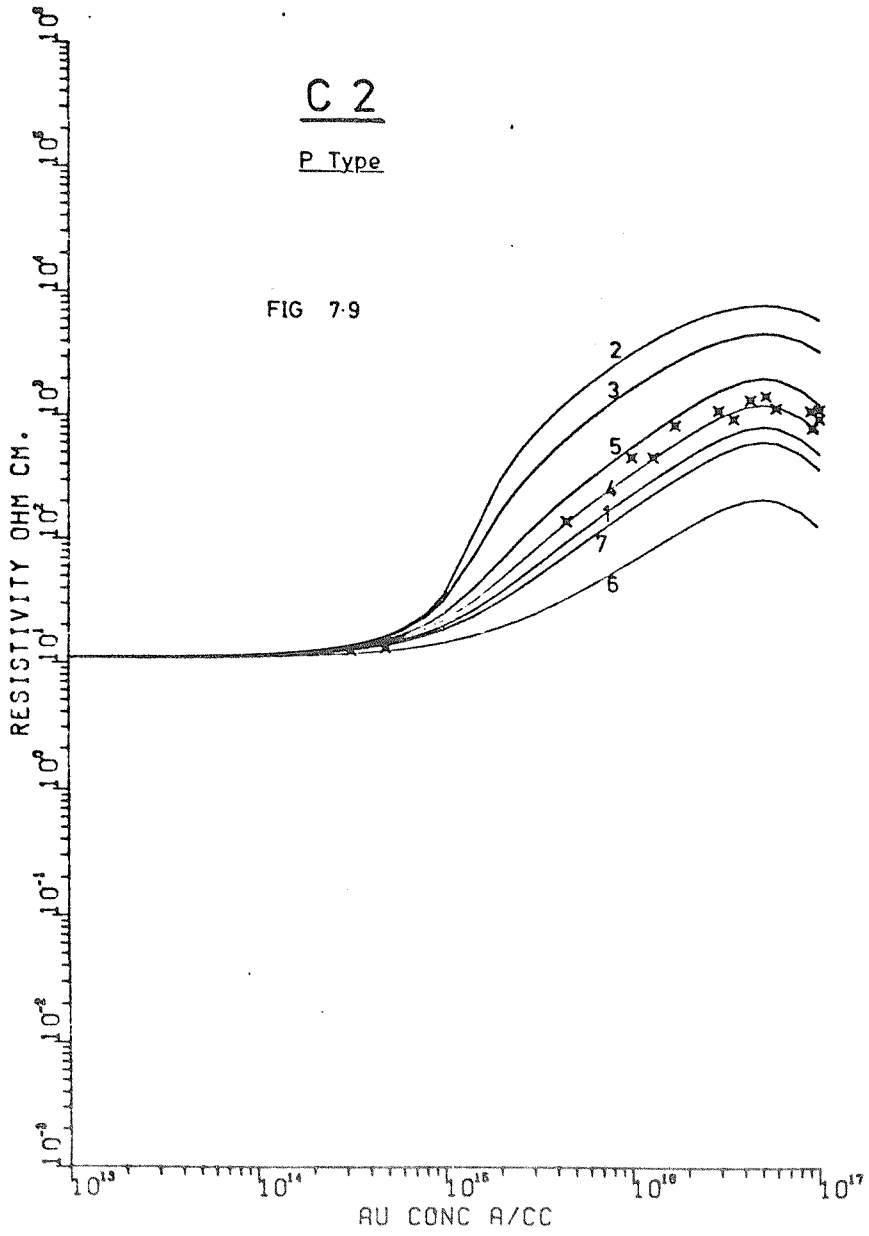


FIG 7·9

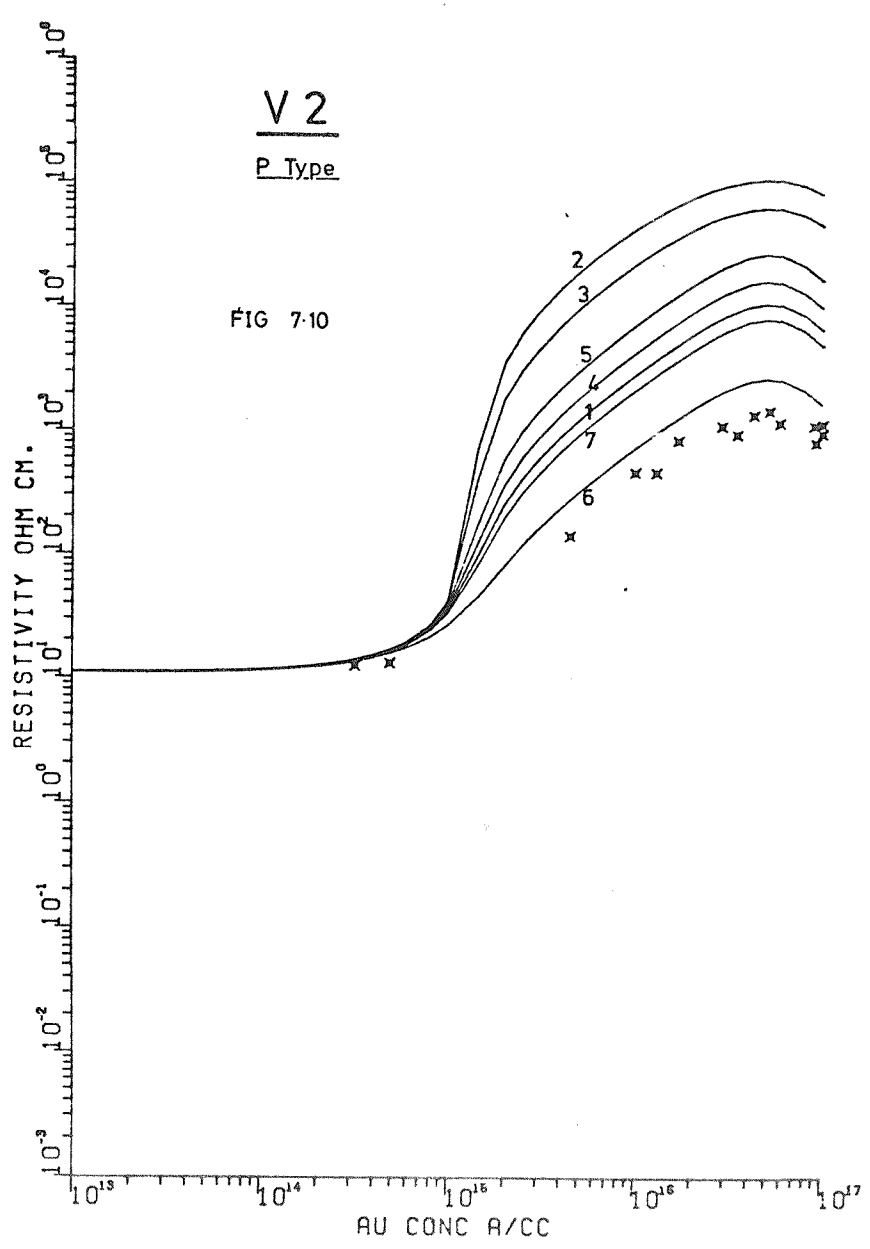
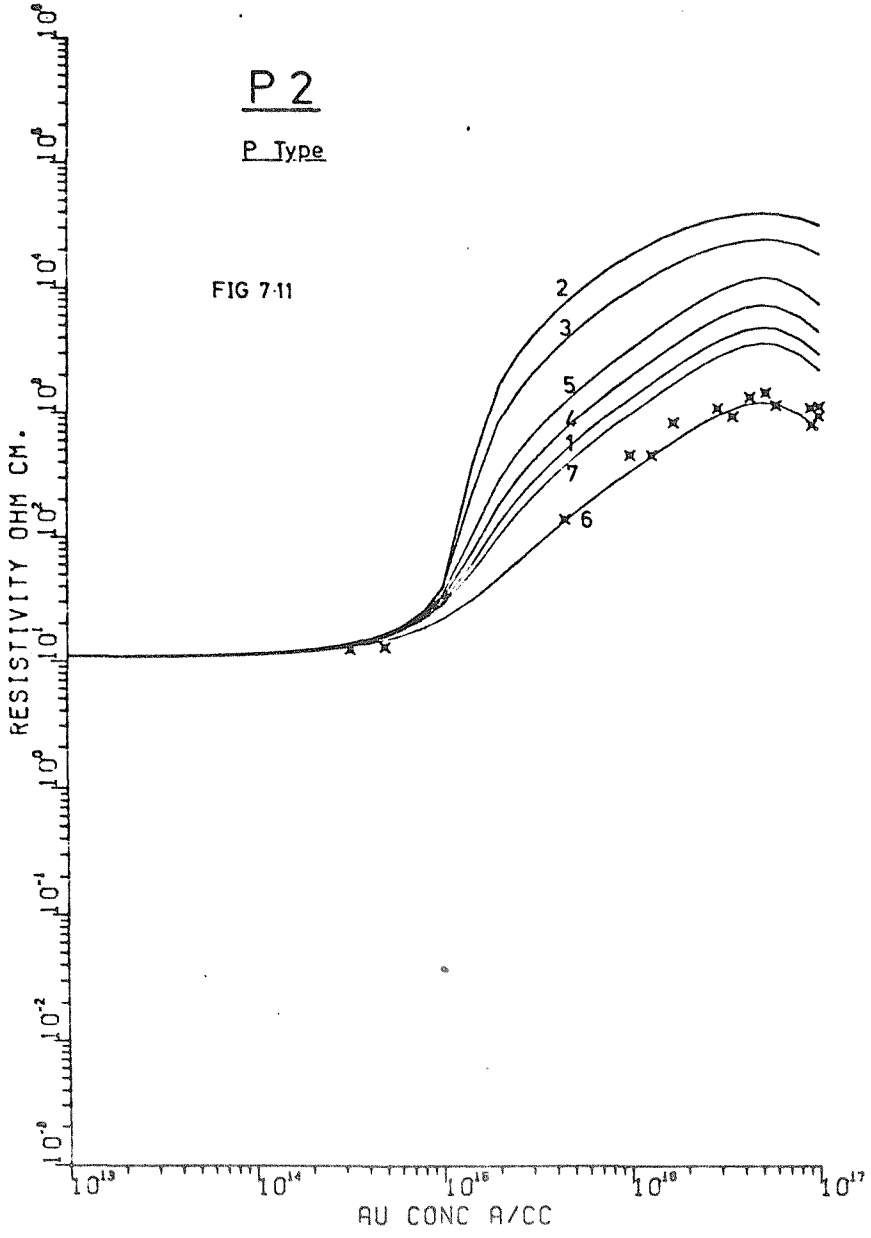
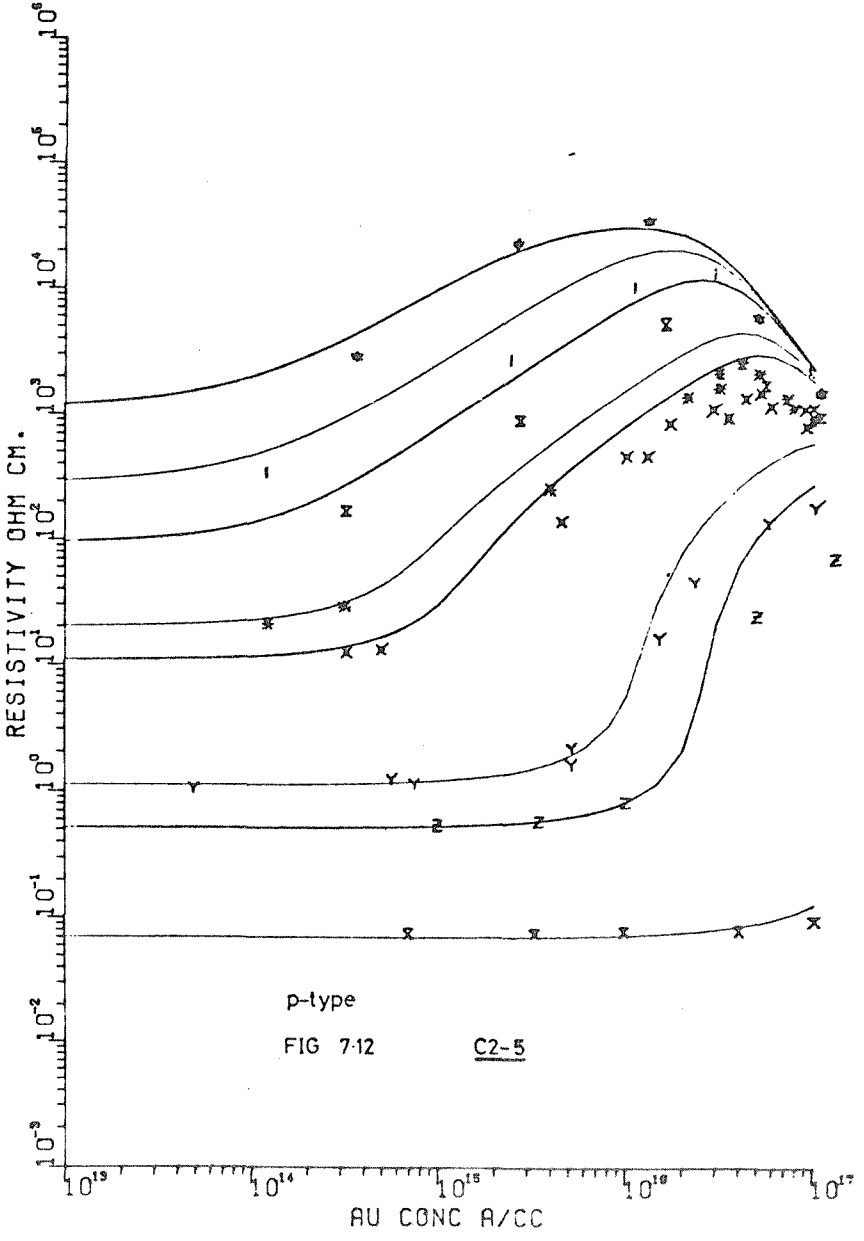


FIG 7·10


P 2
P Type
FIG 7-11
RESISTIVITY OHM CM.
AU CONC A/CC
2
3
5
4
1
7
6


RESISTIVITY OHM CM.
AU CONC A/CC
p-type
FIG 7-12
C2-5

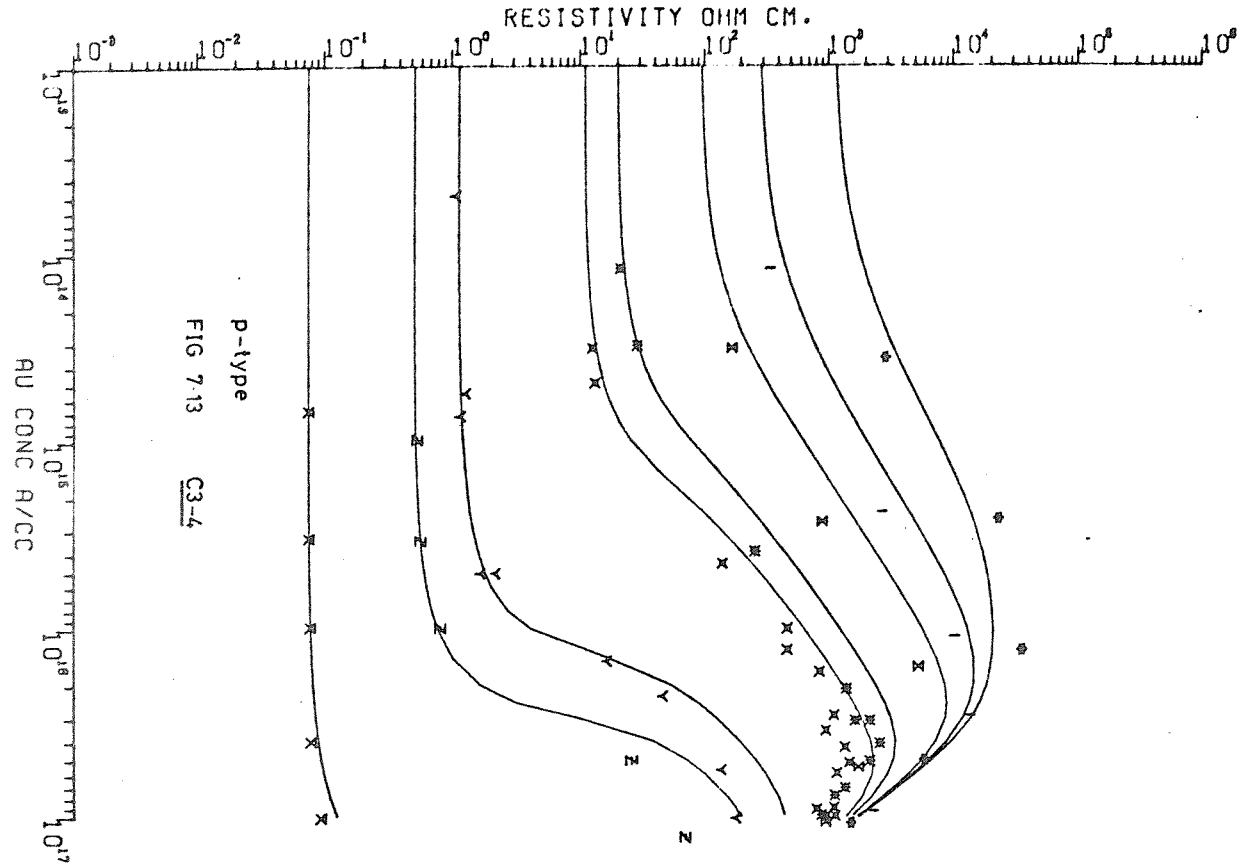


FIG 7.13 C3-4

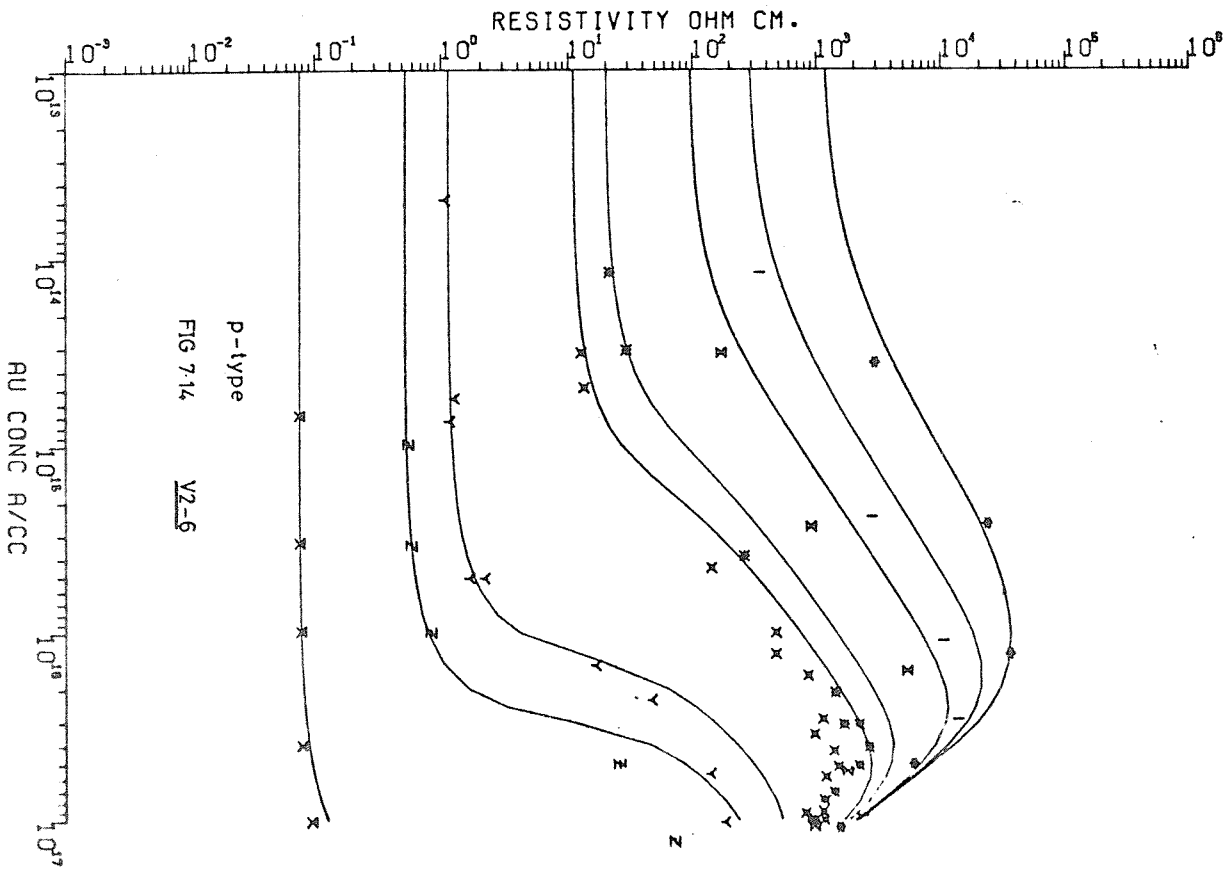


FIG 7.14 V2-6

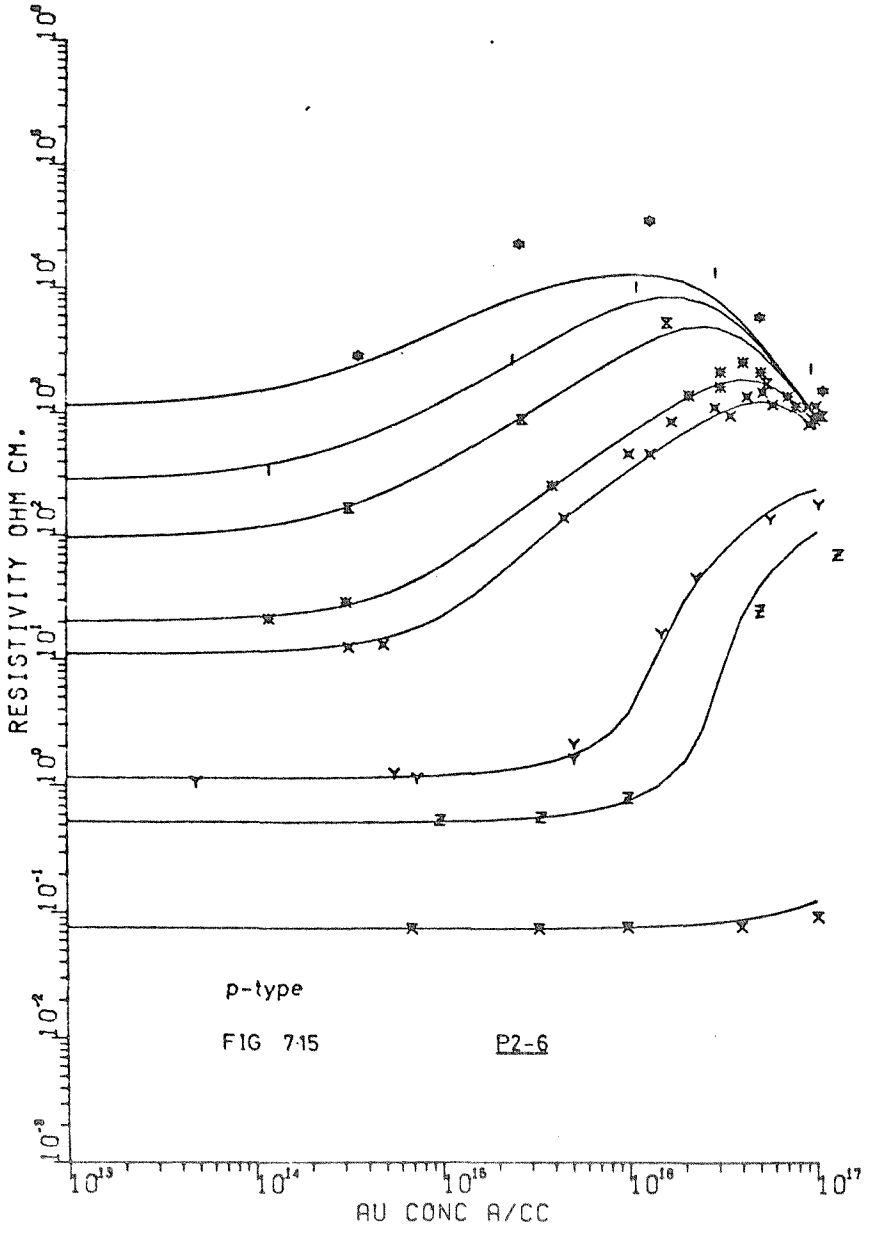


FIG 7·15 P2-6

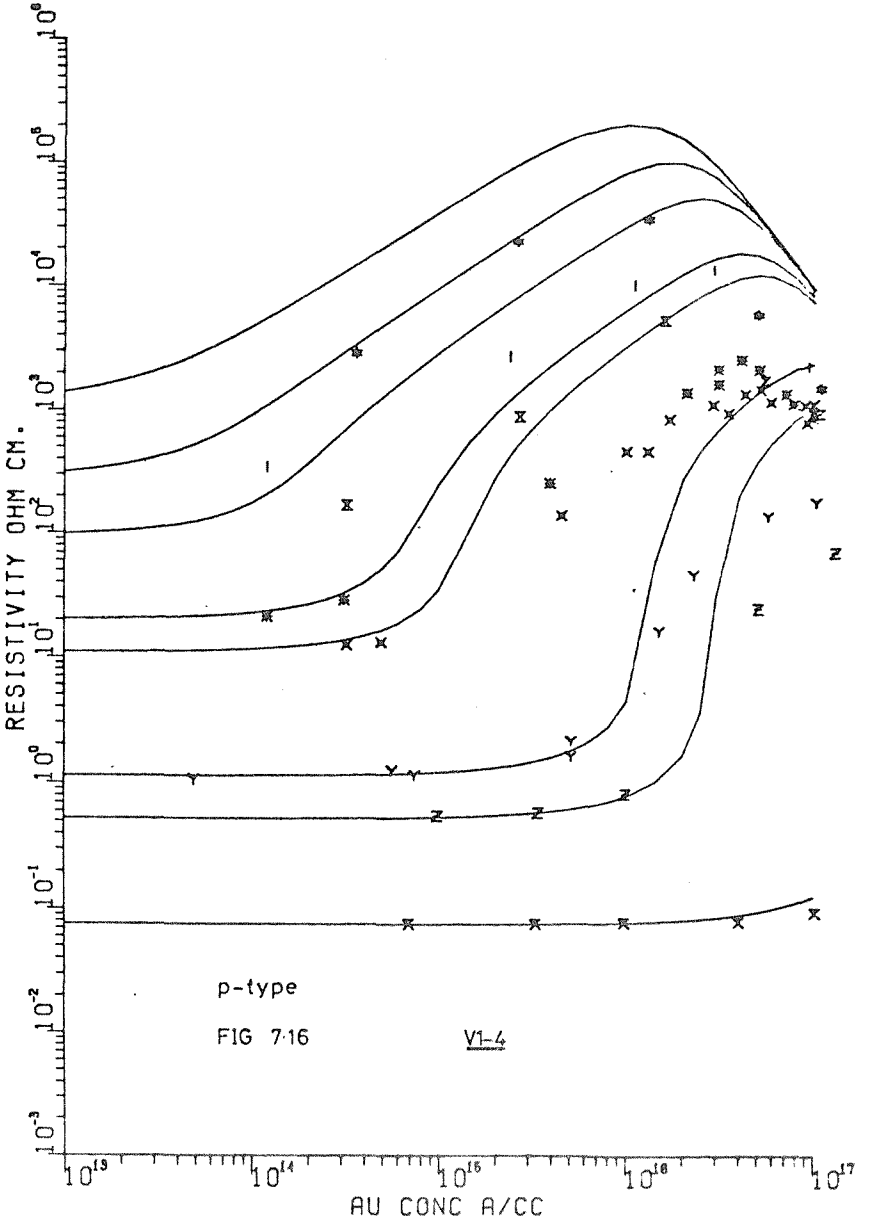


FIG 7·16 V1-4

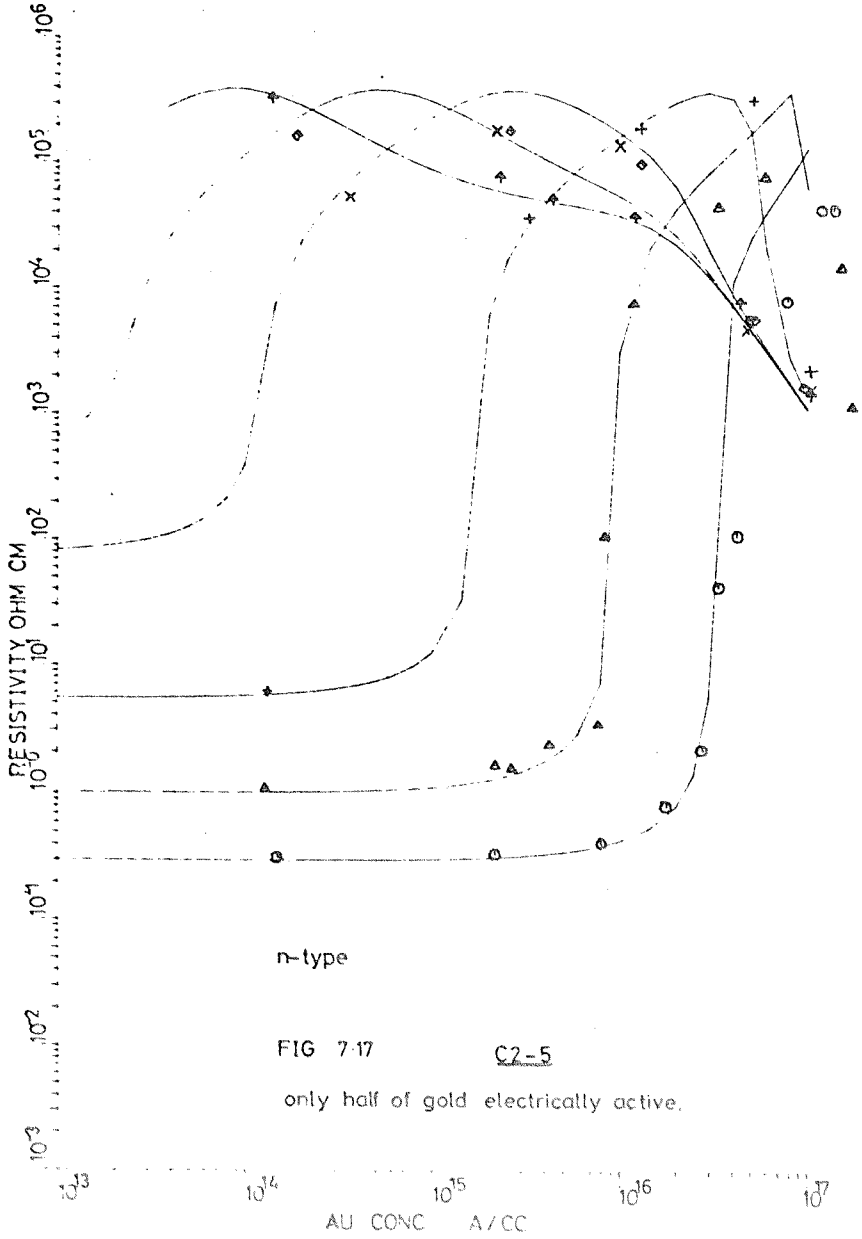
7.4.3 Conclusion

The best theoretical fit to the measured values of the effect of gold on silicon resistivity are obtained for a gold energy level model in which the gold is trivalent with energy levels which remain fixed to the conduction band edge as temperature varies ; $E_c - E_A = 0.553$ eV and $E_c - E_v = 0.84$ eV. The degeneracy factors $g_A = 6$ and $g_D = 2.5$ are obtained by including the split off valence band as well as the two double degenerate valence bands in assessing the degeneracies of the various charge states. Degeneracy factors of 4 and 1.5 also give quite a good fit (trivalent excluding split-off band). It is necessary, however, to make the assumption that approximately half of the gold atoms measured by neutron activation are not electrically active. Figures 7.17 and 7.18 show resistivity versus gold concentration plots of the best-fit gold model (C2-5) for n-type and p-type silicon in which it has been assumed, in the calculation, that half of the gold is inactive. A direct comparison between the predicted and measured values should be made. There are still anomalies in the results which are probably connected with the uncertainty in the measurements of the high resistivity p-type samples and in the concentrations assumed for the gold coupled shallow acceptor level.

In chapter 6 it was shown that the gold model which has provided the best theoretical fit to the measured data here, also gives an excellent fit to the solubility enhancement measurements. The monovalent model in which the gold energy level positions varied in proportion to $E_g(T)$ also gave a good fit to the solubility enhancement problem, however it does not fit the resistivity data as well as the trivalent model. Overall, therefore, the evidence provided by using the gold models discussed in chapter 4 tends to support the trivalent bonding configuration with energy levels fixed to E_c as temperature varies. It is stressed again, however, that it is possible to use any energy level with an appropriate degeneracy factor to give the same answers since they are inseparable (see equation 4.10).

7.5 The shallow interstitial gold donor

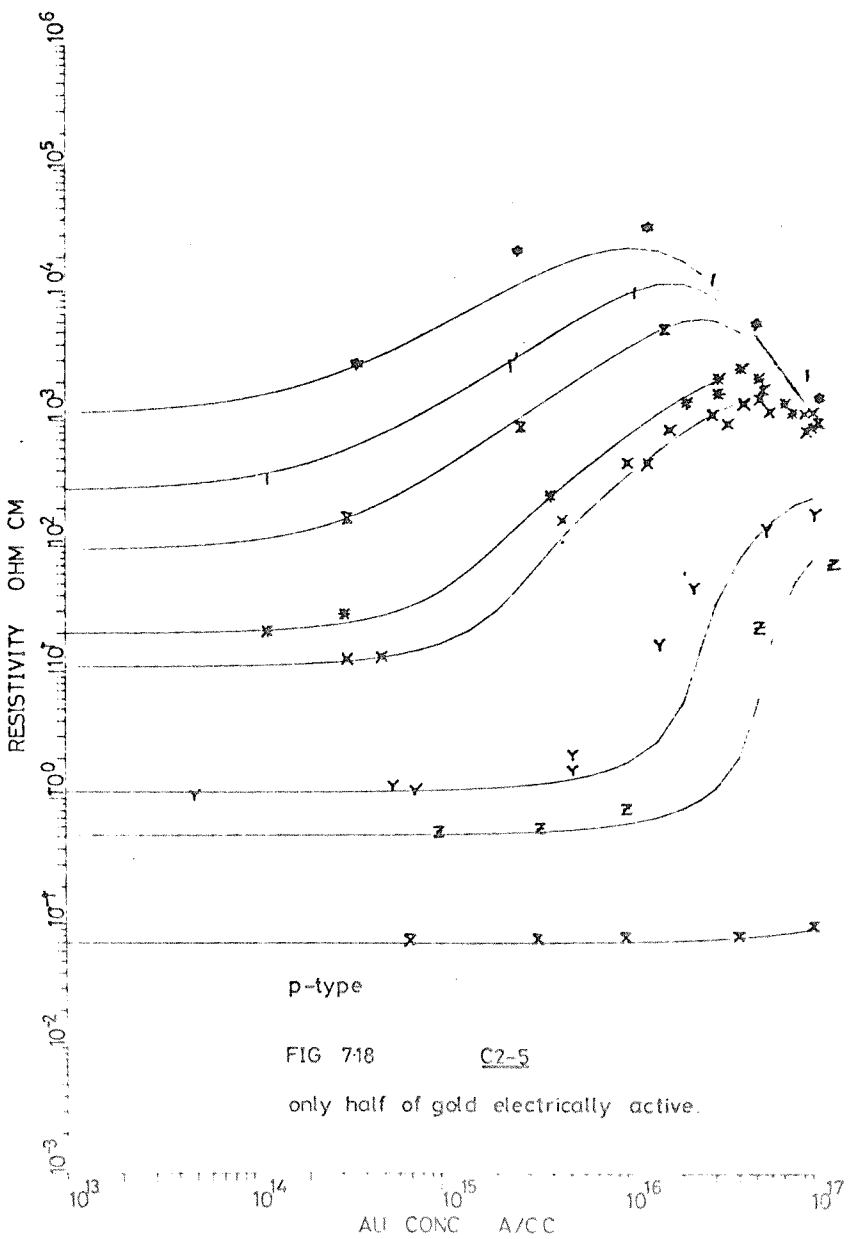
The possibility of interstitial gold behaving as a shallow donor has already been discussed in chapters 4 and 6. Calculations of the effect of



n-type

FIG 7·17 C2-5

only half of gold electrically active.



p-type

FIG 7·18 C2-5

only half of gold electrically active.

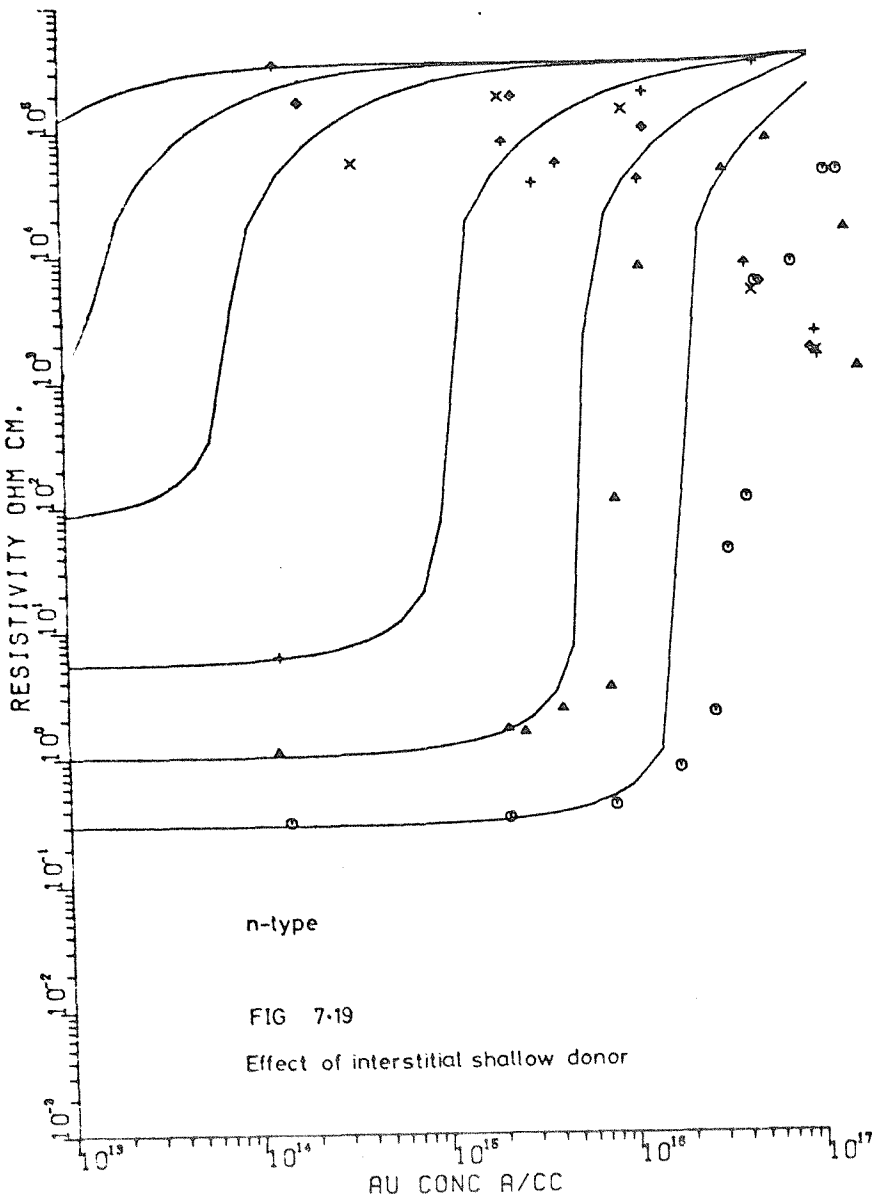


FIG 7·19

Effect of interstitial shallow donor

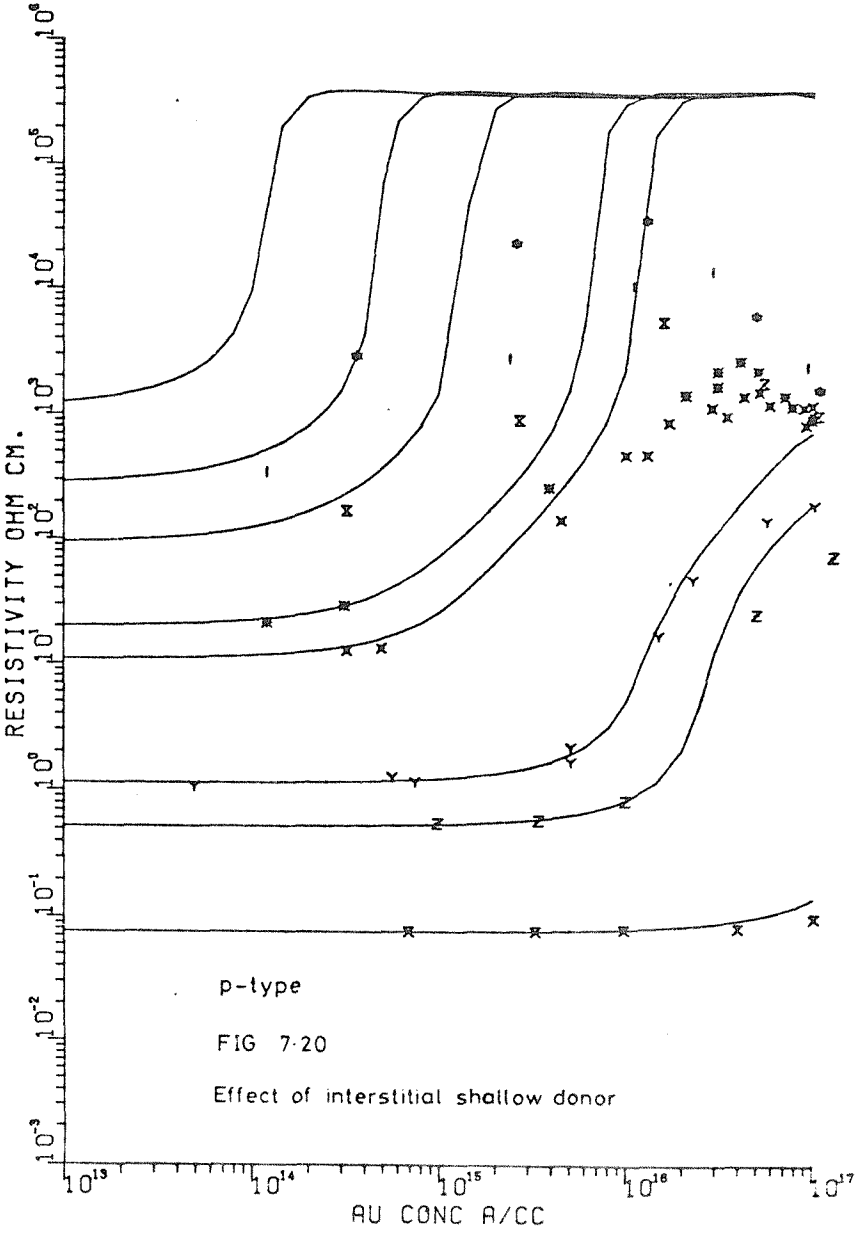


FIG 7·20

Effect of interstitial shallow donor

such an energy level on resistivity show clearly that, even if interstitial gold is a shallow donor at diffusion temperatures, it certainly is not at room temperature. Figures 7.19 and 7.20 illustrate the effect that it would have assuming that $Au_i = Au_s \times 0.1$. If half of the gold were interstitial the result would be even worse!

8. EXPERIMENTAL TECHNIQUES AND RESULTS

In this chapter the experimental techniques which have been employed in the study of gold diffusions in silicon wafers are described. The results from a number of experiments designed to investigate different aspects of the diffusion of gold in silicon in the presence and in the absence of shallow phosphorus diffusions are then described. Finally some results taken from a recent study of oxidation induced defects are used to illustrate the problem of furnace contamination. A discussion of the implications of the various results is given in chapter 9.

8.1 Gold Diffusion Profiling

Diffusion profiles are generally measured using radio-tracer gold or by neutron activation (Brown et al., 1975; Murarka, 1976). These methods, although accurate, are time consuming and expensive. In this work, the spreading resistance technique has been used to measure diffusion profiles of electrically active gold. This method has previously been reported for gold vs depth profiles in silicon by Brotherton and Rogers (1972). The gold distribution profile is inferred, reasonably accurately, from measurements of the variation in resistivity as a function of depth in gold doped and corresponding non-gold-doped-control silicon samples. The effect of gold on resistivity, which was discussed in chapter 7, is used to obtain the gold concentrations which are attributed entirely to the substitutional gold species because, as already shown, the interstitial species cannot be electrically active at low temperatures. (see 7.5)

In order to measure the slowly varying resistivity depth profiles which result from the experiments carried out here, a technique with high spatial resolution was required. Spreading resistance measurements (Mazur and Dickey, 1966) should provide such resolution, however in order to realise the potential of the technique a long period was spent developing a depth-profiling method which combined high spatial resolution and high reproducibility. The spreading resistance technique and the development work carried out on it are detailed in appendix A of this thesis. For the present it is merely necessary to accept that spreading resistance, R_s, which is directly related

to resistivity, may be measured in a localised volume of silicon around a metal-probe to silicon contact. The size of this volume, which defines the spatial resolution of the measurement, is closely related to the area of the probe-silicon contact. Two probes have been used in the experiments described here: the first yielded measurements which were reliable provided there were no abrupt changes in resistivity within a 50μm depth beneath the probe; the second, smaller probe gave reliable information provided there were no sudden changes within a 10μm depth beneath the probe. The effect of sampling the resistivity of the silicon over a volume of size defined by these depths is one of slight smearing-out of sharp variations. It is possible to make mathematical corrections for such fluctuations however in this work it has not been found necessary since the qualitative conclusions drawn from the results are unaffected - where appropriate the effect that making the corrections would have is indicated.

Spreading resistance values as a function of depth were obtained by polishing a shallow angle bevel through the samples to expose the areas of interest on a greatly magnified horizontal scale (see figure A1 in appendix A). The techniques for producing the bevel and measuring the depths of the readings are described in appendix A.

In order to obtain gold diffusion profiles, every gold diffusion was accompanied by control pieces cut from the same silicon wafer. The control pieces were subjected to the same processing treatments as the gold doped samples except at gold diffusion stage where they were given a similar heat treatment in a difference furnace. In cases where gold and phosphorus were both diffused into a wafer four quarters were processed separately :

(1) Phosphorus and gold diffused.
(2) Phosphorus diffused only + heat treatments to simulate Au diffusion.
(3) Gold diffused only + heat treatments
(4) All heat treatments only.

Comparison of (1) and (2) enables the gold profile in (1) to be obtained and comparison of (3) and (4) enables the gold profile in (3) to be obtained.

8.1.1 Inference of gold concentration from resistivity change

The effect of gold on the resistivity of silicon has been described in detail in chapter 7. In this work the problem of main interest is the effect of shallow phosphorus diffused layers on the distribution of gold in the rest of a silicon wafer. The gold concentration in the shallow phosphorus diffused layers could not be detected by resistivity change since it never reaches a high enough concentration to alter the number of carriers significantly. In the bulk of the substrates, which were 10-20Ω cm p-type wafers, the presence of gold in densities greater than 10^{14} atoms/cc causes a rise in the resistivity. Comparison of the resistivity depth profile of the non-gold-doped control sample with that from the corresponding gold doped sample should allow the level of gold contamination to be calculated. It has not proved possible to predict the experimentally observed effect of gold on resistivity in p-type silicon with great accuracy so the experimental data rather than theoretical predictions are used to convert resistivity change to gold concentration. Here the experimental results reported by Thurber et al. (1973) were used. Although an uncertainty of a factor of two exists in the absolute values of gold concentrations thus inferred, relative values of gold concentration should be comparable with much greater accuracy. Because of this uncertainty, values of spreading resistance versus depth have only been converted to gold concentration when really necessary. Where such a conversion does not affect the qualitative conclusions being drawn from the data the results are plotted in the form of the raw data, namely spreading resistance versus depth or in the form of resistivity versus depth.

Resistivity versus gold concentration data - as shown in chapter 7 - was kindly provided by Dr. W. R. Thurber of the National Bureau of Standards, New York. In order to interpolate the relationship between resistivity and gold concentration for any starting material resistivity, the following exercise was carried out. The data was first plotted as shown in chapter 7, i.e. resistivity versus N_{Au} for the experimental starting resistivities. Smooth lines were drawn through the data points and a plot of resistivity versus the background doping level with gold concentration as a parameter was made up. This is shown in figure 8.1. These plots allowed a curve of resistivity versus N_{Au} to be drawn up for

Resistivity vs Background Boron Doping Le
(obtained from

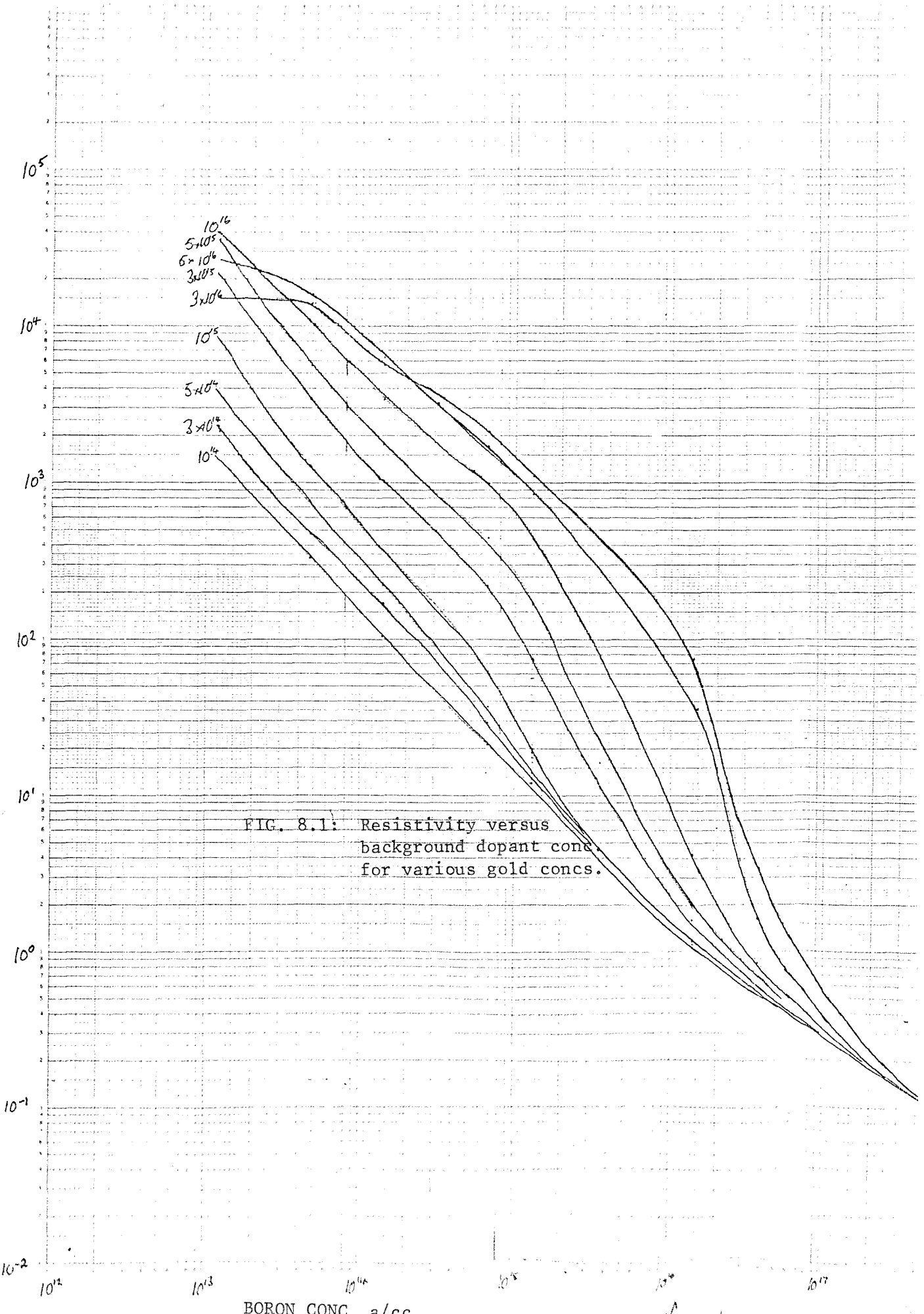


FIG. 8.1: Resistivity versus background dopant conc. for various gold concs.

ρ
Resistivity.
Ω cm

N_B atoms/c

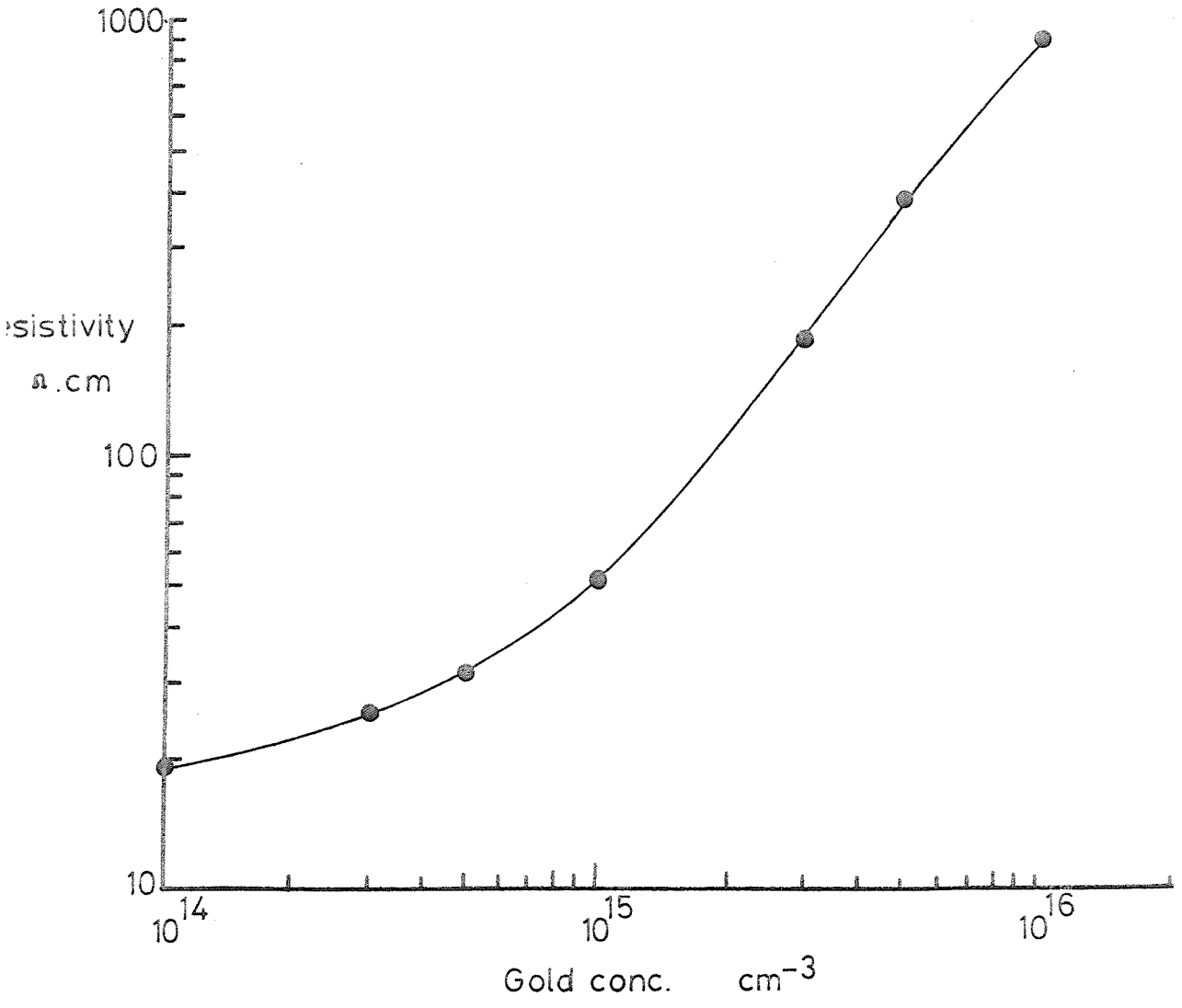


FIG 8.2 Resistivity versus gold conc. for N_{Boron}=7.5 x 10^{14} a/cc.

any chosen background doping level. An example is shown in figure 8.2. Gold concentration versus depth profiles were obtained in the following manner:

(1) The spreading resistance versus depth data was plotted out.

(2) This data was converted from spreading resistance to resistivity versus depth and then smoothed.

(3) The average resistivity of the non-gold doped control sample was converted to carrier concentration using empirical curves due to Irvin (1962).

(4) A curve of resistivity versus N_{Au} (eg figure 8.2) was drawn up for this carrier concentration using figure 8.1.

(5) The gold concentration at any depth for the gold doped samples was obtained using this calibration curve.

8.1.2 Errors

The main sources of error in the gold diffusion profiles presented later in this chapter arise from the conversion of resistivity to gold concentration (see above) and from limitations of the spatial resolution of the spreading resistance equipment. This latter effect is discussed where appropriate in the results. The former effect is mainly due to the uncertainty in the data used to convert resistivity to gold concentration rather than in measuring the resistivities since it is shown in appendix A, that the spreading resistance equipment used here has a reproducibility better than 2%. It is not possible - or meaningful - to quantify this error in conversion from ρ to N_{Au} since it should always be the same. This implies that although the absolute values of N_{Au} may be in error by up to a factor of two, the same factor applies to every measurement making comparisons between measurements valid to a much greater accuracy. The graphical conversion is probably subject to an error of $\pm$ 5% giving an overall error bar of $\pm$ 10% except in regions where the resistivity variation with depth is very rapid or abrupt.

8.2 Shallow Diffusion Profile Measurements

Dopant-depth profiles of the shallow phosphorus diffused layers were obtained from measurements of resistivity versus depth made by the spreading resistance method or by the incremental sheet resistance method.

These techniques and the analysis used to convert the measurements to dopant concentrations are described in appendices A and B.

8.3 Junction Depth Measurements

The measurements of junction depths were carried out by conventional staining techniques (Burger and Donovan, 1967). Shallow diffused n-p junctions were exposed by making a shallow bevel or a cylindrical groove in the silicon surface so as to expose the layers of interest; the bevelled and grooved surfaces were polished with 0.5 micron diamond paste. The p-type regions were stained with an HF/HNO_3 stain made by adding 3 drops of concentrated HNO_3 to 25 cc of 48% HF. Once these regions had been stained the depths of the junctions were measured using a Watson interferometer under sodium light illumination. (λ = 5908 Å).

Junction depths on samples with polished bevels which had been used for spreading resistance measurements were stained with a solution comprising a few crystals of copper sulphate dissolved in a small quantity of very dilute HF. This stain plates copper all over the exposed silicon but the plating occurs much more rapidly on n-type layers allowing them to be identified within a few seconds. For some unknown reason the HF/HNO_3 stain did not work on these polished bevels.

8.4 Rutherford Backscattering Measurements

Rutherford backscattering (RBS) measurements of the concentration and lattice location of gold in heavily phosphorus doped silicon were kindly carried out by Dr. P. L. F. Hemment of the University of Surrey.

The RBS technique, which is described very briefly below, has been reviewed as a method for chemical analysis by Buck and Wheatly (1972).

The silicon sample of interest is bombarded with a collimated beam of 1.5 MeV He^+ ions in an evacuated system. The sample, or target, is aligned in a random crystal direction relative to the beam. Upon collision with the target, the He^+ ions are backscattered, collected and counted. The energy of a backscattered ion may be related to the mass of the impurity with

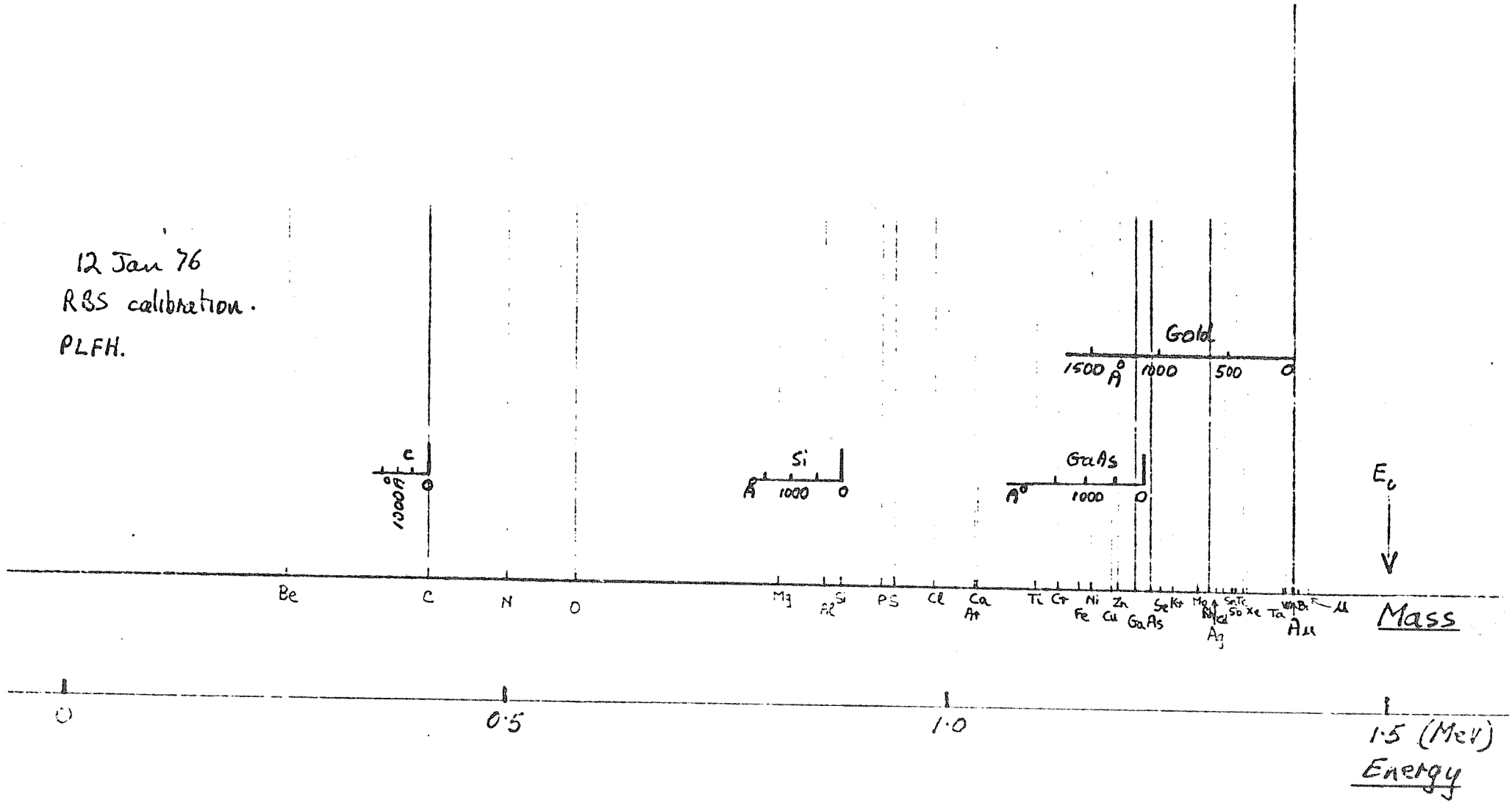


FIG 8·3 Calibration of Surrey RBS System. Mass of backscattering atom ➡ backscattered energy.

which it has collided or its depth within the target specimen. For a given impurity causing backscattering, ions will be collected over a range of energies with the maximum energy corresponding to scattering from atoms of the impurity which lie closest to the surface. The number of backscattered ions of a given energy, collected for a given period, may be related to the density of the target impurity corresponding to that energy.

The energy range is divided into channels of 8.0 KeV and the collection system produces an output indicating the yield of backscattered ions in each channel or energy range increment.

If the example of gold in silicon is considered, the ratio of the density of gold atoms to silicon atoms may be obtained for a given dose of He^+ ions. If the channels corresponding to scattering from the silicon surface layers and scattering from gold atoms in the silicon surface layers have yields of Y_{Si} and Y_{Au}, then the densities of silicon and gold are related in the following manner:

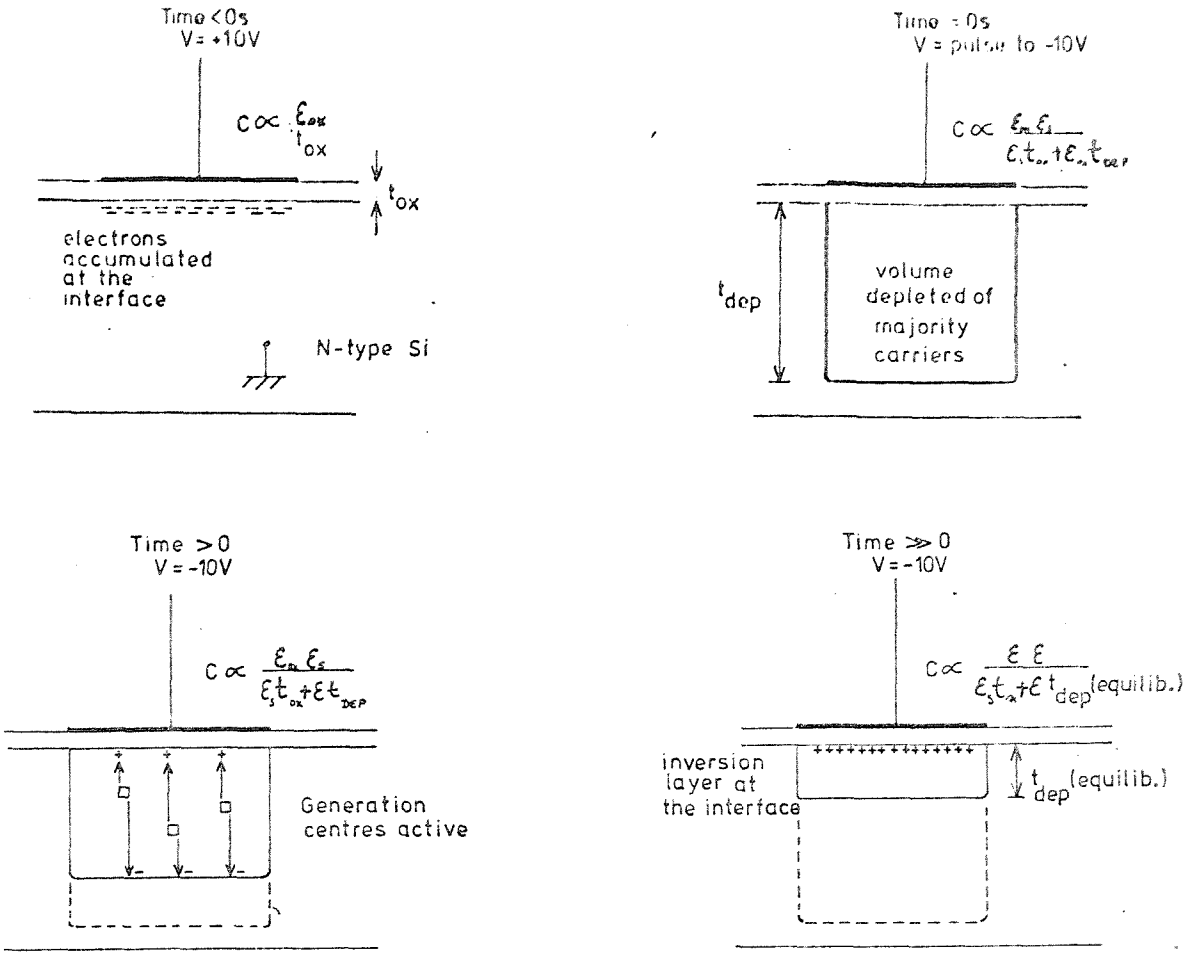
$$N_{si} : N_{Au} = \frac{Y_{Si}}{Z_{Si}} : \frac{Y_{Au}}{Z_{Au}} \qquad 8.1$$

where Z_{Si} and Z_{Au} are the atomic numbers of silicon and gold respectively. The density of gold close to the silicon surface is given by :

$$N_{Au} = N_{Si} \left(\frac{Y_{Si} \cdot Z_{Au}}{Y_{Au} \cdot Z_{Si}} \right) \qquad 8.2$$

where, at 300°K, $N_{Si} = 5 \times 10^{22}$ atoms/cc. The detectability limit for gold is about 10^{17} atoms/cc which means that gold concentrations must be above the solubility limit in intrinsic silicon before they are detectable. This limitation prevents use of the RBS technique for most gold concentration measurements, however it provides an ideal method for obtaining gold concentrations in heavily phosphorus doped layers.

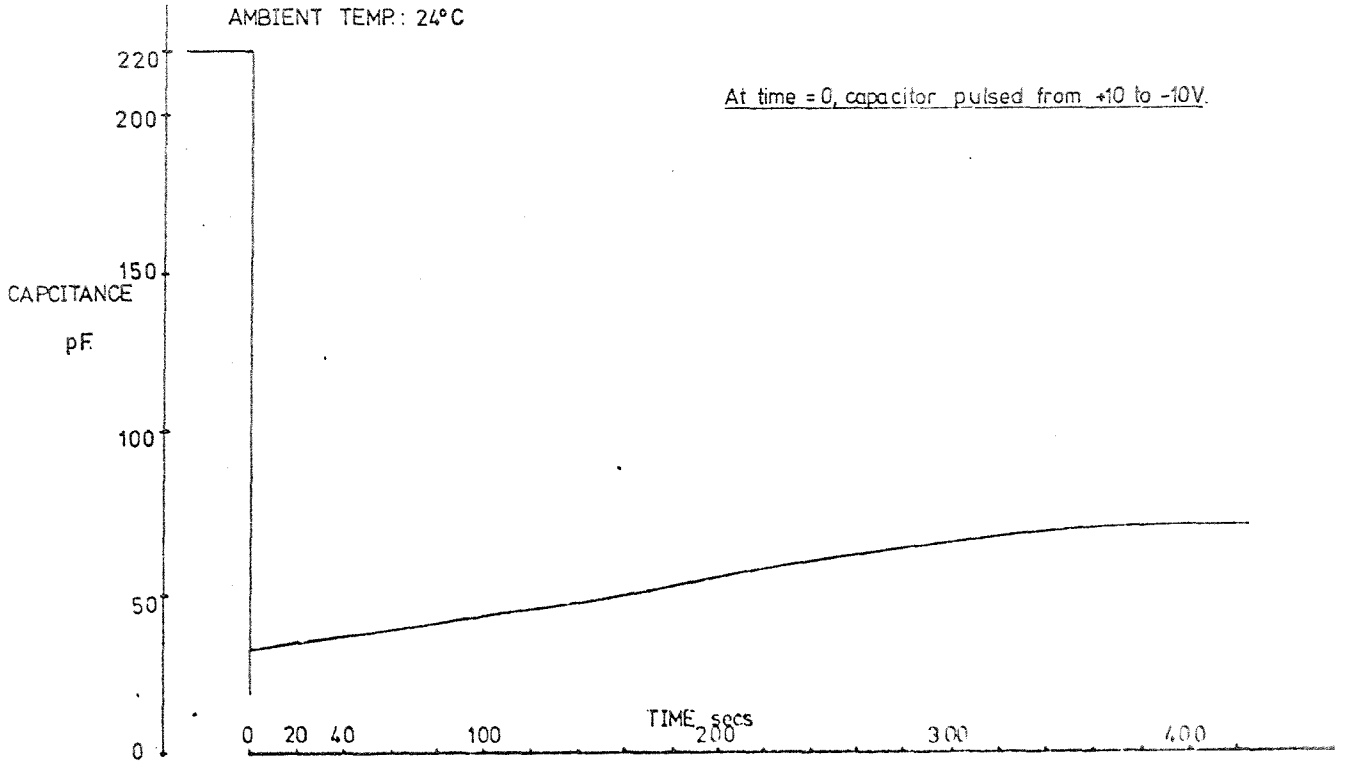
Figure 8.3 is a calibration chart for the RBS system used for measurements in this work. The energy for backscattered ions from the silicon surface is approximately 0.88 MeV; any impurity (such as carbon) with a lower mass will be masked by the silicon depth spectrum. Gold at



DEEP-DEPLETION—to—INVERSION RESPONSE of CAPACITOR

FIGURE 8.4.

TYPICAL CAPACITANCE vs TIME PLOT AFTER A DEPLETING PULSE



the surface, as can be seen, scatters ions at about 1.4 MeV. A continuum of backscattered ions over the energy range 0.88 - 1.4 MeV may be related to the gold depth distribution or to the presence of significant quantities of the elements with atomic masses between silicon and gold.

In the above description, it has been assumed that the target sample was aligned in a random direction to the ion beam, it is possible however, to channel the ions down interstices in the silicon lattice because the He^+ ions are very small compared to gaps in the lattice in certain crystal directions. If the sample is aligned in a channelling direction relative to the ion beam, the backscattered count from silicon atoms and impurity atoms residing on the lattice sites becomes very small; however, a significant count at a particular energy, indicates impurities of the atomic mass corresponding to that energy residing in interstitial or near-interstitial sites. For the same total dose of He^+ ions, the ratio of the backscattered yield at, say the gold energy, in the random direction to the yield at the same energy in the channelling direction indicates the ratio of total gold to interstitial gold. (In order to expose all of the interstitial sites to a channelled beam, several channelling orientations must be tried, however in the limited time available here, it was only feasible to do one channelling experiment).

8.5 Pulsed MOS Capacitor Measurements

A very brief reference to some recent experimental results obtained by this author in a recent study of oxidation induced defects (Unter et al, 1977) is made at the end of this chapter. The results illustrate the effect of furnace contamination during gettering. Since these results are only briefly mentioned, a short description of the experimental technique will suffice.

The pulsed capacitor technique (Zerbst, 1966) gives a simple method of estimating the minority carrier generation rate in a depleted MOS structure. The measurement, which is summarised diagrammatically in figure 8.4, involves monitoring the high frequency capacitance of the MOS capacitor when it is pulsed from accumulation into deep depletion in darkness. The high frequency capacitance gives a measure of the width

of the depletion region which, as shown in the diagram, shrinks to an equilibrium width while generation centres within its volume are active. The time taken for the depletion width to relax from deep depletion to equilibrium (where the surface is inverted) may be related to the generation rate in the bulk - provided there is negligible generation at the $Si-SiO_2$ interface (Zerbst, 1966; Heiman, 1967). A typical response is also shown in the figure.

The relationship between the response, or relaxation, time of the pulsed capacitor and the number of generation centres within the depletion region is derived in the references given above, the results presented here are for comparison with each other and therefore are given as capacitor relaxation times.

8.6 Processing Techniques

The following processing stages were used in the experiments described here :

(1) Silicon wafer cleaning.
(2) Thermal oxidation.
(3) Phosphorus diffusion.
(4) Epitaxial growth.
(5) Standard photomasking.
(6) Gold deposition and diffusion.

Prior to gold doping, the wafers were handled and processed with standard equipment available in the Microelectronics facility at Southampton University. All processing subsequent to gold doping was carried out in specially segregated furnace tubes and using designated handling equipment. This was done to avoid contamination of apparatus which would subsequently be used for handling and processing devices in which long minority carrier lifetimes were desired.

Processes listed as 1-5 above are standard, well documented techniques.

Wafer cleaning was generally carried out using the so-called 'R-C-A' clean (Kern et al, 1970) or by boiling wafers in aqua-regia (3 parts HCl:

1 part HNO_3) which is a well known solvent for gold. In the lists of processing procedures given later in this chapter, wafer cleaning before any furnace treatment is implicit.

Thermal oxidations (which were employed to provide masking oxides both against phosphorus diffusions and against surface diffusion of gold - Huntley and Willoughby, 1970) were carried out at 1100°C or 1200°C in a wet oxygen ambient.

Phosphorus diffusions were carried out as a two stage process: deposition of a phosphorus glass from a $POCl_3$ (0°C) source at 1050°C (± 1°C) or at 1000°C (± 1°C) followed by a drive-in in an oxidising or non-oxidising ambient at another temperature; (precise details are given with each experiment). The furnaces were 2½" internal diameter.

Epitaxial layers, which were required for one of the experiments to be described, were deposited in a vertical, R.F. heated reactor from the vapour-phase reduction of silicon tetrachloride at 1200°C (Grove, 1967, p8).

Gold diffusions were carried out from an elemental source of high purity gold (99.999%) which was evaporated onto the silicon wafers from a molybdenum boat at a pressure of 2×10^{-5} Torr or less. Prior to evaporation, the silicon surface was dipped in buffered HF to remove any trace of oxide which might cause non-uniformity of diffusion. The diffusions, which all took place at 1000°C (± 2°C), were carried out in a double walled furnace tube with a nitrogen ambient. O'Shaughnessey et al. (1976) have shown that when gold is diffused in an oxidising ambient, surface gold is rejected to the top of the oxide layer as it grows. This eventually leads to a situation in which the gold source can be severely depleted.

All wafers which had been diffused with gold were quenched from the furnace temperature to room temperature in less than two seconds. This was done in an attempt to prevent redistribution of gold during cooling.

8.7 Experiments and Results

The experiments presented below are not in chronological order but have

been listed in a logical manner. Most of the experiments were carried out at least twice to ensure that the conclusions drawn were not the result of a freak occurrence.

For brevity and, it is hoped, clarity, each experiment is listed as follows:

Object of the experiment
Processing procedures
Measuring procedures
Results and relevant comments.

A discussion of the implications of these results is then given in the next chapter.

8.7.1 Phosphorus-gold sequential diffusions at 1000°C

This experiment was designed to assess the effect of a shallow, high concentration phosphorus diffused layer at the front of a silicon wafer on the distribution of gold subsequently diffused in from the back.

Material : 10-20 Ωcm, p-type, (111) silicon wafers 350 microns thick. This material was designated by the manufacturer as dislocation and swirl-defect free and was grown by the Float Zone (FZ) technique.

Processing :

i Wafer cleaned.

ii Initial masking oxide; 60 mins wet oxygen, 1100°C.

iii Wafer quartered; quarters numbered 1-4.

iv Oxide removed from fronts of 1 and 2.

v All quarters cleaned.

vi 1 and 2 given 30 mins. phosphorus deposition at 1050°C:

Gas flows : 30 cc/min N_2 through $POCl_3$ at 0°C
10 cc/min O_2
1 /min N_2

3 and 4 given equivalent heat treatment in N_2.

vii 1-4 given 'drive-in' at 1050°C, 30 mins dry N_2 + 20 mins wet O_2.

viii Oxide removed from backs of 1 and 3.

ix Au evaporated onto BACKS of 1 and 3.

x 1 and 3 annealed at 1000°C in 'gold' furnace. Dry N_2.

xi 2 and 4 annealed for same time as 1 and 2 in 'control' furnace.

Five different gold anneal times (steps x and xi) were employed:

Run G51 15 mins Au anneal

Run G52 30 mins Au anneal

Run G53 60 mins Au anneal

Run G54 120 mins Au anneal

Run G55 240 mins Au anneal

Measurements : For each run (G51-55) all four quarters were resistivity-depth profiled by the spreading resistance technique using the small, high resolution probe. In addition the phosphorus layer sheet resistances and junction depths were obtained by grooving and staining quarters 1 and 2 of each run.

Results: The sheet resistances and junction depths of the P-diffused layers were almost identical for the five different gold diffusion times. They were :

Run No.	R_{SH} Ω/□	x_j μm
G51	3.86 ± 0.03	3.6 ± 0.15
G52	3.98 ± 0.03	3.69 ± 0.15
G53	3.98 ± 0.03	3.47 ± 0.15
G54	4.08 ± 0.03	3.62 ± 0.15
G55	4.1 ± 0.03	3.7 ± 0.15

It is reasonable to conclude that the phosphorus layer, which was originally diffused in at 1050°C, does not redistribute appreciably during the lower temperature anneals. Only one of the phosphorus dopant-depth profiles is shown here in figure 8.5 since those obtained from the other samples are almost identical. One of the phosphorus diffused control samples was etched in SECCO etch (Secco d'Aragona, 1972) to see if diffusion

induced lattice damage could be detected. No features which could be associated with such damage were visible after etching in one minute steps for a total of five minutes at room temperature.

The resistivity and dopant depth profiles from the bulk of each wafer are shown in figures 8.6 - 8.15. Two figures for each run are presented. The first depicts the resistivity vs depth data obtained from the spreading resistance measurements on quarters 1 (P + Au), 2(P only), 3(Au only). Each spreading resistance data point has been converted directly to resistivity so that the scatter in the values of resistivity reflects the scatter in the original measurements. (The results for quarter number 4 are not shown because in all cases they overlap those obtained from quarter number 2). The second figure for each run is of gold concentration versus depth for quarters 1 and 3 (gettered and ungettered). In these, smoothed resistivity data was converted to gold concentration by the method outlined in section 8.1.1. The detectability limit for gold contamination is defined by the gold concentration below which there is no effect on resistivity. The shallow diffused layers are not shown on these figures so it is stressed that the phosphorus layer is diffused into the sample surface (left side of the profiles) and the gold source is applied to the back of the sample (right side of the profiles).

The symbols used in the figures are :

■ = quarter no. 1 (P + Au)

● = quarter no.3 (Au)

X = quarter no. 2 (No Au)

These results clearly show the gettering effect of the phosphorus doped layer on the incoming gold. They also show that, after a period of some time between 60 and 120 minutes, the gettering effect of this phosphorus layer ceases.

Transport of gold from the source on the wafer back to the wafer front by surface diffusion was inhibited by the oxide layer, therefore the tip-up in gold concentration at the face opposite the gold source in the ungettered samples is considered to be a phenomenon which acts on gold atoms which have already diffused through the thickness of the wafer. These profiles

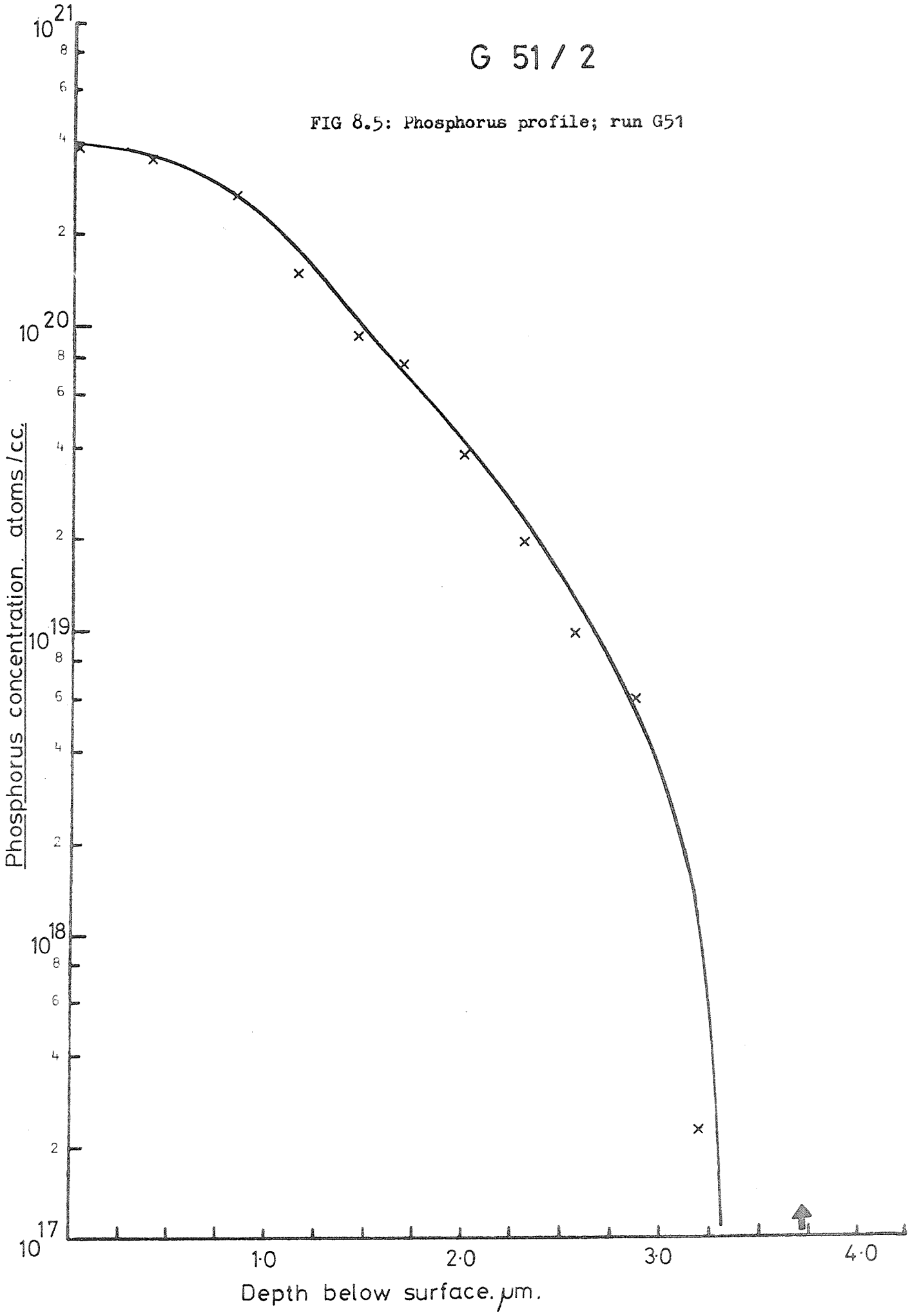


FIG 8.5: Phosphorus profile; run G51

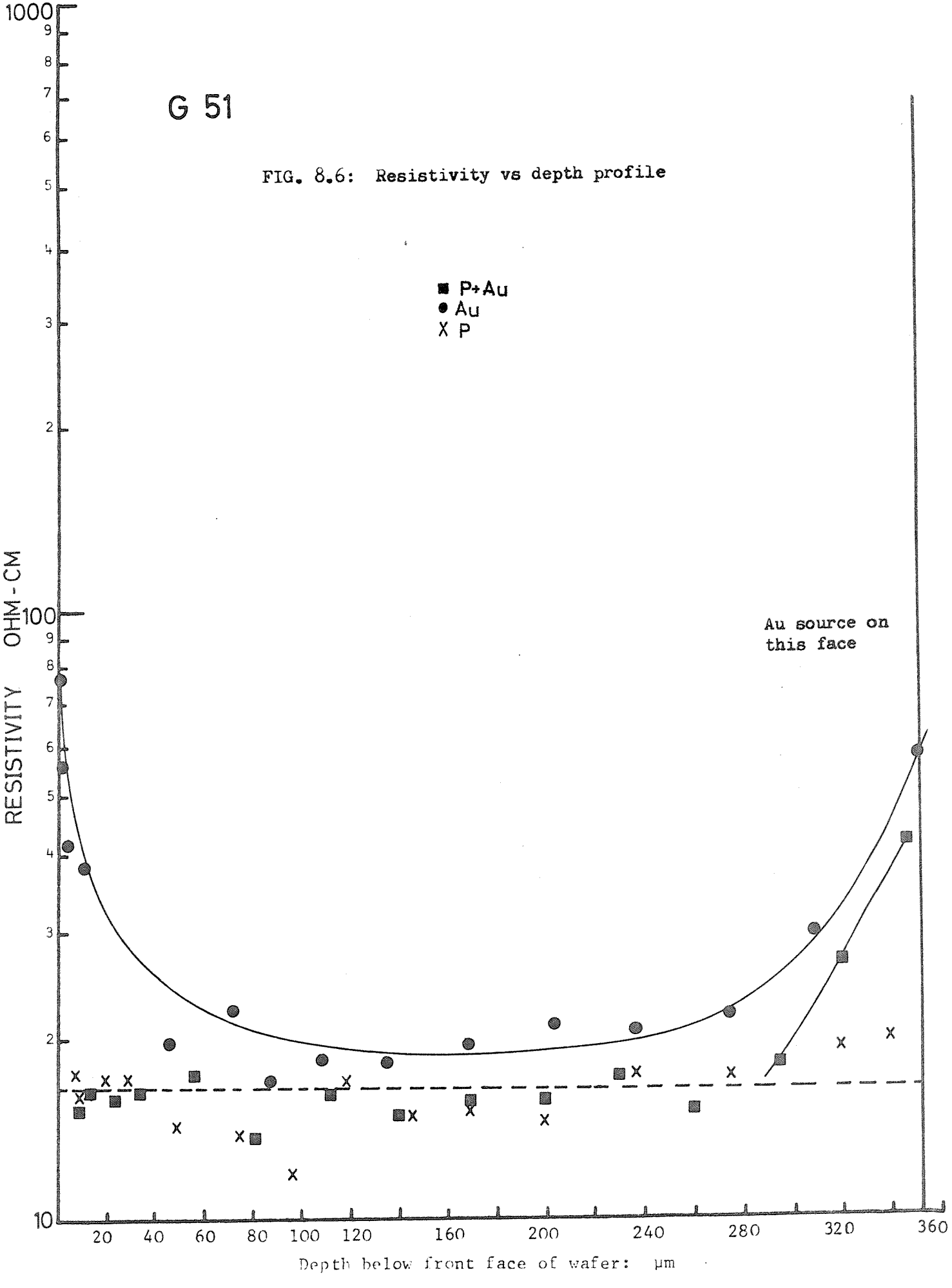


FIG. 8.6: Resistivity vs depth profile

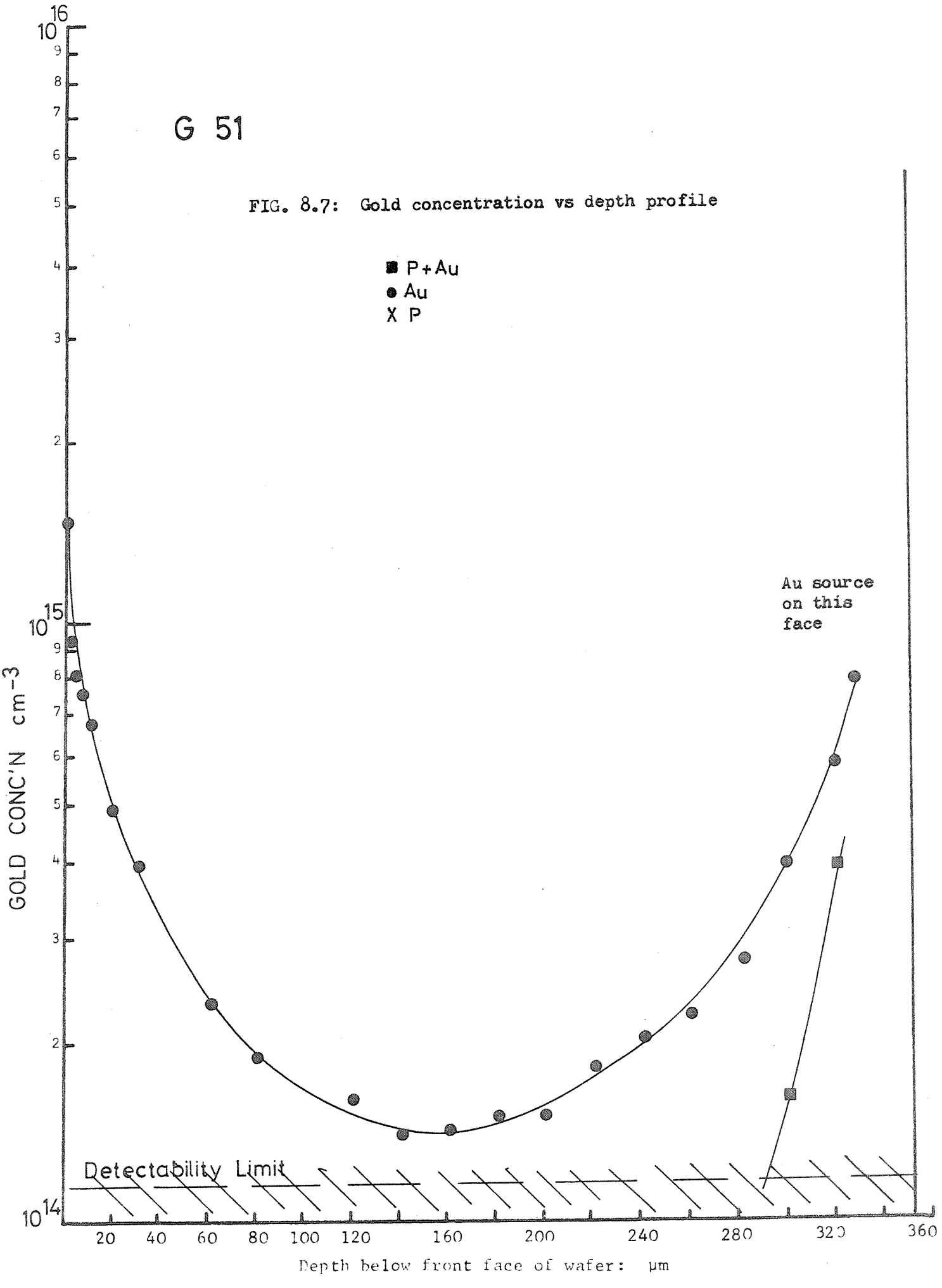


FIG. 8.7: Gold concentration vs depth profile

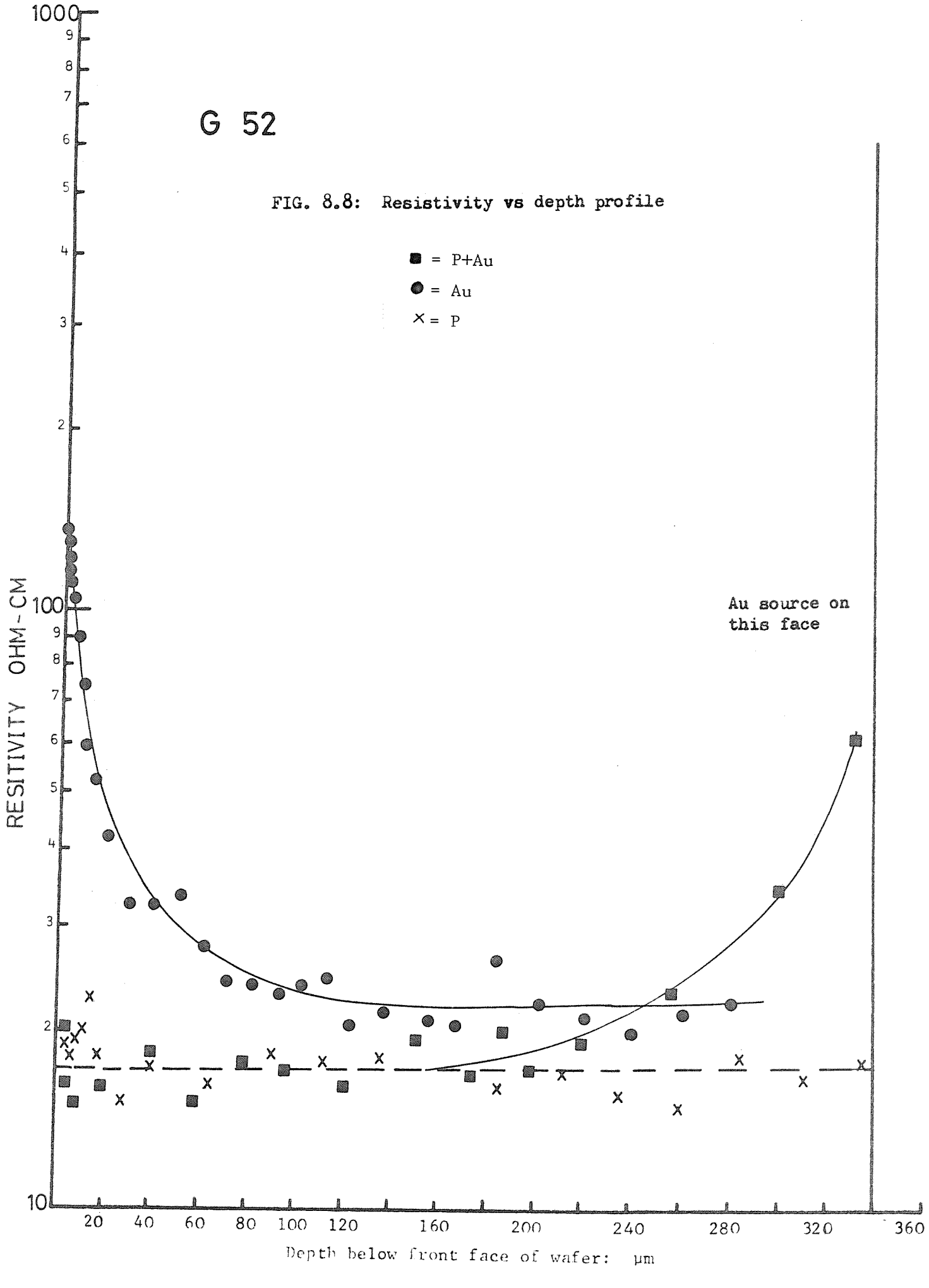


FIG. 8.8: Resistivity vs depth profile

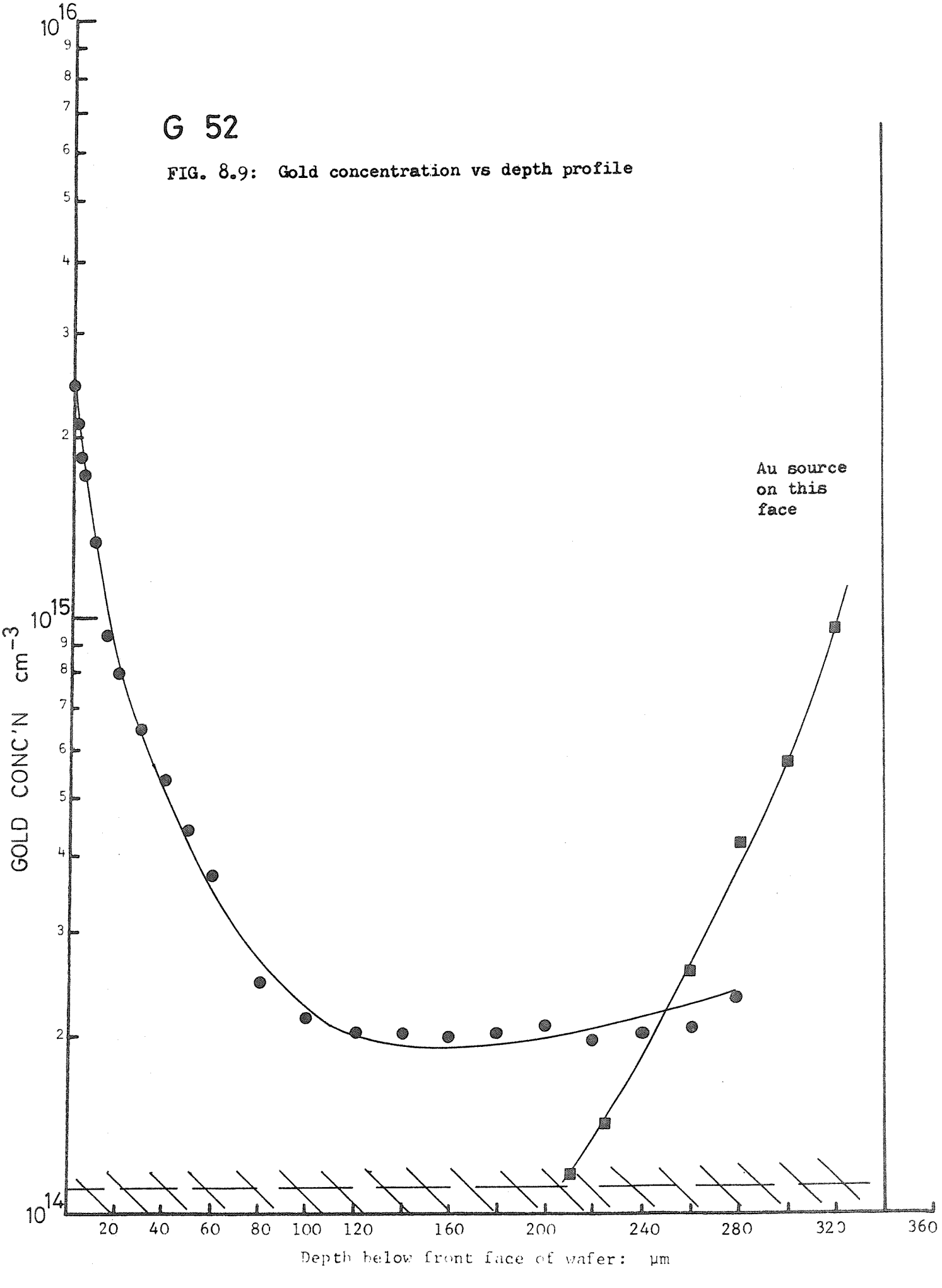
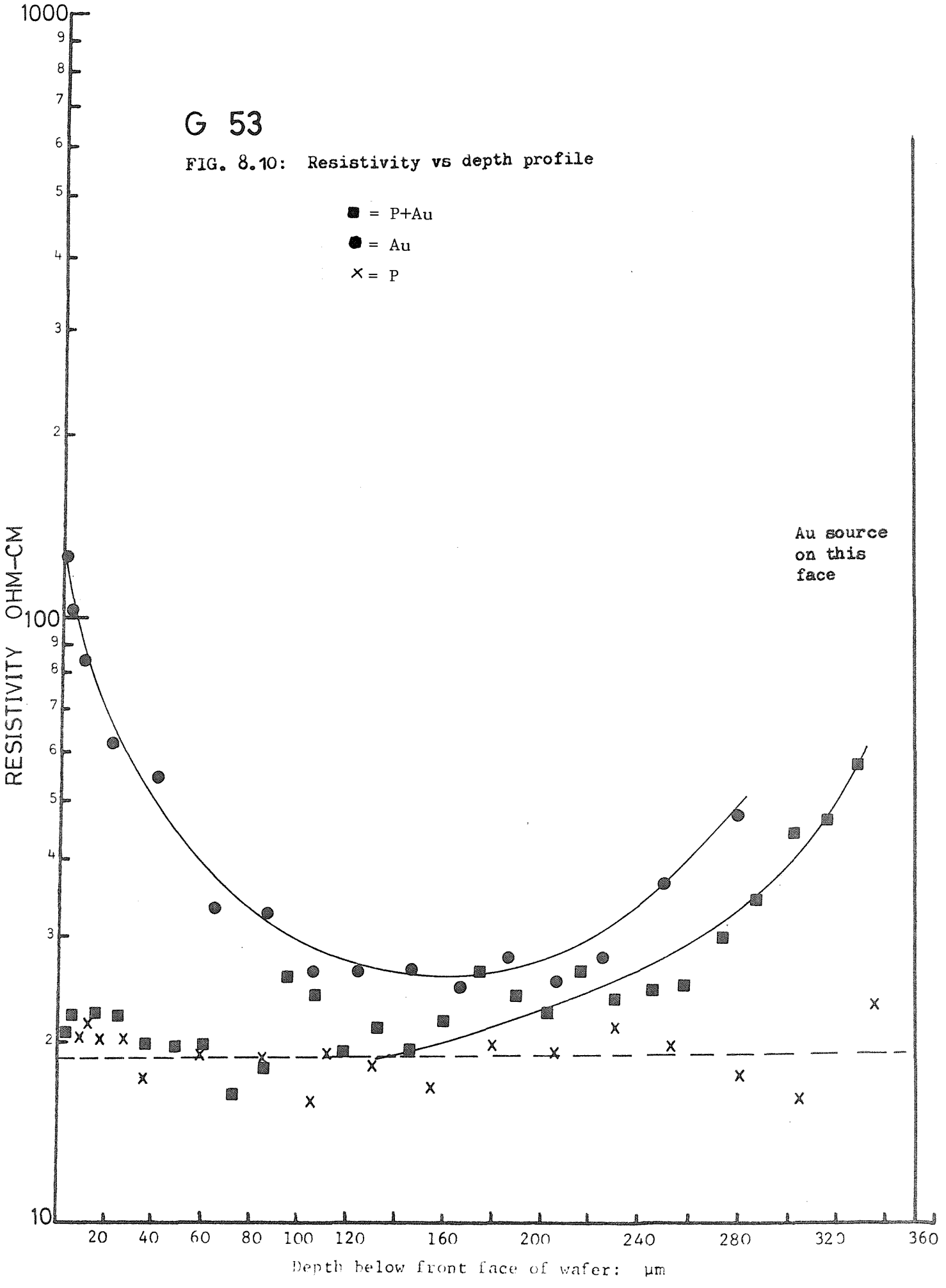


FIG. 8.9: Gold concentration vs depth profile


G 53
FIG. 8.10: Resistivity vs depth profile
■ = P+Au
● = Au
× = P
Au source on this face
RESISTIVITY OHM-CM
1000
100
10
20
40
60
80
100
120
160
200
240
280
320
360
Depth below front face of wafer: μm

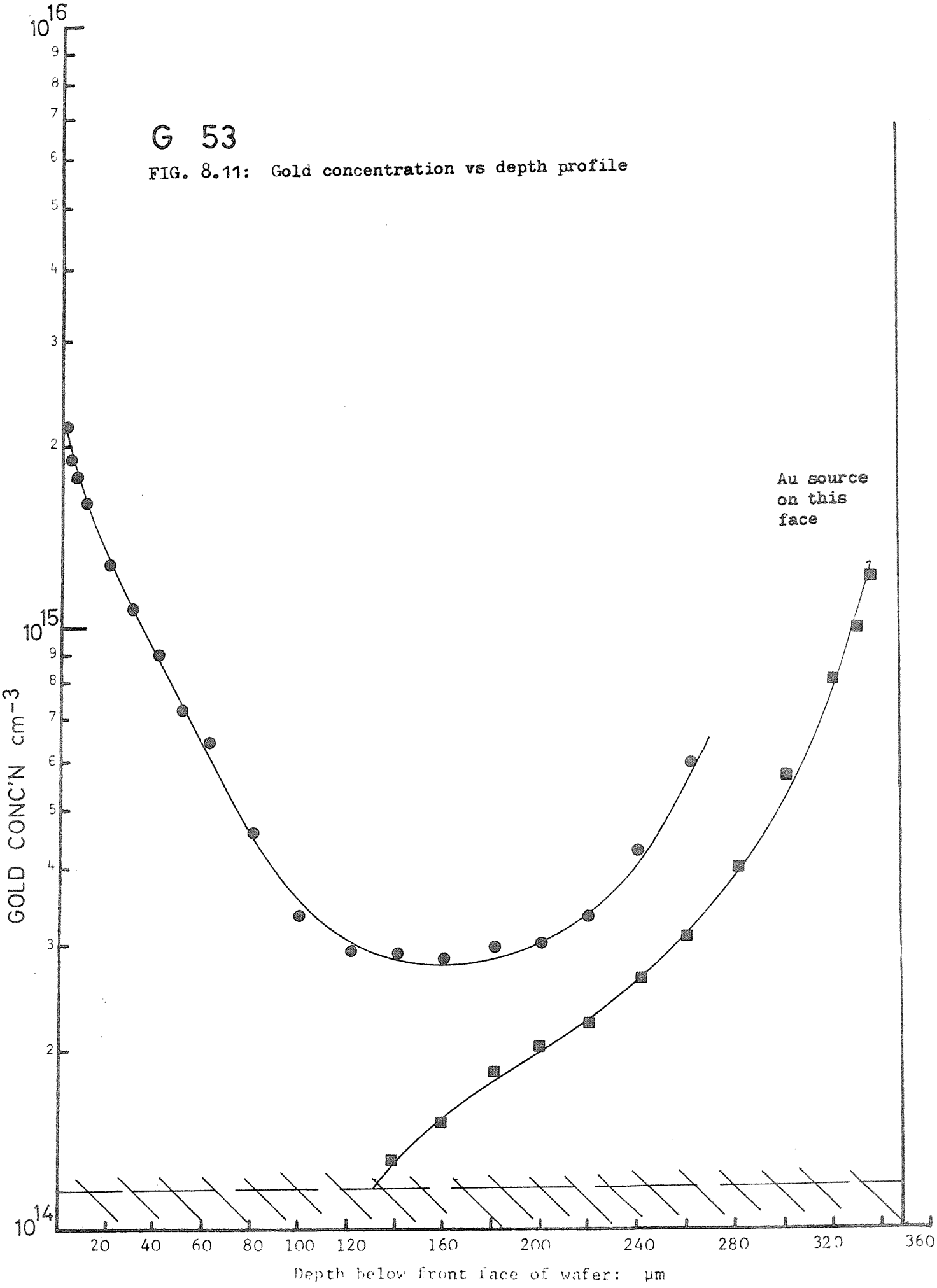
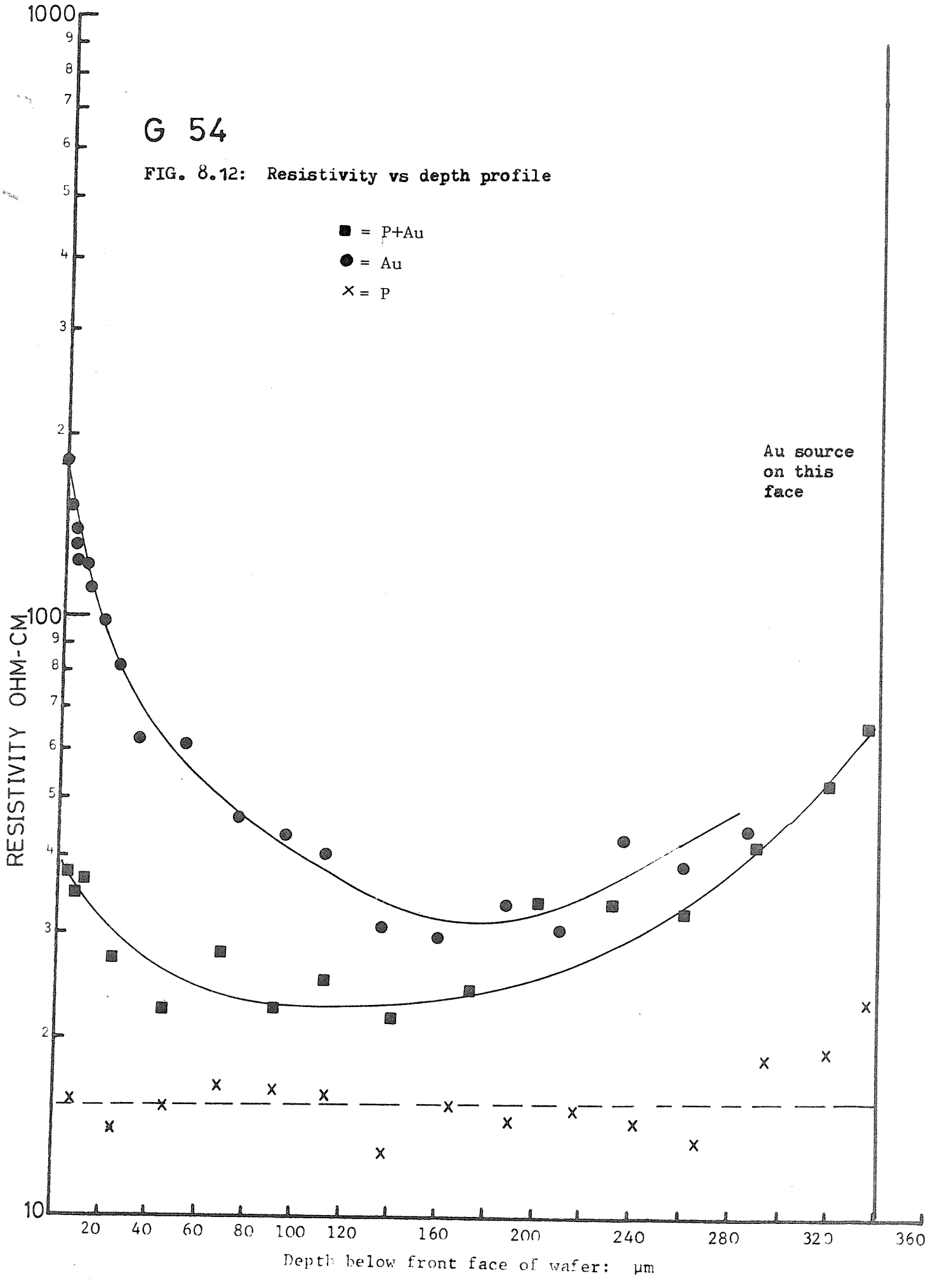


FIG. 8.11: Gold concentration vs depth profile


G 54
FIG. 8.12: Resistivity vs depth profile
■ = P+Au
● = Au
× = P
Au source on this face
RESISTIVITY OHM-CM
1000
100
10
20
40
60
80
100
120
160
200
240
280
320
360
Depth below front face of wafer: μm

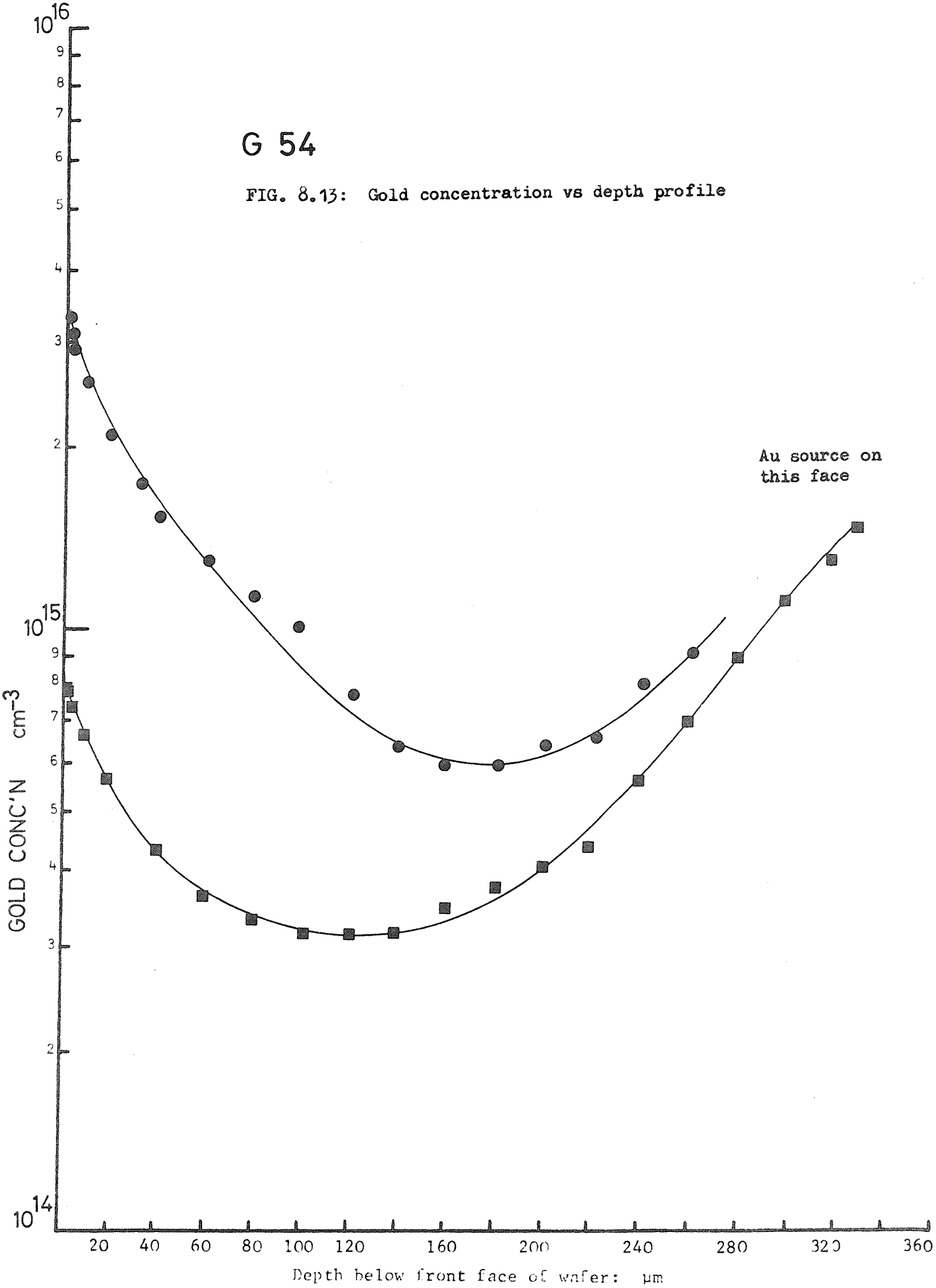
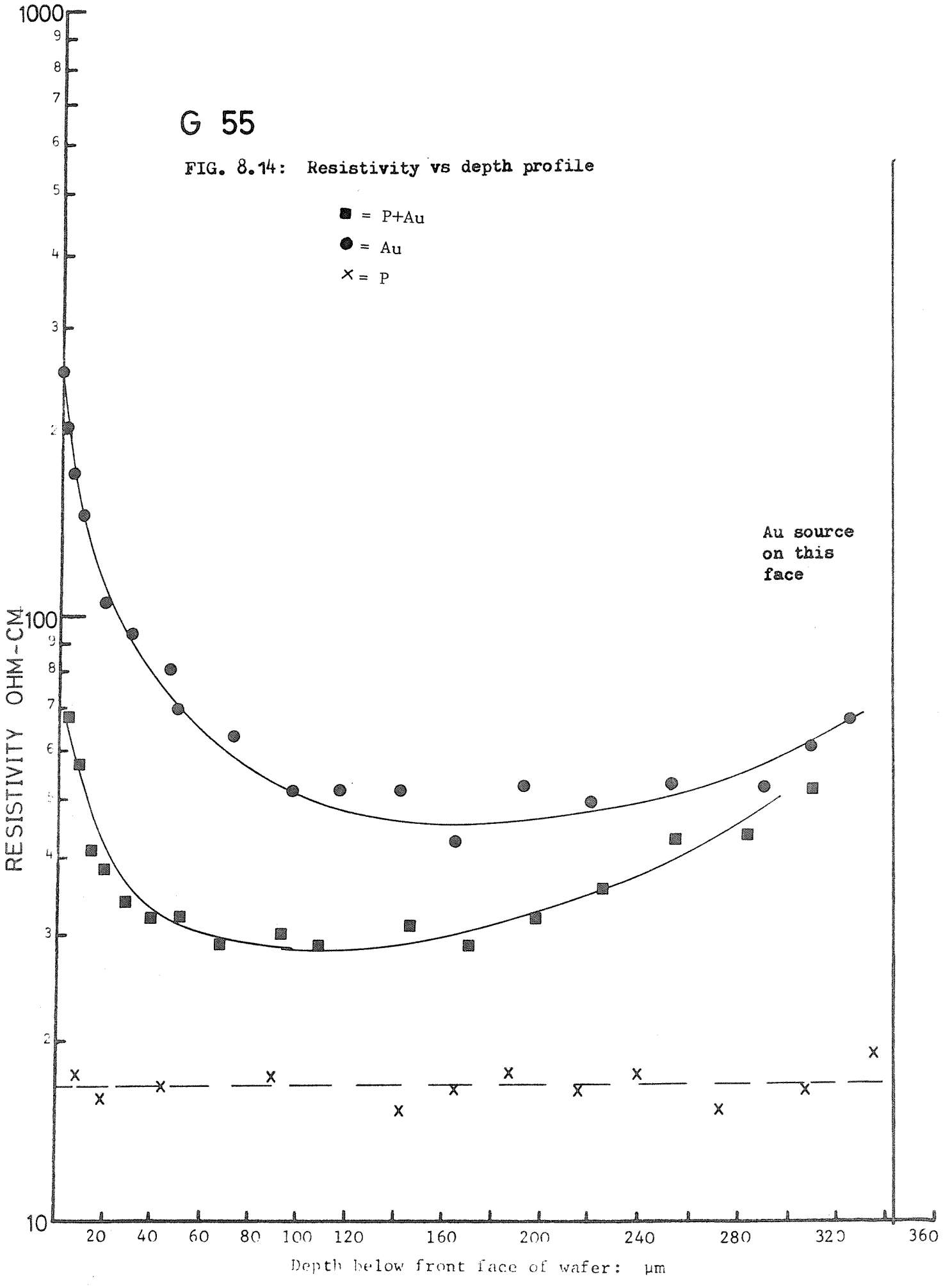


FIG. 8.13: Gold concentration vs depth profile


G 55
FIG. 8.14: Resistivity vs depth profile
■ = P+Au
● = Au
× = P
Au source on this face
RESISTIVITY OHM-CM
1000
100
10
20
40
60
80
100
120
160
200
240
280
320
360
Depth below front face of wafer: μm

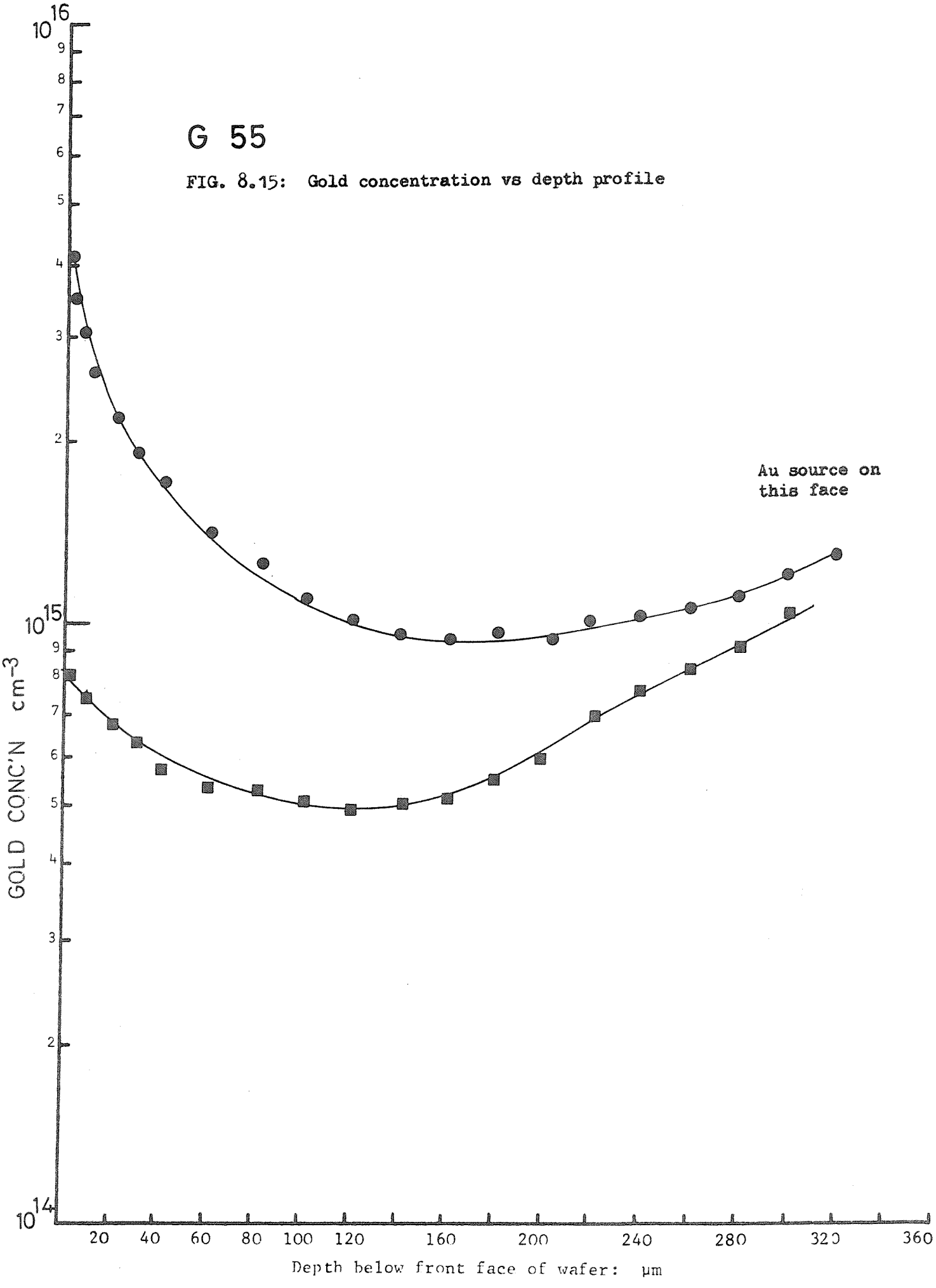


FIG. 8.15: Gold concentration vs depth profile

are considered further in section 8.7.5 and in chapter 9.

8.7.2 Gold-phosphorus sequential diffusions at 1000°C

This experiment was carried out in order to investigate the gettering effect of a phosphorus diffused layer on an already established gold distribution. This does not represent a true gettering situation where gold removal is the object because, during the phosphorus gettering diffusion, the gold source was still present on the back of the wafers, The results will indicate whether the newly diffusing-in phosphorus layer can getter faster than the gold can diffuse in from the back of the wafer.

A specially set up phosphorus furnace was used in this experiment to avoid gold contamination of the standard furnace.

Material : 10-20 Ω cm, p-type, (111) silicon wafers 250 microns thick. This material is designated, by the manufacturer, as dislocation free and was pulled by the Czochralski (CZ) technique.

Processing :

i Wafer cleaned.

ii Initial masking oxide 55 mins, wet oxygen, 1200°C.

iii Wafer quartered; quarters numbered 1-4.

iv Oxide removed from backs of 1 and 2.

v Au evporated onto backs of 1 and 2.

vi 1 and 2 annealed at 1000°C in gold furnace, 60 minutes in dry N_2.

vii 3 and 4 given equivalent heat treatment to vi

viii Oxide removed from fronts of 1 and 3.

ix 1 given phosphorus gettering diffusion in special furnace at 1000°C. (Different times for each run). Gas flows: as step vi in section 8.7.1.

x 3 given same treatment as step ix at 1000°C in standard phosphorus furnace.

xi 2 and 4 given equivalent heat treatments in gold furnace and control furnace respectively.

Results from two different runs are presented here :

G11 30 mins phosphorus getter, at step ix

G12 60 mins phosphorus getter at step ix.

Measurements: The quarters from each run were prepared for spreading resistance depth profiling. Unfortunately quarter number 3 (P only) for run G12 was broken. The spreading resistance probe used was the larger one so some smearing of sharp changes in resistivity must be expected in these results.

Results: Figure 8.16 shows the shallow phosphorus diffusion profile for run G11, quarter number 3. The spreading resistance - depth profiles in quarters 1, 2 and 4 of G11 are plotted in figure 8.17. The following symbols are used :

● = quarter no. 1 (P+ Au)

■ = quarter no. 2 (Au only)

X = quarter no. 3 (P only)

Since the effect of increased gold concentration is to raise the resistivity and hence the spreading resistance, comparison between the profiles indicates the relative amounts of gold in the two gold doped quarters. Despite the presence of an infinite source of gold at the back of the wafer, the gettering effect of the shallow phosphorus diffused layer is quite marked, indicating that the rate of gettering is greater than the rate of diffusion of gold from the gold source. In fact, the interstitial gold concentration in the vicinity of the phosphorus diffusion has been depressed far enough below its local equilibrium with substitutional gold to drive substitutional gold atoms into interstitial sites - hence the dip in the measured gold profile which is that of substitutional gold only.

Figure 8.18 shows similar profiles for run G12, where the effect of gettering for longer is even greater. This indicates that the rate of input of gettering centres (phosphorus-vacancy pairs?) at the front of the wafer is higher than the rate of take up of gold (gold interstitial diffusion rate limited) by the layer.

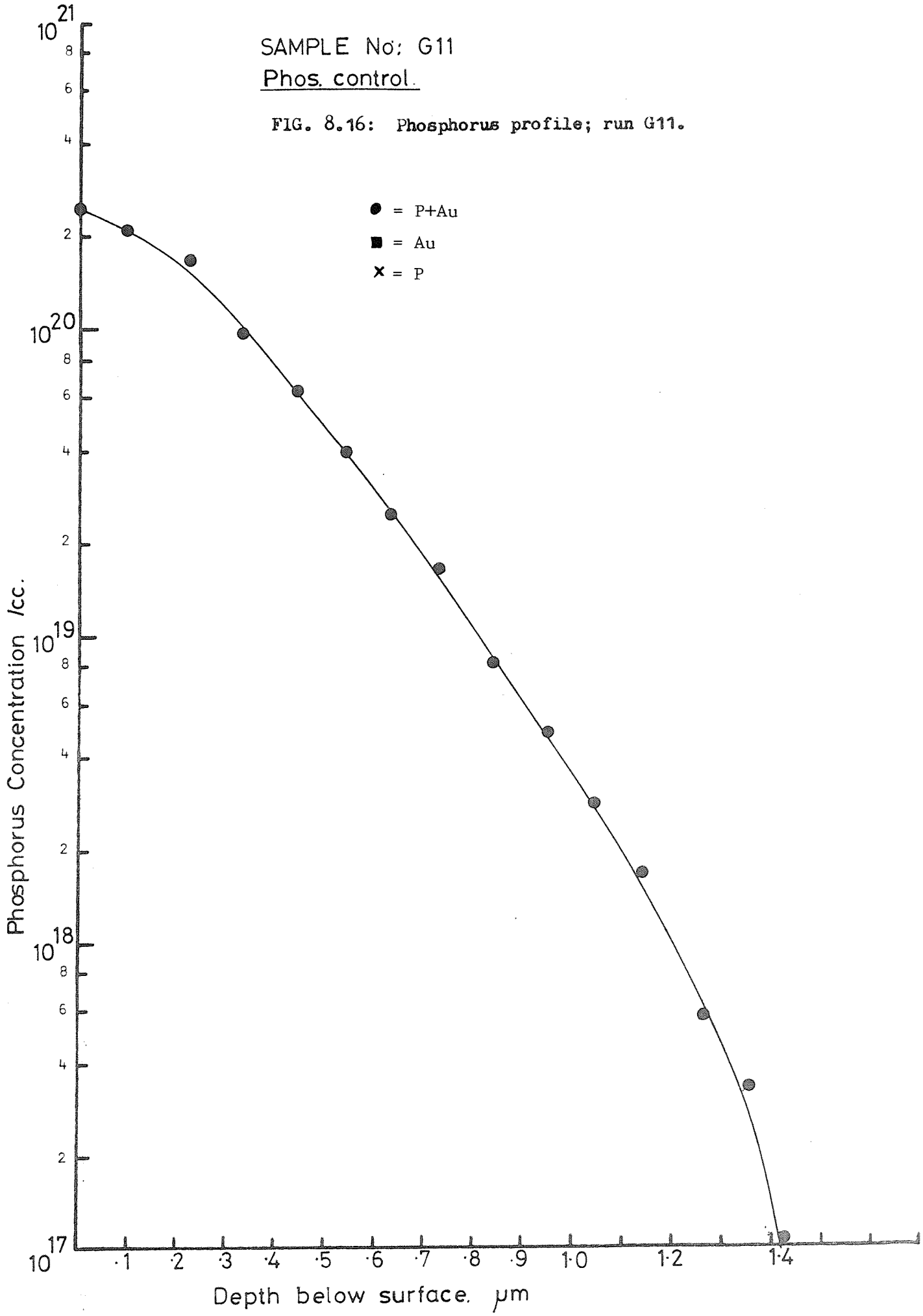


FIG. 8.16: Phosphorus profile; run G11.

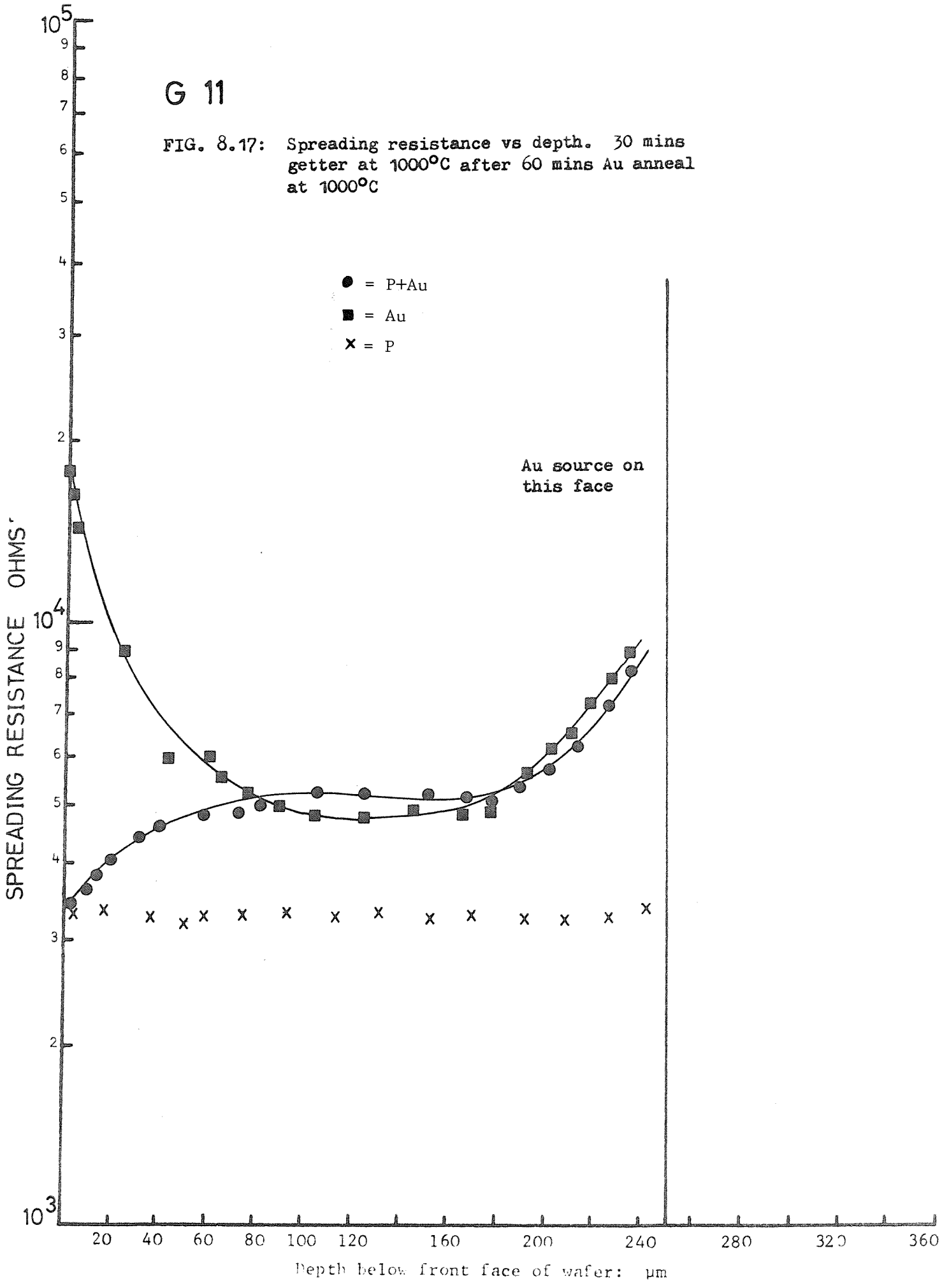


FIG. 8.17: Spreading resistance vs depth. 30 mins getter at 1000°C after 60 mins Au anneal at 1000°C

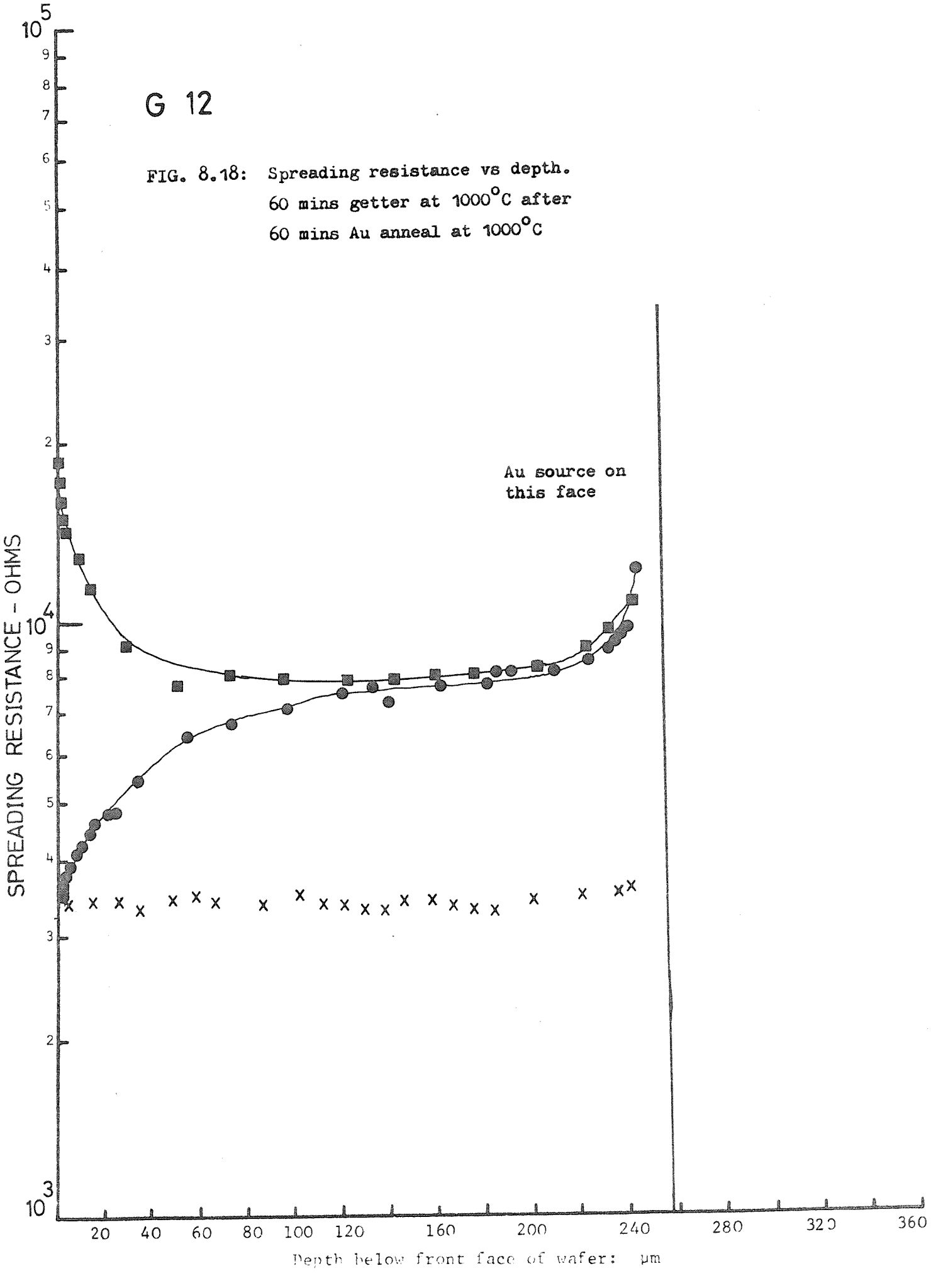


FIG. 8.18: Spreading resistance vs depth.
60 mins getter at 1000°C after
60 mins Au anneal at 1000°C

8.7.3 Attempted gettering of a limited amount of gold

In order to study a realistic gettering situation, attempts were made to getter a fixed total quantity of gold with a shallow phosphorus diffused layer. This experiment was similar to that described above except that removal of the gold source before the gettering treatment was tried. After the gold diffusion (step vi above) the gold source is a gold-silicon alloy on the back of the wafer. This had to be removed by a technique which did not cause lattice damage to the back of the wafer since any such damage would itself getter the gold (Nakamura et al. 1968) and result in a very uncontrolled experimental situation.

Removal of the gold source was attempted by subjecting the sample to a silicon etch after masking the front face with Apiezon wax. After etching, the wax was removed with trichloroethylene and the back of the wafer was protected against further contamination with a layer of 'spin-on' silicon oxide (spin-on oxides and doped oxides are commercially available as diffusion masks and diffusion sources). The silicon etches were based on mixtures of HF, HNO_3 and acetic acid in various proportions. (The addition of acetic acid to the HF:HNO_3 mixture produces an etch which results in a fairly well polished silicon surface). Amounts of silicon ranging from 10 to 50μm were removed before the wafers were subjected to a second annealing step to see if the gold source had been removed efficiently. In almost all cases the anneal resulted in a further significant rise in resistivity reflecting an increased gold concentration. This occurred even when the etched wafers were subjected to repeated boils in aqua regia before application of the spin-on glass to remove any gold back plated from the etching solution (Brown et al, 1975). Figure 8.19 illustrates a typical result. The profile of a sample before etching is compared with the profile of another part of the same sample after etching and a further 60 minute anneal at 1000°C in dry N_2. This is one of ten experiments carried out in which the results were extremely variable.

In view of these difficulties, this experiment was abandoned. A possible solution would have been to diffuse the gold in from a vapour-phase source (such as gold chloride in an evacuated sealed capsule -

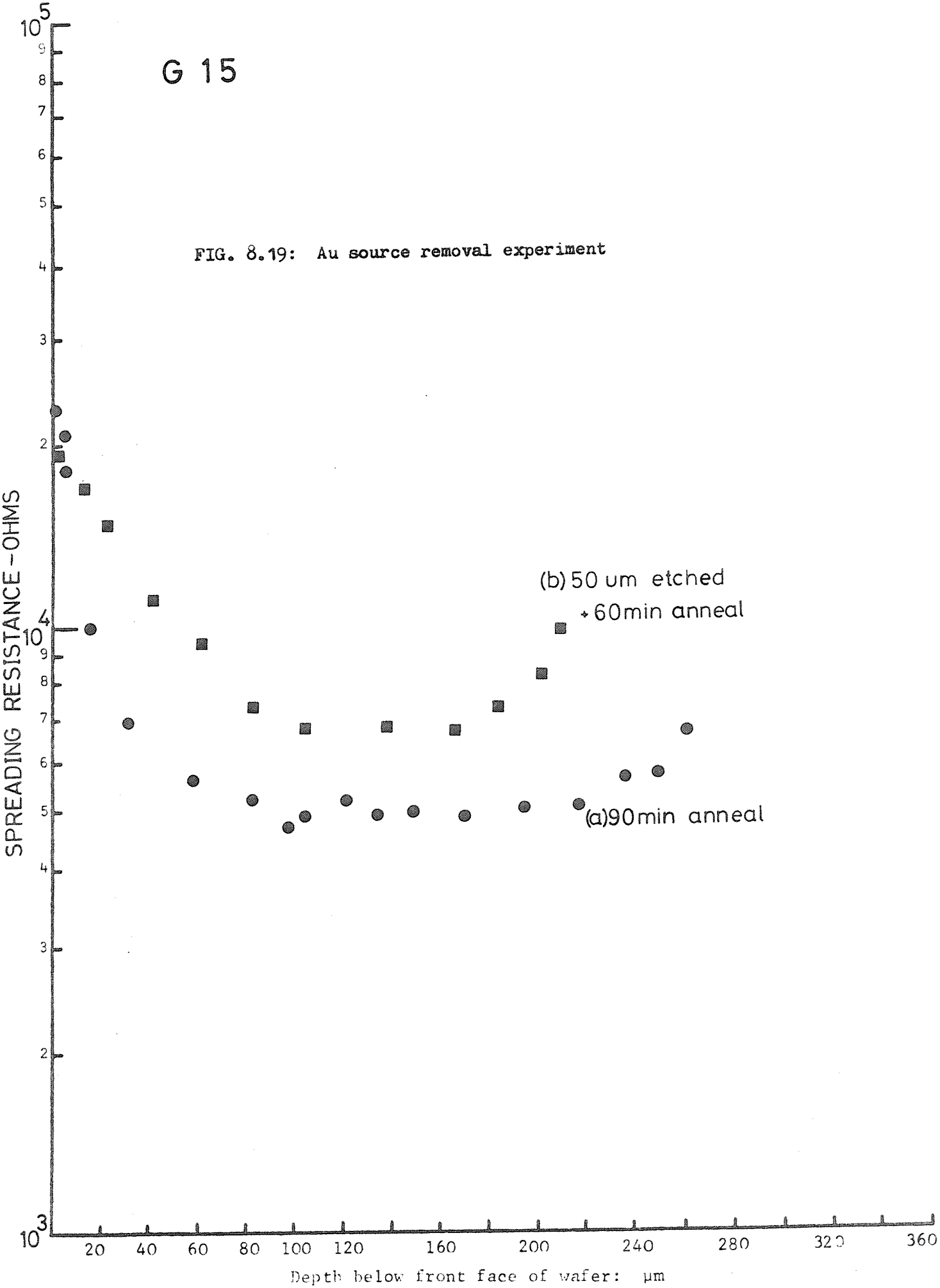
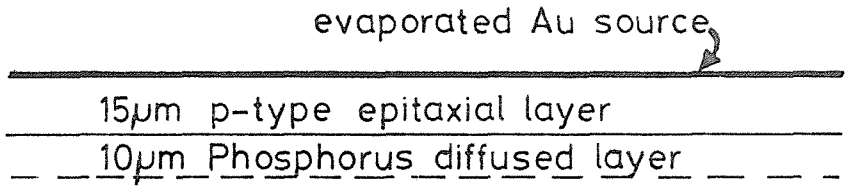


FIG. 8.19: Au source removal experiment

Wolley et al, 1967) to prevent formation of gold rich phases at the wafer surfaces.

8.7.4 Penetration of phosphorus doped layers by gold at 1000°C

The diffusion of gold through a heavily phosphorus doped layer was investigated in this batch of experiments. The structure sketched below was employed.



The phosphorus doped region is under an epitaxial layer in much the same way as buried collector structure in planar bipolar transistors. The 15 micron epitaxial region was intended as a buffer between the gold source (which forms a gold-silicon alloy with the silicon surface) and the phosphorus doped region. Annealing of this structure resulted in diffusion of gold through the thin epitaxial layer into the phosphorus doped layer. Subsequent profiling of the wafer bulk by the spreading resistance technique yielded information about the rate of penetration of this layer by the gold.

Material : 10-20 Ωcm, p-type, (111) silicon wafers 250 microns thick. This (CZ) material is designated by the manufacturer as dislocation free.

Processing :

i Wafer cleaned.

ii Initial masking oxide : 55 mins wet oxidation 1200°C.

iii Wafers halved.

iv Oxide removed from front of only half of each piece

v Phosphorus deposition on both halves (note from iv that only half of each piece is unmasked). 1050°C, 30 minutes.
Gas flows: as step vi in section 8.7.1.

vi Phosphorus drive-in : 200 minutes in dry O_2 at 1200°C.

vii All oxide removed from fronts of both halves.

viii Approximately 15 microns, p-type, epitaxial silicon grown on the fronts of the wafers.

ix Au evaporated onto front of one half.

x Piece in ix annealed in Au furnace for appropriate time in dry N_2 at 1000°C.

xi Other half given equivalent anneal to x in control furnace.

Four different anneal times were employed:

Run G21	15 minutes
Run G22	30 minutes
Run G23	60 minutes
Run G24	120 minutes

Measurements: The samples were cut up and profiled by the spreading resistance technique using the high resolution probe.

Results: Figure 8.20 illustrates the spreading resistance versus depth profile through one of the non gold doped control samples. The rise in spreading resistance in the epitaxial layer is caused by the proximity of the p-n junction to the probe (see appendix A) and is an artifact of the measurement. The very low spreading resistance values obtained in the phosphorus doped layers - indicative of heavy doping - are shown in the inset. The uniform spreading resistance versus depth profile in the substrate is as expected.

Since the penetration of gold into the substrate is the item of interest here, the spreading resistance profiles in the substrate alone

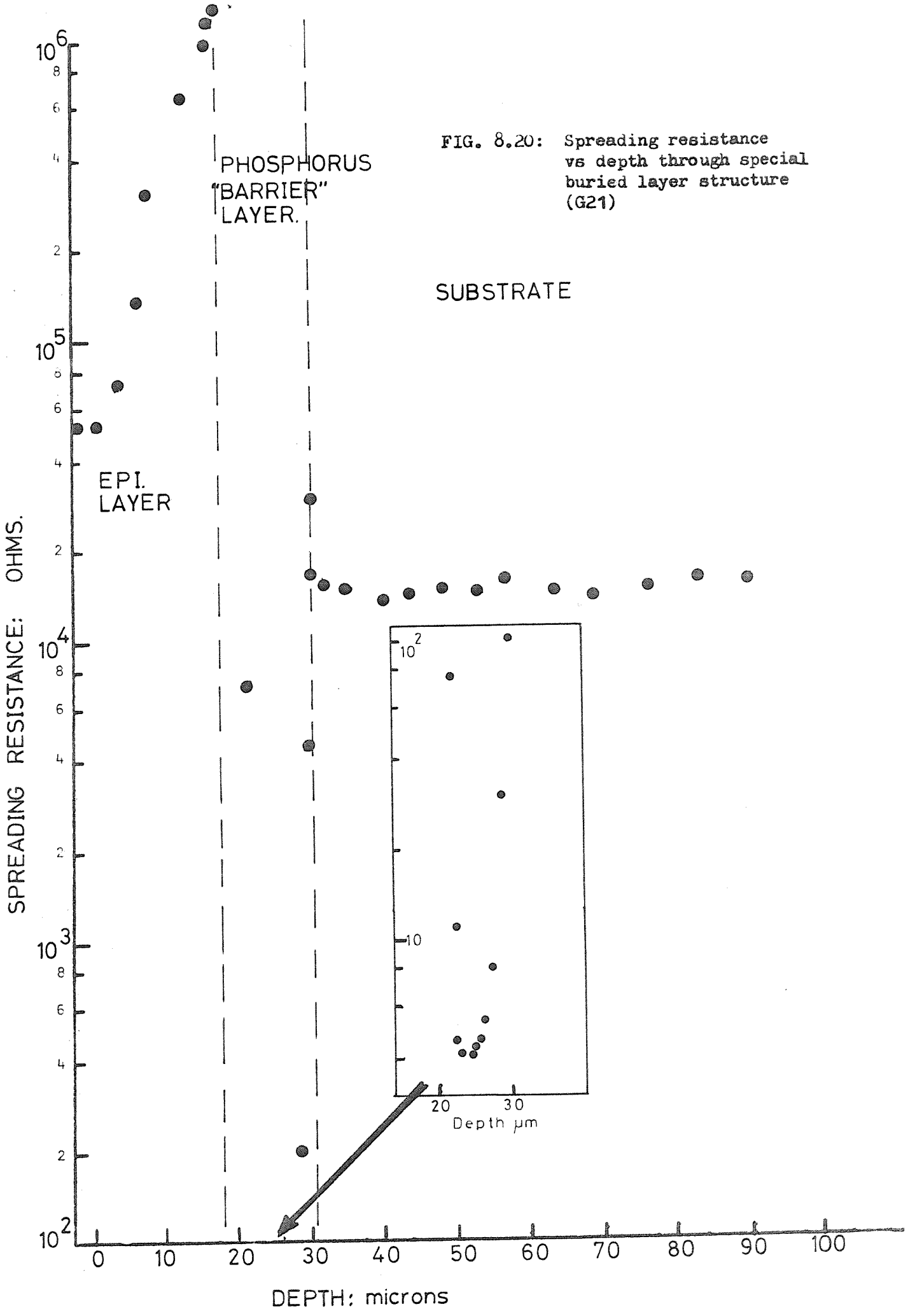
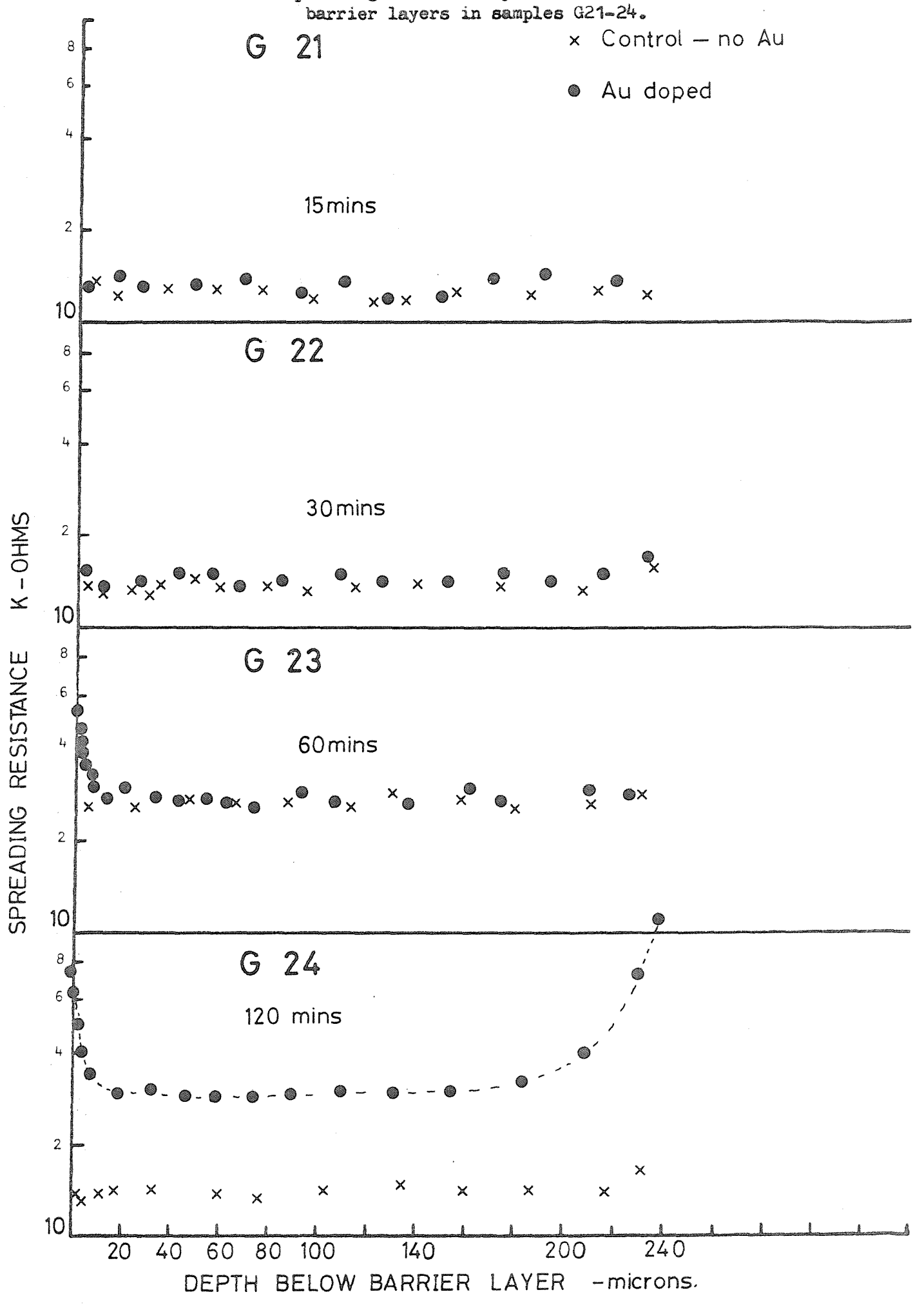


FIG. 8.20: Spreading resistance vs depth through special buried layer structure (G21)

FIG. 8.21: Spreading resistance profiles below barrier layers in samples G21-24.



are illustrated in figure 8.21 for the non gold-doped control samples and for the gold doped samples. Runs G21-24 are all illustrated in this figure.

It is apparent from these profiles, that penetration of the phosphorus buried layer is not measurable until approximately 60 minutes annealing time. For this time (G23) there is evidence of a small amount of gold close to the phosphorus layer and after 120 minutes gold is detected throughout the substrate. These results indicate that, either the gold diffusion rate is significantly slowed down in the phosphorus layer (as suggested by Lambert and Reese, 1968) or that for a period the conversion rate of interstitial to substitutional gold is so high that it effectively prevents any gold from reaching the substrate. This latter interpretation is favoured since, as will be shown in chapter 9, the phosphorus layer would be saturated with gold after a period between 30 and 60 minutes and then present no barrier to interstitially diffusing gold. This appears to be the case when G23 and G24 are compared since the level of contamination in G24 is roughly that expected after 60-90 minutes of gold diffusion at 1000°C. If the barrier layer were merely slowing the interstitial gold down, this rate of rise would not be expected.

8.7.5 Comparison of gold diffusion profiles in defect free and non-defect free silicon wafers at 1000°C

The diffusion of gold in thin silicon wafers in the absence of shallow diffused layers has been discussed in chapter 3. There exists some disagreement about the rate limiting step in the rise of gold concentration as a function of time. Huntley et al. (1970 and 1973) suggested that there should be significant differences between gold profiles in dislocation free wafers and gold profiles in heavily dislocated wafers if the process $Au_i + V \rightarrow Au_s$ was vacancy (V) supply limited. Experimentally however, much smaller differences than predicted were observed. Similar results obtained by Brotherton et al. (1972) led them to suggest that the reaction between Au_i and vacancies was the rate limiting step. Huntley (1972) does suggest, however, that the dislocation free material may have contained sources of vacancies (other than dislocations) in the defects known as the swirl pattern (see chapter 9 for a more detailed discussion of the nature of the

swirl pattern).

Recently, silicon wafers which are specified as zero-dislocation and swirl free have become available and in this section a comparison of profiles obtained in such wafers with profiles obtained in dislocated wafers is presented.

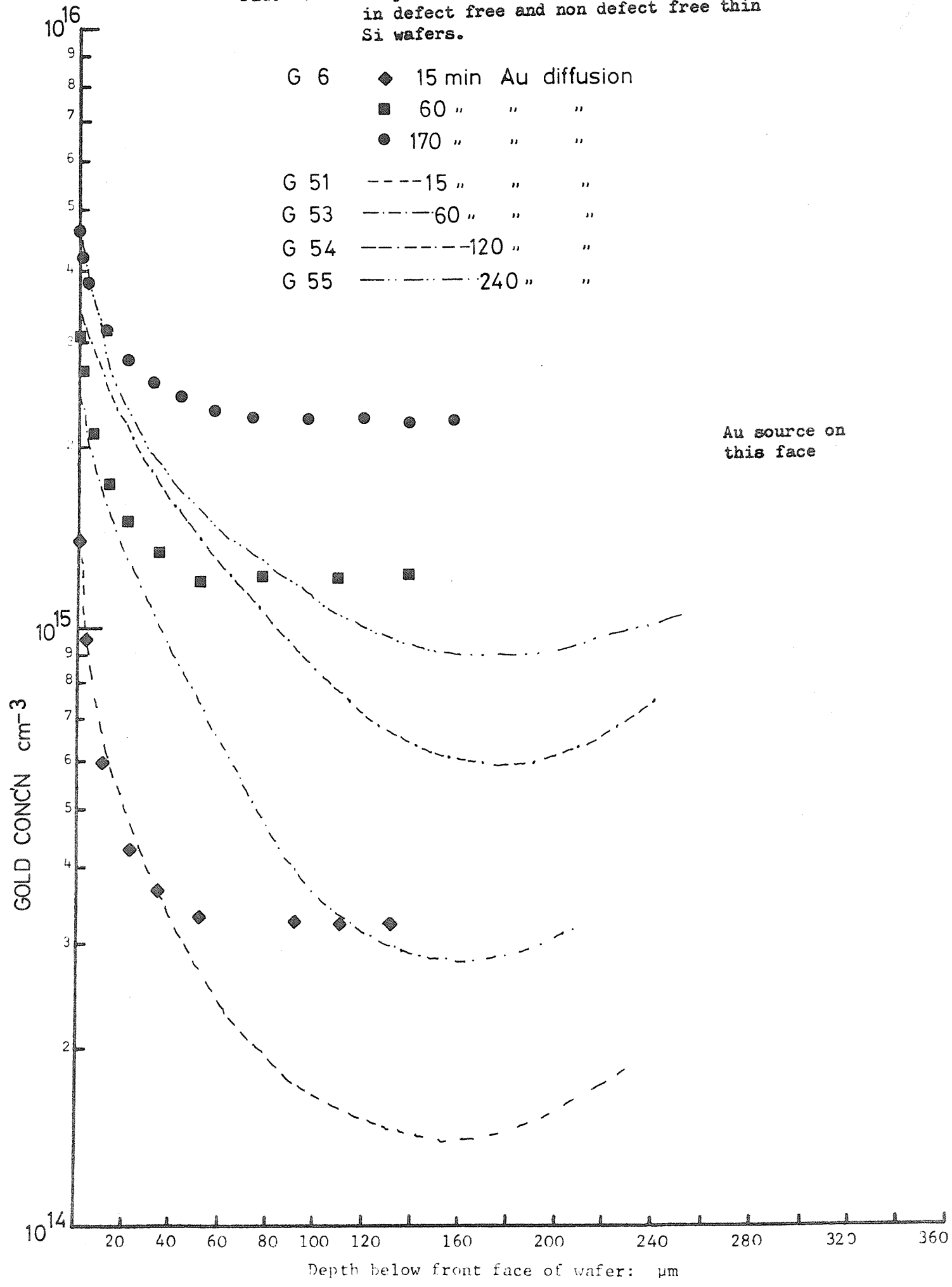
The gold profiles in the defect free wafers are taken from the experiments detailed in 8.7.1. (Runs G51-55). The other profiles were obtained from an old 250 micron thick, FZ, p-type, (111) silicon wafer. of 10-20 Ω cm. This wafer exhibited dislocation etch pits when SIRTL etched (Sirtl et al. 1962), suggesting a dislocation density of 10^2-10^3 per cm^2. The wafer was processed in a similar manner to the gold-only controls of G51-55.

Processing :
i Wafer cut into four quarters.
ii Initial masking oxide, 55 mins wet oxygen, 1200°C.
iii Oxide removed from backs of three quarters.
iv Au evaporated onto the backs of these quarters.
v Au anneals at 1000°C in dry N_2 for 15 mins, 60 mins, and 170 mins.

Measurements: These samples were measured by the spreading resistance technique and the gold profiles inferred from the measurements by comparison with the resistivity of the fourth quarter of the wafer. The non gold doped quarter was SIRTL etched. The larger, lower resolution spreading resistance probe was used for these measurements.

Results: The profiles obtained in this experiment, together with those obtained on runs G51-55, are shown in figure 8.22. The effect of the lower resolution spreading resistance probe would be to raise the values of gold concentration close to the surfaces of the samples; the bulk profiles would not be affected appreciably. The differences between the two materials is very marked, and tends to support the Huntley et al, model of vacancy supply rate limitation. The reasons for this are detailed in chapter 9.

FIG. 8.22: Comparison between gold distribution in defect free and non defect free thin Si wafers.



8.7.6 Rutherford backscattering measurements of gold in shallow phosphorus diffused layers

RBS measurements were made on the 'phosphorus + gold' doped samples of runs G51-54 to obtain information about the gold concentration within the phosphorus doped layers.

The samples were cleaned by removing the surface oxide in buffered HF and boiling the samples in aqua-regia. The aqua regia boil was alternated with HF dips five times and after the sixth aqua regia boil the samples were washed in double distilled water and dried. This extensive clean was to ensure that any accumulation of gold on the wafer surface was removed since the RBS technique is very sensitive to surface contamination.

The spectra obtained are shown as yield versus channel number (→energy of backscattered ions) in the following figures :

Figure 8.23: G51, 12.0μC dose. Sample aligned in slightly channeling direction.

Figure 8.24 : G52, 12.0μC dose. Sample aligned at random

Figure 8.25 : G53, 12.0μC dose. Sample aligned at random

Figure 8.26 : G54, 12.0μC dose. Sample aligned at random

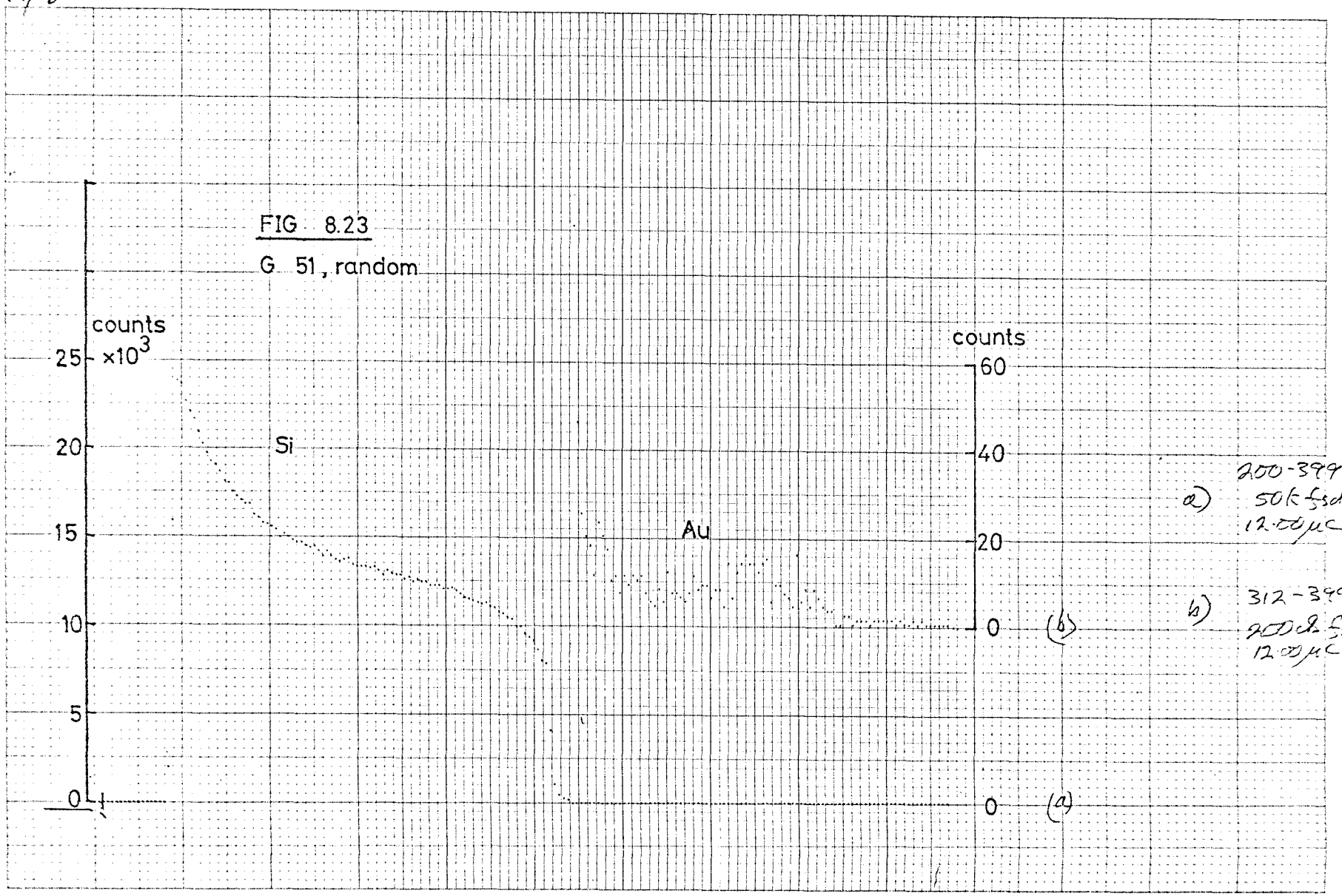
Figure 8.27 : G54, 16μC dose. Sample aligned in <110> channeling direction.

It should be noted that two different scales are used on each of these diagrams : 500 counts per small division for the silicon part of the spectrum and 2 counts per small division for the gold part of the spectrum on 8.23 - 8.26, and 200 counts per small division in the silicon spectrum on Figure 8.27.

Study of these spectra yields the following information: for G51, the rather curved edge of the silicon spectrum suggests that the sample was not aligned in a true random orientation to the H_e^+ beam. The channeling so caused is minimal and probably does not affect the results very much.

Comparison of figures 8.23, 24, 25 and 26 shows a steadily increasing amount of gold close to the surface of the silicon. There is quite a lot

3/5/77/8
TU
N°1.

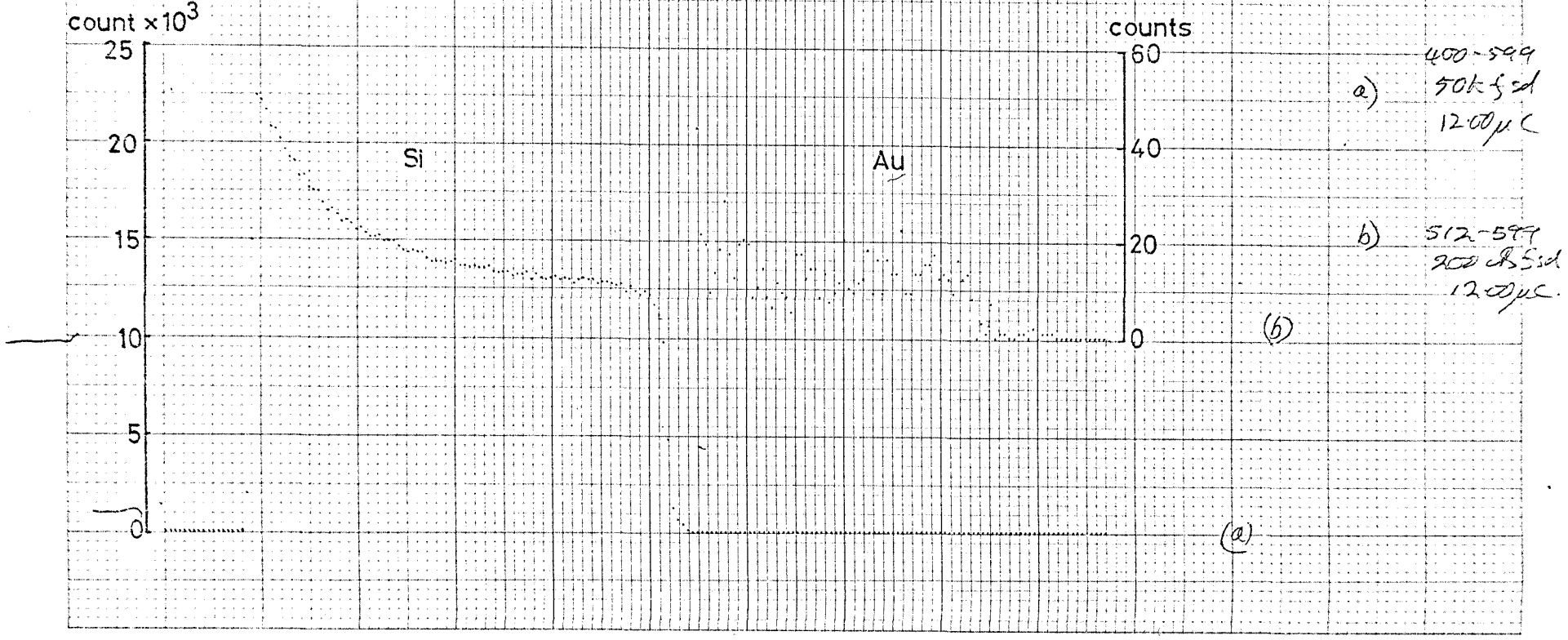


SENSITISED COATINGS LTD., COULSDON, SURREY

3/5/77/9
TU
N°2

FIG 8.24

G 52, random



a) 400-599
50k 5cl
12·00μC

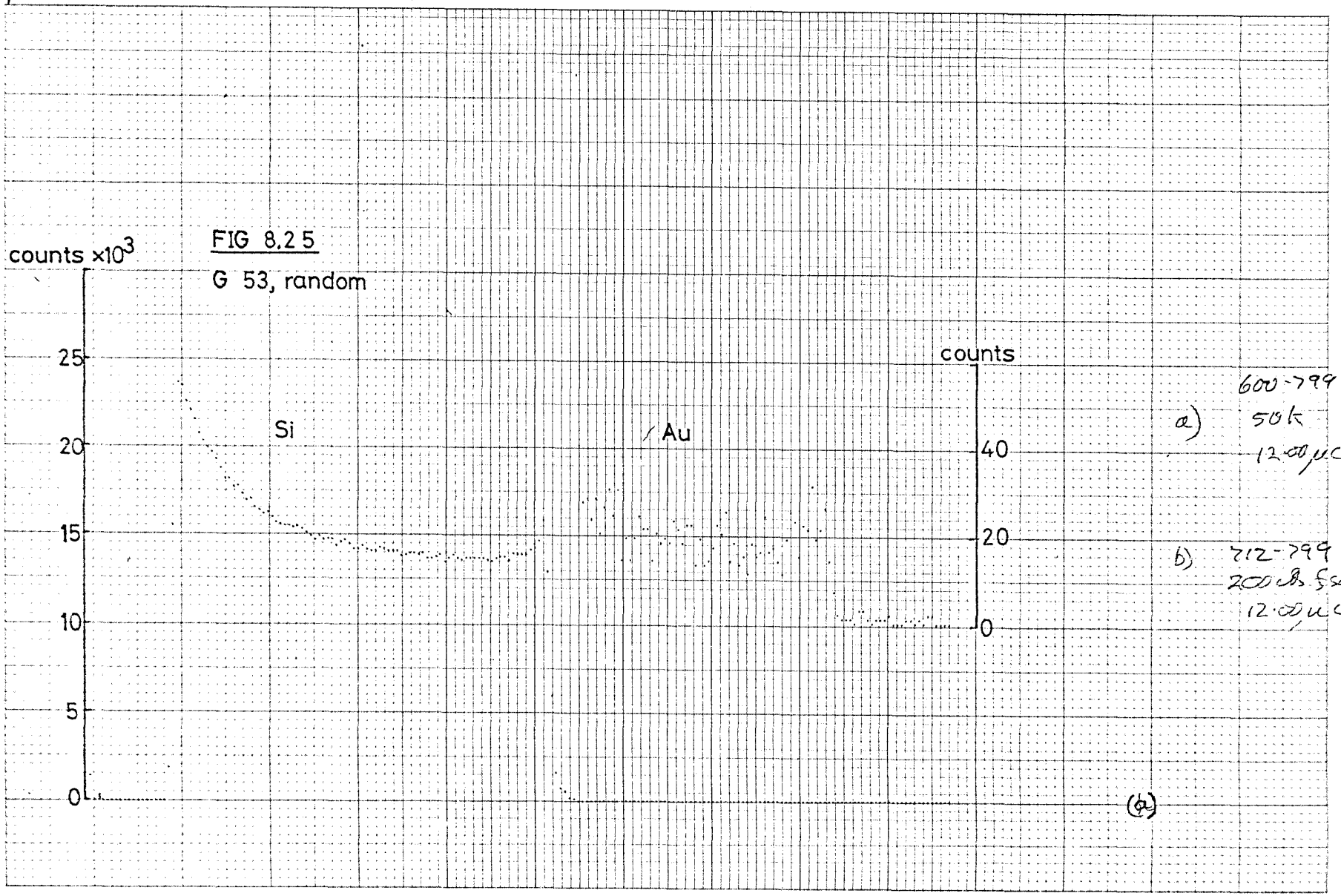
b) 512-599
200 ch 5cl
12·00μC

SENSITISED COATINGS LTD., COULSDON, SURREY

3/5/77/10

TU

N°3



SENSITISED COATINGS LTD., COULSDON, SURREY

3/5/77/0

N°4

T.U.

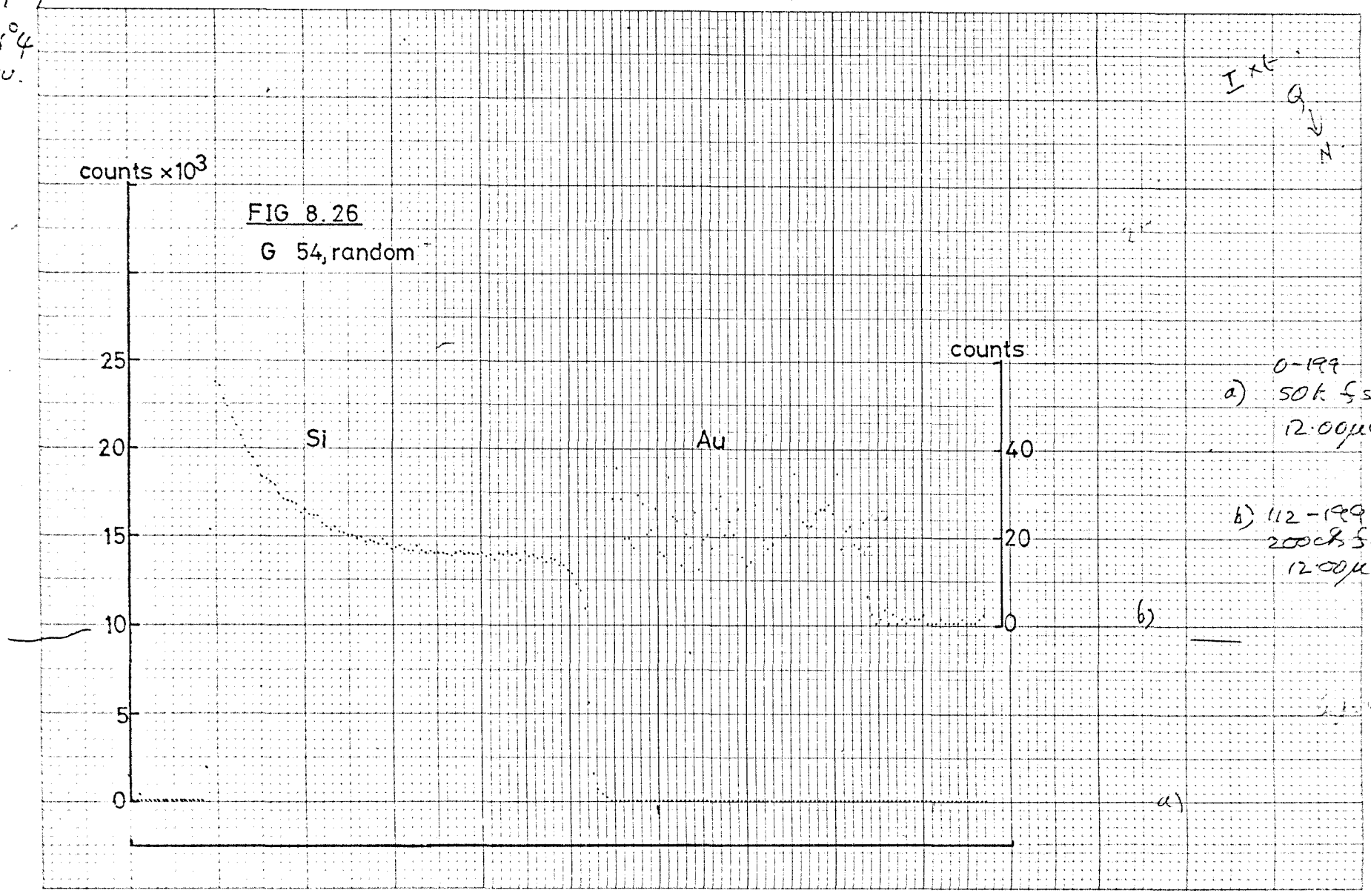


FIG 8.26

G 54, random

I x t

Q

N

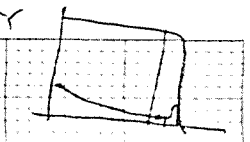
a) 0-199
50k f.s.d
12.00μC.

b) 112-199
200ck f.s.d
12.00μC.

b)

a)

4/5/77/1
N°4 TU·
channelling
L107

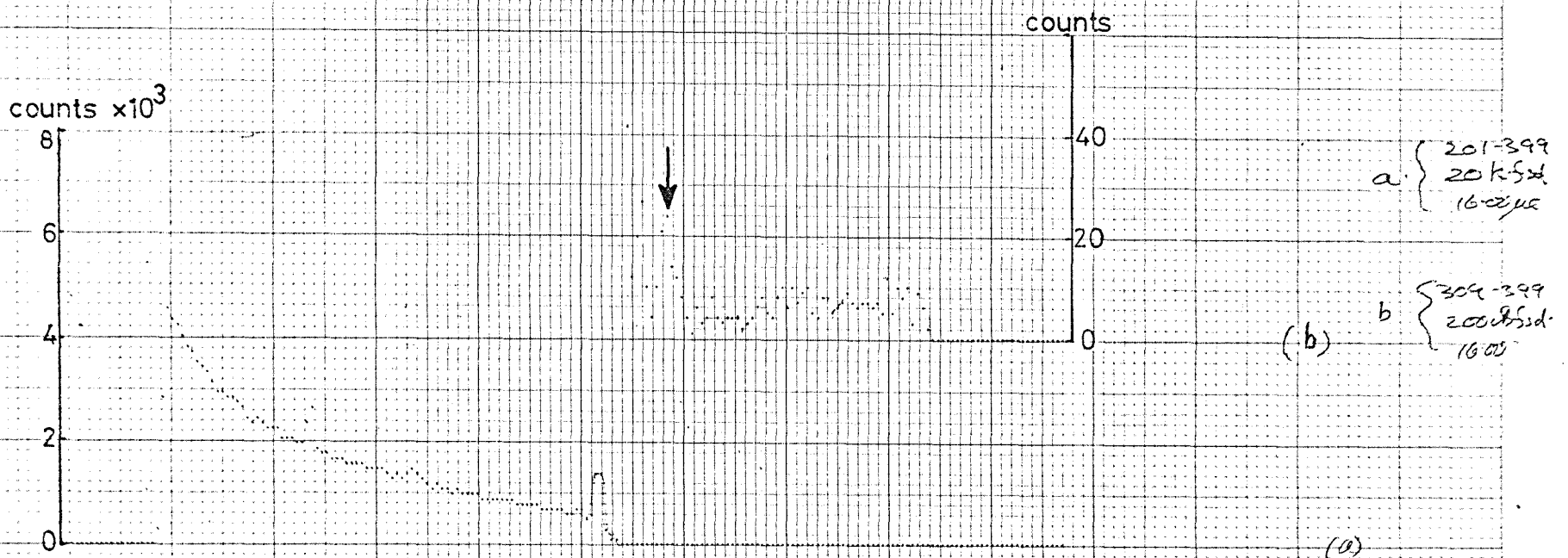


$\chi_{min} = \frac{0.6}{8.0} \times 100\ \%$

$= 7.5\%$ -

FIG 8.27

G 24, channelling.



SENSITISED COATINGS LTD., COULSDON, SURREY

of scatter in the gold counts but the spectra are fairly flat overall, suggesting that within the depth range resolved (3000 Å) the gold concentration profile is flat. The corresponding phosphorus profile over the same depth (figure 8.5) is also fairly flat so this result is not surprising.

Mean values of the gold yield were taken from the print out corresponding to these spectra and equation 8.2 was solved to give the following surface concentrations of gold :

G51	7.7×10^{17}	gold atoms/cc	(15 mins gold diffusion)
G52	1.6×10^{18}	gold atoms/cc	(30 mins gold diffusion)
G53	2.2×10^{18}	gold atoms/cc	(60 mins gold diffusion)
G54	2.9×10^{18}	gold atoms/cc	(120 mins gold diffusion)

The gold profiles from G54 (figure 8.13) indicate that the phosphorus layer was probably saturated with gold. Time did not permit an RBS study of G55, which may have confirmed this. The value of gold concentration from G54 is used in chapter 6, table 6,3, where the model proposed agrees well with this measured value.

The channeling spectrum from G54 has two interesting features. The most important is the amount of interstitial gold. The yield for the 16μC dose of He^+ is 7, which, when compared to the yield for the 12μC dose in the random direction, indicates that 80% of the gold measured in the first 3000Å depth of G54 is on substitutional lattice sites. Any definite conclusion about the significance of this must be made while bearing in mind that substitutional gold could well move into interstitial sites during the cooling of the sample.

The second feature of interest is the small peak in the backscattered spectrum at the position arrowed on figure 8.27. This corresponds to the presence of chlorine on the surface of the sample. As already stated, the samples were cleaned by boiling in aqua regia which contains HCl. The washing step carried out after cleaning was obviously not adequate to remove all traces of aqua regia.

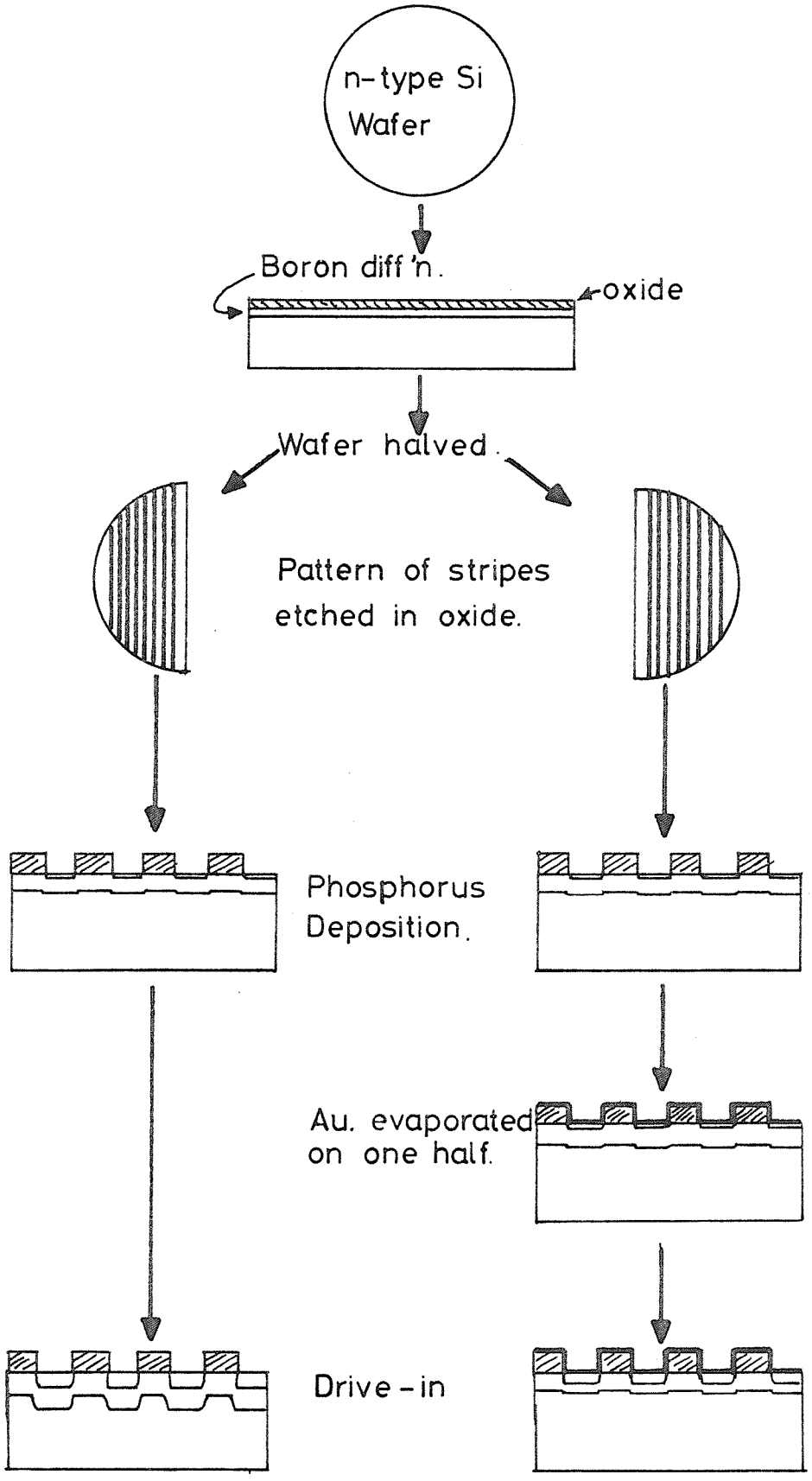


FIG. 8.28: Processing sequence for push-out experiments

8.7.7 Rutherford backscattering studies in heavily boron doped silicon

In order to try and resolve the question of whether the gold solubility enhancement observed by Brown et al. (1975) in heavily doped p-type silicon was due to enhancement of the substitutional species or of the interstitial species, an RBS study of one of the wafers used in their experiments was carried out. (This sample was kindly supplied by Dr. M. Brown). The sample in question was uniformly doped with 9×10^{19} boron atoms/cc and was gold diffused to saturation at 1573°K. The radiotracer gold saturation concentration measured by Dr. Brown was 1.7×10^{17} atoms/cc. This is close to the detectability limit for gold in silicon by RBS. A long exposure to the He^+ beam at a random sample orientation (3 hours) yielded very few counts in the gold region of the spectrum. Analysis of the data suggested a gold concentration of 2.7×10^{17} atoms/cc + 25% or -100%. Since there was such large uncertainty the channeling experiment was abandoned.

8.7.8 Influence of gold on base push out

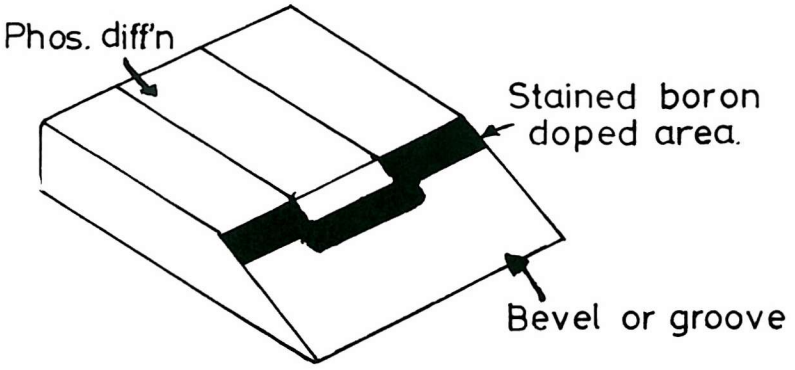
It was postulated in chapter 6, that if an 'extra' solubility enhancement of gold in a phosphorus doped layer is caused by a vacancy excess, the saturation of the phosphorus layer with gold should "mop-up" the excess vacancy population through the reaction : $Au_i + V \rightarrow Au_s$. Indeed, it was further proposed that this reaction may in fact be Au_i + (E-centre) $\rightarrow$ AuP. This gettering reaction was linked with the phenomena of base push out and the phosphorus kinked profile since it has been suggested that they are attributable to E-centre break-up. If the push-out effect is caused by a vacancy excess from the phosphorus diffused layer, then simultaneous diffusion of phosphorus and gold might be expected to reduce the magnitude of the push-out effect normally caused by the same phosphorus diffusion in the absence of gold.

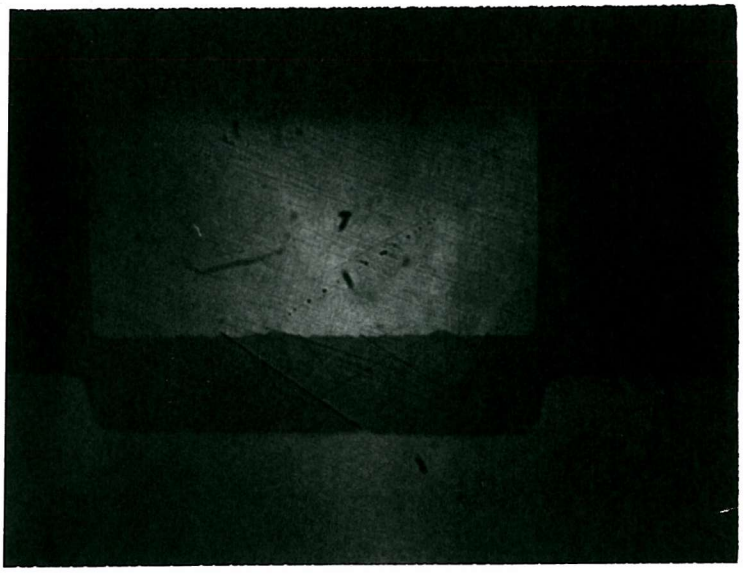
Figure 8.28 shows the processing sequence used to test this possibility. It is also listed below.

Material : n-type, (111), 2-5 Ω cm dislocation free wafers.

i Wafer clean

ii Boron deposition from boron-nitride planar source at 960°C in dry N_2 for 30 mins.





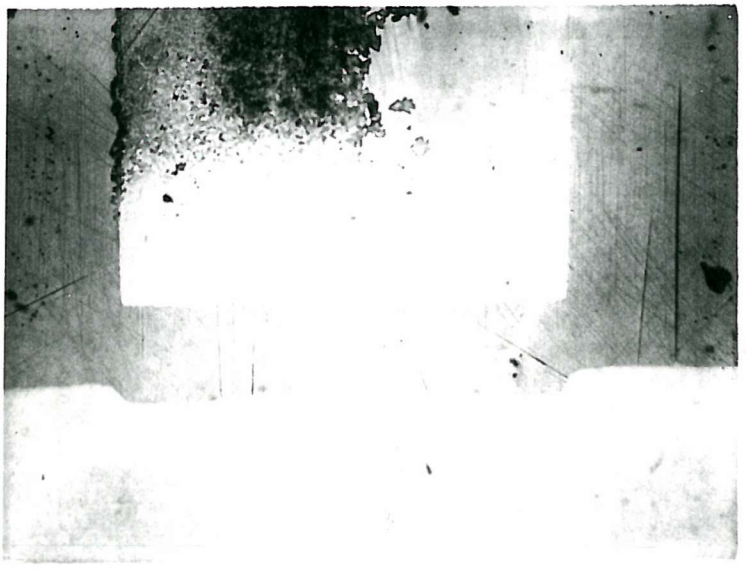


FIG 8.29 Typical result of push-out experiment.

iii Boron drive in at 1200°C 8 mins dry O_2
10 mins wet O_2.

iv Wafer halved.

v Pattern of 300μm wide stripes etched in oxide on both halves to provide windows for phosphorus diffusion.

vi Phosphorus deposition 1050°C, 15 minutes (Gas flows: see step vi in section 8.7.1)

vii Phosphorus glass removed from both halves by dip in buffered HF.

viii Very thin Au layer evaporated onto front face of one half.

ix Au deposited half given 30 mins Au + P drive in at 1050°C in dry N_2.

x Non-Au deposited half given similar drive-in.

Measurements : Both of the samples were grooved and the p-type regions were stained. The results are shown in figure 8.29 where it can be seen that there are very distinct differences between the samples. (There is no scale on the photographs because, without knowing the bevel angle it is meaningless).

There is evidence of push-out on both samples although the magnitude is very much less on the Au+P diffused half; In addition the phosphorus boron junction depth is much smaller in this half. The junction depths were : (all $\pm$ 0.04).

Sample	P-B jcn depth	B-substrate (not under P)	B-substrate (under P)	Amount of push out
No gold	1.8 μm	2.1μm	2.51μm	0.41μm
Gold	1.3μm	2.0μm	2.24μm	0.24μm

The push-out effect does not appear, from these results, to have been entirely prevented by the addition of gold to the system, however, account must be taken of the amount of push-out caused during the phosphorus deposition stage. A sample was processed in the way described above up to step vi. The structure was then grooved and stained to obtain the amount of push-out after deposition only. The results were:

P depth = 1.1μm. B (not under P) = 2.0μm. B (pushed out) = 2.29μm. This amount of push-out, 0.29μm, was the same within experimental error as the total push out observed for the gold doped drive-in situation. This indicates that the simultaneous diffusion of gold with the phosphorus 'emitter' diffusion has totally inhibited the push-out effect.

The experiment, which has been repeated a number of times with the same result, is open to the criticism that placing a gold layer on top of the phosphorus diffusion could have caused the observed result by compound formation rather than by inhibiting any vacancy excess. To investigate whether or not this was so, the whole experiment was repeated except that, at stage vii the phosphorus glass was not removed. Instead, the oxide masking layer on the back of the wafer was removed and the gold source evaporated onto the back. The drive-in cycle at 1050°C was repeated as before and the sample was grooved and stained with the following result : (all $\pm$ 0.04μm).

P depth = 1.5μm. B (not under P) = 2.05μm. B(pushed out) = 2.36μm. The amount of push out in this case is 0.3μm - only slightly more than the case with gold on the front.

The effect is considered to be the result of an interaction between diffusing gold and diffusing phosphorus. The slightly larger push out observed in the latter experiment is not surprising since, compared to the first experiment, the supply of gold is limited by the rate at which it can diffuse across the wafer.

The other result of interest in these experiments is the retardation of the phosphorus-boron junction depth in the gold doped samples. Once again, less retardation has occurred for the sample with gold on the back. This retardation,which is almost certainly associated with the inhibition of push-out, is discussed in chapter 9.

8.7.9 The effect of post-oxidation gettering on MOS capacitor relaxation times

The two results presented here illustrate an effect of gettering on wafers which, in the absence of the getter, appear to have very low contamination levels and hence very low leakage currents. In each sample phosphorus gettering is applied to the back of half of the wafer only, the other half (which is not separated) receives the same heat treatment in the same furnace but with an oxide mask on the back as well as on the front.

Material : 2-5 Ωcm, n-type, (100), FZ, 2" diameter wafers. These wafers are dislocation and swirl free.

Processing :

i Wafer cleaned. (R.C.A. clean with electronic grade chemicals)

ii 'Gate' oxidation in HCl cleaned furnace tube. 60 mins dry oxidation at 1100°C with 1 or 3% HCl in the gas ambient.

iii In situ anneal in N_2 for 15 mins. Wafer cooled rapidly in N_2.

iv Oxide removed from half of back of wafer-

v Wafer cleaned (R.C.A. clean as in i)

vi Gettering step at 1050°C 10 minutes.
Gas flows: 30 cc/min N_2 through $POCL_3$ at 30°C
10 cc/min O_2
1 /min N_2.

vii Aluminium evaporation from electron beam system.

viii 90 minute anneal at 420°C in H_2:N_2,60:40 mixture.

The aluminium was evaporated through a metal mask to give a matrix of capacitor dots 1mm in diameter, 2 mm apart. The low temperature anneal at step viii is intended to remove fast surface states from the Si-SiO_2 interface (Yeow, Y.T. 1974) and to remove radiation damage induced by the E-beam evaporation system (Grove, 1967, p 143).

Measurements: Alternate capacitors in two rows across the width of each wafer were measured in a direction orthogonal to the border between

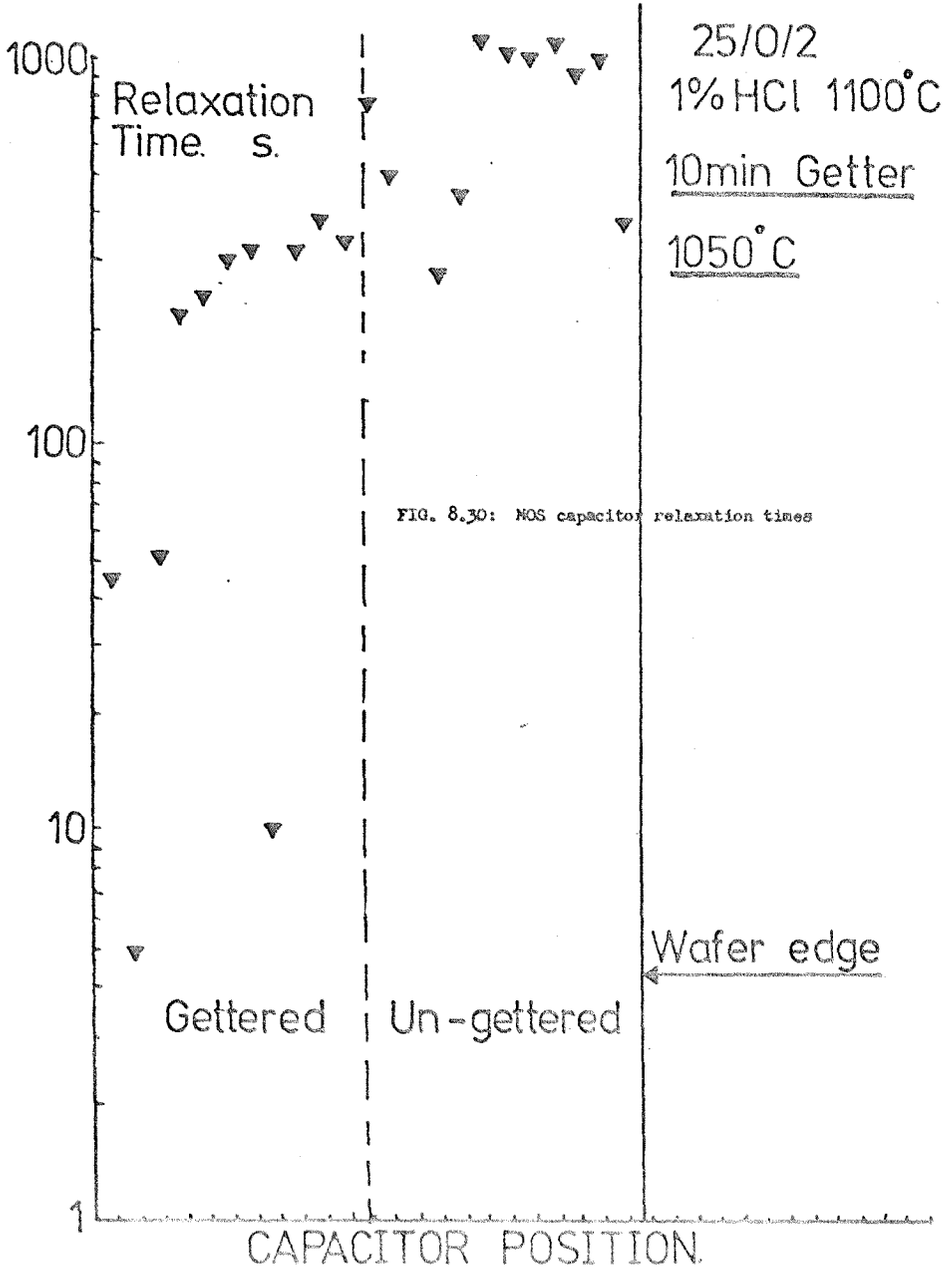


FIG. 8.30: MOS capacitor relaxation times

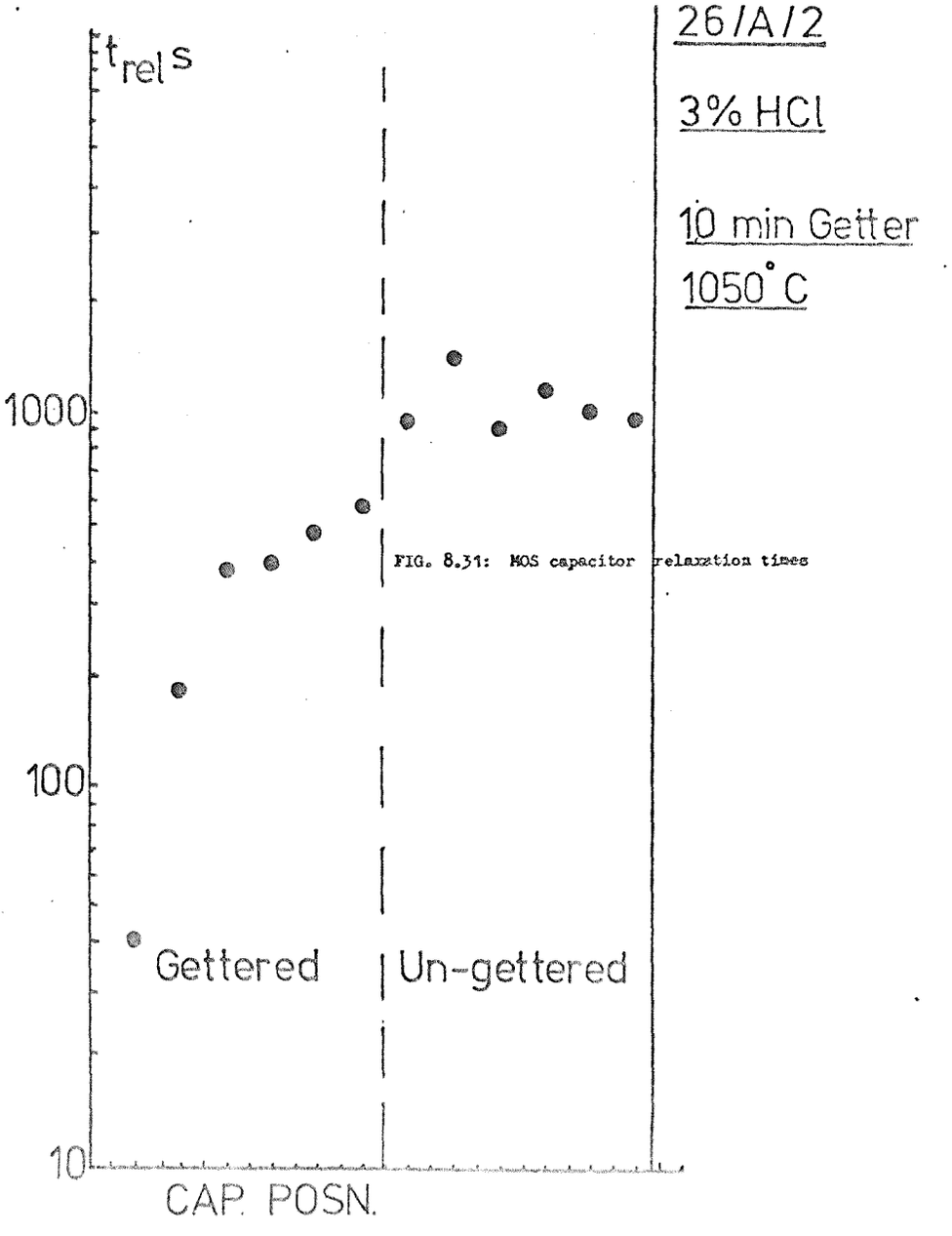


FIG. 8.31: MOS capacitor relaxation times

the gettered and ungettered halves of the wafer. In this way half of the capacitors were on the gettered sides of the wafers and half on the ungettered sides.

Results : Figures 8.30 and 8.31 show results from two wafers processed in the manner described above. The relaxation times are plotted as a function of the position of the capacitors across the 2" diameter wafers. The only difference in processing between the two wafers is the percentage of HCl gas present in the oxidising ambient - however, this does not appear to have had an effect. In both cases the gettered halves of the wafers show distinctly lower relaxation times which may be attributable to contamination entering through the backs of the wafers during gettering. Obviously no definite conclusion can be drawn from this one experiment except perhaps that it is not safe to assume that phosphorus gettering will always bring about an improvement when low levels of leakage current are sought.

9. DISCUSSION OF EXPERIMENTAL RESULTS

In the discussion of the interaction between shallow phosphorus diffused layers and gold diffusion in silicon wafers (Chapter 6) it was indicated that certain aspects of the dynamics of the problem needed clarification as they had not been previously considered in any detail in the literature. By measuring gold profiles of electrically active (substitutional) gold in the experiments described in chapter 8, it has been possible to identify (at 1000°C) that the rate limiting step of the gettering process is the diffusion rate of interstitial gold (Au_i); it has been shown that the gettering layer behaves as an infinite sink for gold until it is almost saturated and that the flux of gettering centres into the wafer is greater than the flux of gold atoms into the gettering layer. These results all suggest that the simple model of the gettering process dynamics which was developed in Chapter 6 is indeed based on valid assumptions ; viz. a gold diffusion rate limited process in which the gettering layer acts as an infinite sink for gold with the flux of gettering centres into the gettering layer greater than the flux of gold into the gettering layer.

These conclusions arise as follows:

9.1 Discussion of results from runs G51 - 55

In the experiments described in 8.7.1, there is an infinite supply of gold at the back of the wafer and essentially a fixed amount of phosphorus at the front. The gold diffuses into the wafer as an interstitial species (assuming, of course, that the dissociative mechanism is correct). The gold concentration measured in the experiments is assumed to be substitutional gold only, Au_s, - this being the electrically active species - so before its presence can be detected, an interstitial gold atom must react with a vacancy. This process has very obviously occurred in all of the wafers which have received gold diffusions only since gold is detected throughout all of them with especially high concentrations close to the wafer front surface.

In the wafers which had received a phosphorus diffusion prior to gold diffusion, the gold profile shapes are quite different and for a period of over 60 minutes no gold is detectable close to the front face of the wafer. There are two possible explanations of this 'dip' in the gold profile ; either there are no vacancies available for the reaction $Au_i + V \rightarrow Au_s$ or the phosphorus layer is withdrawing gold from the wafer bulk faster than the gold can diffuse in from the source at the back. The former possibility is rejected for the following reasons : if a vacancy depletion had occurred, the only cause would have been the shallow phosphorus diffusion gettering vacancies from the bulk. Quite to the contrary however, in almost all of the studies of shallow phosphorus diffused layers reported in the literature it has been concluded that a vacancy excess is produced by the layer - not a vacancy depletion (Jones et al. 1976 ; 1975 and 1974 ; Lee, 1974; Yoshida et al. 1974 ; Lee et al. 1972; Schwettmann et al., 1972; Hu et al. 1969). Additional evidence against such a possibility is provided by the inhibition of push-out observed when gold and phosphorus are simultaneously diffused - see 8.7.8. The conclusion must be that the phosphorus layer is gettering gold faster than the gold can diffuse to it and, therefore, the gettering process is gold interstitial diffusion rate limited.

This conclusion is further supported by application of the simple model described in section 6.5.6. The capacity of the phosphorus layer for gold may be estimated from figure 8.5. using the method outlined in 6.5.4. The result is a capacity of 3.5×10^{14} atoms/cm^2 for the layer. Equation 6.72, which represents the gettering rate of the layer while it is acting as an infinite sink, predicts a gettering rate of 10^{11} atoms/cm^2/ second, at 1000°C for a 350μm thick wafer. The values for D_i and C_i, which are 8×10^{-6} cm^2/sec and 8×10^{14} atoms/cc, were taken from Wilcox et al. 1964. If the gettering layer were to behave as an infinite sink until saturated, the time required for saturation would be 58 minutes. Bearing in mind that the values of D_i and C_i for gold are open to some doubt and that it is almost certainly incorrect to assume that the layer is an infinite sink until saturated, this time of 58 minutes is quite close to the observation that gettering in this experiment ceased sometime between 60 and 120 minutes.

Consideration of the values of gold concentration obtained from the R.B.S. measurements on the phosphorus doped layers of these samples sheds further light on the gettering rate and the efficiency of the sink. The simple prediction of equation 6.72 is that the amount of gold gettered, increases linearly with time. Experimentally, the gold concentration at the phosphorus layer surface is 7.7×10^{17} atoms/cc after 15 minutes and 1.6×10^{18} atoms/cc after twice the time, thus bearing-out the prediction. After 60 minutes however, the increase is only to 2.2×10^{18} atoms/cc, indicating that the sink is no longer as efficient ; however, since the layer contains almost 75% of its total capacity for gold by this stage the result is perhaps hardly surprising. Certainly, the phosphorus layer does appear to be an infinite sink until at least 50% filled.

A further calculation may be carried out on this data using equation 6.72, to see how closely the measured values of gold concentration in the phosphorus layer compare with the predicted values. Assuming that at any time, t, the gold profile within the phosphorus layer is the same shape as the saturated profile

$$\left(\text{i.e.} \quad \frac{C_s}{C(x)}\left[t\right] = \frac{C_s}{C(x)}\left[\text{saturated}\right]\right)$$

and using the gettering rate of 10^{11} atoms/cm^2/second already calculated, the gold concentration at the surface after 15 minutes should be 1.2×10^{18} atoms/cc. This is within a factor of 1.6 of the measured value. Once again this is considered as a good correlation when the uncertainties in D_i and C_i are considered. Indeed all of the results obtained here suggest that one or both of these parameters is an overestimate, a conclusion supported by recent work of Huntley et al.(1973ii).Fitting the parameters to the results gives a $D_i.C_i$ product of 2.5×10^9 atoms/cm/second, compared to 4×10^9 atoms/cm/second from the values of Wilcox et al.(1964).

The RBS channeling experiment in G54 is another strong indicator that this gettering process is mainly associated with gold on substitutional lattice sites. This result is an excellent agreement with others reported in the literature (Meek et al. 1975 ; Chou et al. 1975) who measured ~90% of the gettered gold to be on substitutional sites in saturated phosphorus diffused layers.

9.2 Buried "barrier" layer experiments

A predicted consequence of the very high rate of reaction between diffusing interstitial gold and gettering centres within the phosphorus diffused layer is that such a layer should act as a barrier to diffusing gold until its efficiency as a sink for gold is very much reduced. Experiments described in 8.7.4 demonstrate that, within the limitations imposed by being unable to detect gold below densities of 10^{14} atoms/cc, this effect is observed.

As pointed out in section 8.7.4, there are two alternative explanations for the results : either, as suggested above - that the diffusing species, Au_i, is converted to Au_s (or Au_sP pairs) more rapidly than it can diffuse through the layer - or that the diffusion rate of Au_i is significantly slowed down in the layer (Lambert and Reese, 1968). A rough calculation, given below, shows that during the period for which gold is not detectable in the substrate, the barrier layer will have become saturated with gold.

The thickness of the buried barrier is 12μm and its 'surface' concentration inferred from the resistivity measurements is roughly 5×10^{20} phosphorus atoms/cc. The capacity of the layer for gold may be estimated conservatively by assuming the buried layer has a triangular profile (first order approx. to erfc). This gives a gold capacity of about 3×10^{15} atoms/cm^2. Using equation 6.72 the time required to saturate the buried layer with Au_i diffusing through the 15μm epitaxial layer is $\simeq$ 40 minutes. It appears, therefore, that G23 (60 minutes Au diffusion) would be expected to be saturated whereas the samples diffused with gold for a shorter period would not. The observed gold in G23 and G24 suggest very strongly that the layer is only penetrated when saturated. Before this occurs, the rate of conversion of gold from Au_i to the relatively immobile gettered state (Au_s or Au_sP pairs) is high enough to prevent any interstitial gold from diffusing right through the layer. Once saturation and hence penetration of the phosphorus doped layer has occurred, the rate of rise of gold concentration in the bulk is apparently the same as that observed in the absence of the buried layer (compare figure 8.21 with the ungettered control samples in figures 8.17 and 8.18). This

behaviour is more consistent with the reaction proposed here than with a simple slowing down of Au_i.

9.3 Gettering of an already established gold concentration

In the model described in chapter 6, it was shown by a brief calculation, that the diffusing flux of phosphorus from an infinite source entering a typical "gettering" layer at 1000°C was greater than the interstitial diffusion rate limited flux of gold entering the layer across a wafer of typical thickness. The veracity of this is shown in the results of runs G11 and G12 (figures 8.17 and 8.18). Once again, the gettering rate must be greater than the gold interstitial diffusion rate to produce a dip in the gold profile. Indeed in both of these experiments the take-up of interstitial gold by the phosphorus doped layer reduces the concentration, Au_i, in the vicinity of the gettering layer to the extent that the substitutional gold concentration - already established prior to gettering - becomes depressed due to the reversal of the reaction $Au_i + V \rightarrow Au_s$. Indeed the situation truly becomes one of "dissociative diffusion" since the substitutional gold atoms dissociate into $V + Au_i$. In run G12, the dip in the substitutional gold concentration extends much further into the bulk than in G11 showing that, rather than becoming saturated in any way, the phosphorus layer is gettering even more after 60 minutes.

When the gettering layer was limited in runs G51-55, the volume of the wafer which was depleted of gold became progressively less as the phosphorus layer was saturated with gold coming from an infinite source. On the other hand, in G11 and G12 where both phosphorus and gold are supplied from infinite sources, the depleted volume becomes larger showing that the input of gettering centres can more than cope with the input of gold. It can be seen from equation 6.72 , that if the wafer is thinner, the gold flux increases, for this reason no generalisation from the result discussed here can be made. In addition, the calculation in Chapter 6 was only carried out for the temperature at which these experiments were done. Examination of the relative temperature variations of the diffusion coefficients of phosphorus (Sze, 1969, p31) and interstitial gold (Wilcox et al. 1964) indicate that at lower temperatures the gold flux

could become greater than that of phosphorus for the same wafer thickness etc. The unreliability of the gold diffusion data makes an evaluation of the effect at all temperatures something of an academic exercise, however the experimental technique developed here should enable further work - which will hopefully clarify matters - to be carried out with ease.

This effect is of particular importance in device processing where, when deliberate gold doping is desired, the gold diffusion from a source on the back of a wafer is carried out simultaneously with the phosphorus 'emitter' diffusion. Such a process (which economises on furnace steps) is common and was used for the structure investigated by Brotherton et al. (1972). In high speed bipolar switching devices the lifetime 'killer' is required in the base region. The model in Chapter 6 and the result discussed here show how difficult it may be to achieve this situation reproducibly - the wafer thickness could well be the parameter which determines how much gold there finally is! The solution to the problem would appear to lie in using thin wafers or - more suitably - in ensuring an adequate supply of gold. Perhaps a good way of doing this would be to place the gold source on the same side of the wafer as the phosphorus diffusion. This would have the additional advantage of inhibiting base push-out (see section 9.4 below) thus enabling better control of the base width.

9.4 The "push-out" effect experiments

The result obtained in section 8.7.8 is of considerable interest. As with most diffusion interactions no concrete conclusions can be made without knowing the precise diffusion mechanism of at least one of the components. However in this case the evidence for gold diffusing by the interstitial - substitutional mechanism is very strong (Huntley et al. 1973ii;Wilcox et al.1964) and it is known from this work, as well as references previously cited, that almost all of the gold atoms in a saturated phosphorus diffused layer reside on lattice sites. Assuming that this mechanism is correct and that, from the discussion above, the reaction rate for $Au_i + V \rightarrow Au_s$ is extremely high in the phosphorus doped region one may make a number of inferences from the results.

The push out effect is almost certainly caused by a vacancy excess generated, in some way, by the phosphorus diffusion. Some or all of these vacancies are annihilated by reacting with interstitially diffusing gold ($Au_i + V \rightarrow Au_s$) and this reaction takes place within the phosphorus doped region. If there is an adequate supply of gold the push out effect does not occur at all and if there is only a limited supply of gold, it is strongly retarded. Although the push-out effect has been modelled as a vacancy excess and many of the authors cited here and in 8.7.8. have suggested the vacancy as its cause, no evidence as direct as the result presented here has been published.

A retardation of the phosphorus-boron junction depth was also observed in these experiments and in the case where the interstitial gold supply was limited, the retardation was not as great. Lee (1974) has shown how closely the push-out effect and phosphorus kinked profile are linked (see Chapter 6) so it is reasonable to conclude that the retardation of the phosphorus diffusion is, in fact, retardation of the kinked part of the profile to a degree linked with the observed inhibition of push-out. Further experiments would be required to confirm this.

It is considered that this result provides very strong evidence for the gettering model proposed in Chapter 6. The cause of push-out has been explained quite well in terms of the E-centre break-up model by Jones (1974) and more recently by Fair (1977) who also considers the kinked phosphorus profile. The inhibition of push out and, apparently, of the phosphorus diffusion profile tail by the interaction of diffusing gold with the phosphorus layer, links the mechanisms closely, thus supporting the model, proposed here, that the gettering process involves interaction of gold with E-centres. The fact that absolutely no push-out is observed when there is a flux of gold into the phosphorus layer which can match the flux of gettering centres (i.e. when the gold is on the wafer front surface) provides good justification for assuming, in equation 6.58 , that the gold solubility will be enhanced by an amount closely related to the vacancy excess (χ). If this were not the case, then some push-out would still be expected.

The best way to examine this theory would be to vary the flux of gold entering the phosphorus doped layer by using wafers of differing thicknesses with the gold source on the opposite face to the phosphorus. At some wafer thickness between the one used in section 8.7.7 and zero thickness, the point at which the flux of gold entering the phosphorus layer is adequate to prevent push-out will be reached. The gold flux could be measured by Rutherford backscattering or with the aid of radiotracers and the phosphorus flux could also be measured. Comparison of these should indicate what proportion of the phosphorus flux is giving rise to excess vacancies.

The only alternative to the push-out effect being caused by vacancies seems to involve assuming that gold does not diffuse by the interstitial-substitutional mechanism. The result that nearly all of the gold within the phosphorus layer is on substitutional sites, as well as the evidence described in chapter 3, indicates that this is highly unlikely.

9.5 Gold diffusion profiles in thin silicon wafers

The very marked differences between the gold profiles obtained in defect free and dislocated thin (250-300μm) silicon wafers have not been observed before although, as discussed in chapter 3, such differences were predicted by Huntley et al. (1970) on the basis of a gold diffusion model in which the rate limiting step for the reaction $Au_i + V \rightarrow Au_s$ was the vacancy supply. It was assumed that in dislocated silicon, vacancies would be supplied by climbing dislocations and that in dislocation free silicon, vacancies would be supplied by diffusion from the sample surfaces.

Measurements by Huntley et al. (1973i) and by Brotherton et al. (1972) only yielded small differences between profiles in silicon wafers which were nominally 'dislocation-free' and in silicon wafers which contained dislocations. (The former measurements were made by radio-tracer techniques yielding the total gold concentration ; the latter were made by the spreading resistance technique yielding the electrically active gold concentration). These results led Brotherton et al. to conclude that the concentration of substitutional gold was limited by the rate of the reaction between interstitial gold and vacancies rather than by the

vacancy supply, with the qualification that a small but noticeable effect due to the presence of dislocations had occurred.

Huntley et al. did not favour the reaction rate limited model, and proposed that their result could possibly be explained by the presence of vacancy sources other than climbing dislocations in the wafers. These sources, now generally known as "swirl defects"because of their distribution in a swirl pattern in silicon wafers, are incorporated into the silicon during crystal growth. Since the time of Huntley et al's. work, swirl defects have received increasing attention (Vieweg-Gutberlet, 1973) due to their identification as leakage current sources in silicon vidicon devices (Shiraki, 1971), p-n junctions (Lesk, 1973) and MOS devices (Unter et al. 1977). De Kock (1970) suggested that the swirls were vacancy clusters ; further study (De Kock et al. 1975 ; Petroff et al. 1975) has indicated that the defects are probably microdislocation loops caused by the condensation of excess point defects (vacancies and silicon self-interstitials) whilst cooling from the melt during crystal growth. Techniques for growing silicon free of such defects have been reported (Herrmann et al. 1975) and it is such recently available material that was used in G51-55.

The much lower central gold concentrations in these wafers compared with those obtained on non-defect-free material must constitute strong evidence for Huntley et al's postulation that the swirl defects in their samples gave rise to excesses of vacancies.

It is worth noting that Brotherton et al. did consider the possibility of vacancy supersaturation as a result of swirl defects. Using the report of Sanders and Dobson (1969) that an oxidising silicon surface acts as a sink for vacancies, they oxidised dislocation free wafers for an extended period prior to gold diffusion. This step produced no change in the observed rate of rise of gold concentration with time. More recent studies of the mechanism of oxidation has produced much evidence to suggest that, rather than acting as a sink for vacancies, an oxiding surface acts as a source of silicon self interstitials (Hu, 1975 and 1974). Hu (1977) shows that mutual annihilation between self interstitials and vacancies is unlikely since the activation energy for such a process is higher than the activation energy of formation of their respective clusters.

Clearly there is an area for further investigation here since, if the gold diffusion process is limited by the vacancy supply or by a combination of vacancy supply and reaction rate, the profiles are not only bound up with the properties of gold diffusion but also with those of vacancy diffusion in silicon.

Huntley et al. (1970) predicted that, for defect free material,the only sources of vacancies are the sample surfaces - which are effectively an infinite source. The vacancy depth distribution resulting from such a situation would be a complementary error function profile becoming linear when the vacancy diffusion length, $2\sqrt{D_v t}$, is greater than the wafer thickness and with a constant surface concentration equal to the equilibrium vacancy concentration. If the gold-vacancy reaction rate is high, then the substitutional gold profile will follow the vacancy profile with a constant surface concentration equal to the gold solubility limit. Examination of the profiles shown in figure 8.22 suggests that, although they do follow a shape close to the expected complementary error function, the gold surface concentrations rise with time. This rise is probably largely brought about by the limitations of the measurement technique. As already explained (section 8.1) the spreading resistance reading is influenced by the resistivity of silicon within a volume beneath the probes which is defined by the probe-silicon contact area. For a given value of resistivity at the contact interface the actual measured value will be lower for a steeper fall in resistivity with depth beneath the probe (see appendix A for a more detailed discussion of this effect). In the profiles shown in figure 8.22, the surface concentration of gold is artificially lower for the shorter diffusion times since the profiles are steeper. Correction of the profiles for this effect would, therefore, both raise the surface concentrations and bring them much closer together.

Overall, these profiles do tend to look like the ones predicted by Huntley et al. (1970) for gold diffusion limited by vacancy diffusion from the wafer surface and do, therefore, support this model rather than the gold-vacancy reaction rate limited model. It must be stressed that these results obtained at 1000°C do not necessarily imply that the same result would be obtained at any diffusion temperature since the reaction-rate and vacancy diffusion rate may exhibit widely differing temperature dependences.

10. SUGGESTIONS FOR FURTHER WORK

The spreading resistance technique has been shown to provide excellent resolution of gold profiles throughout the bulk of silicon wafers for gold concentrations above the substrate doping level. The use of this method could be extended to the study of electrically active profiles of other deep level dopants which affect the resistivity. Gold is not the only deep level dopant which has been suggested as a dopant for extrinsic silicon infra-red detectors; dopants such as indium, sulphur and platnum have all been proposed but little is known about their diffusion behaviour. Spreading resistance profiling should provide a quick and simple means of assessing structures diffused with these dopants if the lack of resolution below the substrate doping level is not a limitation.

The study of gold profiles in defect free silicon wafers and the investigation of the gold-phosphorus interaction should be extended to other temperatures since the dynamics of the processes may be different. R. B. S. measurements coupled with spreading resistance depth profiling have been shown to provide an excellent picture of the gold distribution in the samples used in the experiments described here ; it would be interesting, however, to use junction transient experiments (see chapters 4, 5 and appendix F) to obtain information about the gold distribution in the junction regions of the n^+p structures studied. A large area diode with a metal guard ring - no. 211. TU.DL. - has been designed for this purpose although time did not permit any experiments to be carried out.

A lack of reliable quantitative data concerning the relative concentrations of interstitial and substitutional gold has been highlighted and further work in this area would be of value.

The effect of gold on the anomalous diffusion effects associated with phosphorus 'emitter-type' diffusions could be studied further. Possible experiments using silicon wafers of different thicknesses were described in chapter 9. Phosphorus profiles should be measured - by incremental sheet resistance or a similar technique - to gain a better idea of whether the gold does affect formation of the 'kink' in the phosphorus profile.

APPENDIX A

THE SPREADING RESISTANCE MEASUREMENT TECHNIQUE

In order to carry out some of the resistivity profiling described in this thesis a substantial period was spent investigating factors which affect the resolution, reproducibility and accuracy of spreading resistance measurements. Most of the results should be applicable to optimising the performance of any spreading resistance apparatus although a number of improvements to the particular set-up used in this work are described.

The general properties, usefulness and limitations of the technique are first reviewed briefly. A detailed consideration of all aspects of spreading resistance measurements is not made since some are not directly relevant to the work described here. During the currency of the investigations described in this appendix the proceedings of the first 'Symposium on Spreading Resistance', organised by the U.S. National Bureau of Standards, became available. Many of the problems encountered with the apparatus used here had been encountered by other workers using different equipment. Comparison with papers from the symposium indicates that the reproducibility of results obtained here is comparable with or better than measured by many other workers ; the reason appears to lie in the sample preparation technique used here. Interested readers are directed to the published proceedings of the symposium for a detailed assessment of the state of the art in spreading resistance measurements (N.B.S. special publication 400-10; December 1974).

A1. Spreading Resistance - Introduction

A number of techniques for the measurement of silicon resistivity and resistivity-depth profiles are commonly used, these include four point probe measurements, junction and MOS capacitance - voltage techniques, use of mercury probes and optical methods (Runyan, 1975). Probably the best known and most widely used of these is the four point probe sheet resistance measurement technique (Runyan, p.75). The use of potential probes in this way is direct and easy, however it has a very limited spatial resolution. That is to say, the sheet resistance measured in a particular position depends on the average resistivity of a large volume around and beneath the probes (large

compared to the contact area). A typical four point probe makes four small contacts about 0.5 mm apart ; the area around the probes over which the sheet resistance is averaged is approximately 100 mm^2 and in depth can be up to 0.5 mm depending on the thickness of the sample, this yields a large sampling volume of 50 mm^3.

For the measurement of dopant-depth profiles, use of the four point probe in conjunction with successive layer removal - known as the differential sheet resistance technique - is well known (Tannenbaum, 1961 ; Shaw, 1973, p.182 ; Appendix B of this thesis). This method requires that the resistivity of the layer is uniform for a large distance surrounding the probes in a plane parallel to the sample surface and that the removal of a thin layer (generally less than 0.1μm) from the sample surface has a significant effect on the average resistivity of the whole sample. With a shallow diffused layer where there is an electrical boundary beneath and close to the probes (i.e. the junction), this latter requirement is normally met and, provided the sample satisfies the first constraint of lateral uniformity, the technique may be employed. If the barrier does not exist, any change in the number of carriers caused by removal of a thin layer is probably insignificant when compared to the total number of carriers in the measurement sampling volume. Use of the differential technique to measure isotype epitaxial profiles and to measure resistivity profiles which vary slowly as a function of position across, or depth through, a sample is impossible. Indeed this limitation applies to almost all of the other methods of resistivity measurement which have been mentioned unless very complicated procedures are employed.

The spreading resistance technique was proposed by Lark - Horowitz et al (1959) and described for resistivity measurements and depth profiling by Mazur and Dickey (1963 and 1966), specifically to overcome the resolution problems which were described above. It is based on the well known phenomenon of constriction resistance which occurs at small area contacts between conductors (Llewellyn-Jones, 1957 ; Holm, R.A. 1958). Mazur and Dickey's method involves the placing of a very small probe as a point contact onto the silicon surface. If a current is passed through the probe into the silicon almost all of the potential

drop from probe to silicon occurs within a very small volume around the probe point ; the V/I quotient is the 'spreading resistance' which is related to the resistivity of the silicon if the probe is of negligible resistivity. The spatial resolution of the measurement is determined by the small volume around the contact of which the measured resistance is a weighted average. Various probe configurations which are discussed in the literature may be used to implement the technique (Mazur and Dickey, 1966 ; Schumann and Gardner, 1969). In this work the three-probe configuration is used ; see section A.3.

Spreading resistance measurements undoubtedly provide the easiest-to-implement technique for measuring the two types of resistivity profiles mentioned above (laterally across the silicon surface and slowly varying depth profiles) and offer the highest resolution of all techniques other than recently developed scanning electron microscope measurements (Kamm et al 1977). The main disadvantages lie in difficulties encountered with obtaining reproducible measurements (a problem largely solved here) and in the fact that the measurement is destructive because the probes leave a small damaged area on the silicon surface.

Resistivity versus depth profiles are obtained by making spreading resistance measurements on a specially prepared specimen. The sample to be profiled is bevelled at a shallow angle (usually less than 4^{o} and typically less than 1^{o}). The layers to be evaluated are thus exposed on a greatly magnified horizontal scale and the measurements may easily be made by probing down the bevel. The scheme is illustrated in figure A1. In the equipment used in this work, the bevelled sample is mounted on a X-Y table beneath the probe system. The table is moved by micrometers calibrated in 2μm divisions.

A2. The Spreading Resistance of a Single Probe

The first model of the contact between a probe with an idealised hemispherical tip and the flat silicon surface was based on the assumption that the probe retained its shape and pierced the silicon. Subsequently a consideration of the deformation of both materials suggested that a more realistic geometry for the contact interface was that of a circular

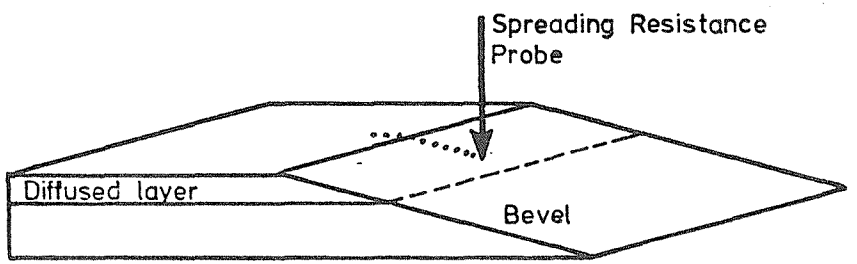


FIG. A1: Resistivity depth profiling by spreading resistance measurements on a bevel.

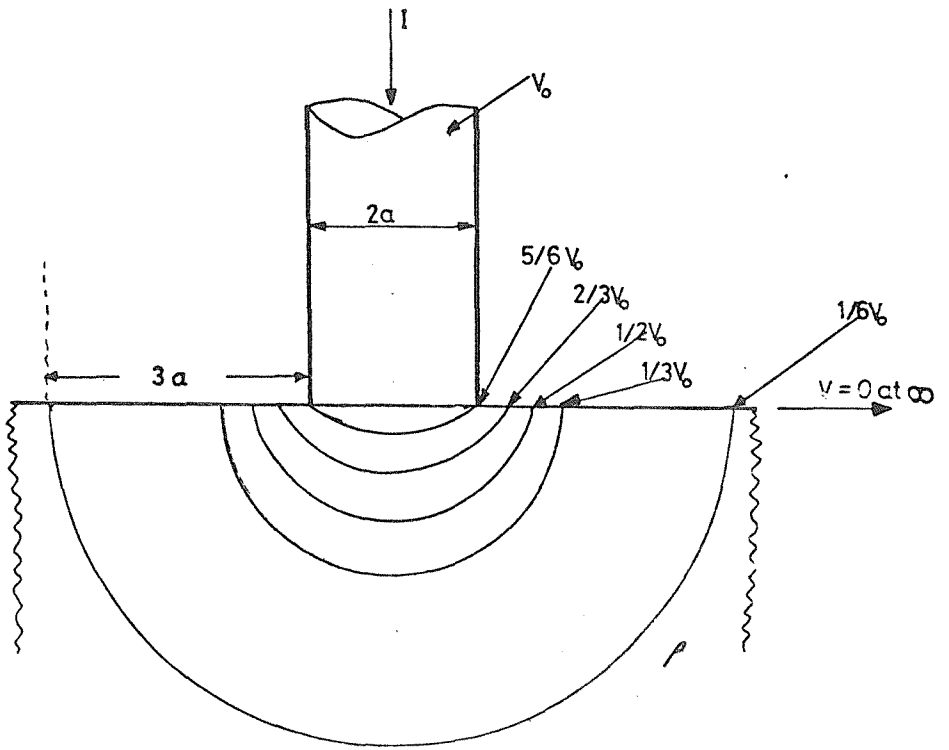


FIG A2 Potential distribution of a flat circular probe on a semi-infinite medium

disc of radius 'a' (see ref. 16 of Mazur and Dickey, 1966). This configuration leads to the following relationship between the current I, flowing through the probe and the potential drop, V, across the contact between probe and silicon:

$$V = \frac{\rho I}{4a} \qquad \text{A1}$$

where ρ is the resistivity of the uniformly doped silicon wafer and the resistivity of the probe is negligible. The measured spreading resistance, R_s, is apparently related to the resistivity of the silicon by the very simple formula :

$$R_s = \frac{\rho}{4a} \qquad \text{A2}$$

In reality the contact area is not a flat circular disc at all ; the real nature of the probe- silicon contact will be considered shortly but equation A2 may be used to give a good indication of the spatial resolution of the technique.

Figure A2 shows the potential distribution resulting from the situation described above ; it was obtained by solving Laplace's equation under appropriate boundary conditions (Gardner et al, 1967 ; Shaw, 1973, p 188). As may be seen, practically all of the potential drop occurs within a distance of a few probe radii from the probe centre. The spatial resolution of the measurement is defined by this potential distribution. The spreading resistance reading depends mainly on the resistivity of the material contained within the hemispherical region defined by a radius of 3 - 4a and is almost independent of what lies at greater depths.

Typical values for 'a' reported in the literature are in the range 5 - 10 μm yielding sampling volumes of less than 1.4×10^{-4} mm^3.

Mazur and Dickey showed that the ideal relationship between V and I predicted by equation A1 was not obeyed. However the usefulness of the technique was not lost since they pointed out that R_s could be related to ρ empirically. A number of uniformly doped silicon wafers of known resistivity were measured by the spreading resistance technique and

a 'calibration curve' was drawn up to relate any spreading resistance measurement to a resistivity (see A2.4).

In practice, the non-ideality of the probe tip and its interrelation with the surface quality of the silicon must be considered before accurate high resolution measurements can be made. Several factors concerned with the electrical characteristics are also important and, along with the calibration curve, are considered first.

A2.1 Non-ohmic Nature of the Contact

Metal-semiconductor contacts are generally non-ohmic due to the work function difference between metal and semiconductor. In the case of the spreading resistance probe on silicon there may be a reverse biased barrier in series with the spreading resistance or a forward biased barrier across which excess carriers may be injected thus reducing the resistivity locally. This problem can be circumvented by the use of the calibration curve provided it is ensured that the applied bias is not great enough to disturb the equilibrium situation. If the applied voltage V is appreciably less than kT/q, this condition is met. At room temperature, kT/q is approximately 25 mV. The applied voltage on the apparatus used in this work is maintained below 10 mV, thus ensuring a near linear I-V characteristic. During the measurements, this may be tested by reversing the polarity of the applied voltage ; if the measured spreading resistance does not change then it may be concluded that the applied voltage is too small to disturb the equilibrium zero-bias barrier (whether it is foreward or reverse biased). In practice, this condition has always been met by the equipment used in this work for the range of resistivities measured.

A2.2 Electric Field at the Contact

Due to the small size of the contact, the electric field in the spreading resistance region may be high enough to alter the carrier mobility and thus modify the local resistivity. Mobility in silicon is independent of electric field below field strengths of about 2×10^3 V/cm (Grove, 1967). In the experimental arrangement the potential

across probe and silicon is always less than 10 mV which ensures an electric field well below 10^3 V/cm.

A2.3 Contact Heating

The small contact area of the spreading resistance probe could cause heating of the silicon at moderate currents : the heat could then cause a thermoelectric voltage, an alteration of carrier concentration or of mobility in the silicon. Mazur and Dickey (1966) calculate that the maximum temperature rise is less than $0.1^{\circ}C$ for a probe-silicon voltage of 15 mV and a contact radius of 4μm. This obviously has a negligible effect since, as discussed in A4.2 and A4.3, drift free readings are possible.

A2.4 Barriers between Probe and Silicon

The presence of a resistive or insulating barrier layer, such as silicon dioxide, between the probes and the silicon could affect the results appreciably. Visual evidence from the damage left on the silicon surface after measurements shows that any such layer is pierced by the probe. Measurements are also shown to be reproducible for a given method of preparing the samples.

A2.4i The Calibration Curve

The calibration curve, which relates the measured spreading resistance of a uniformly doped sample to its resistivity, is unique to a given probe. This is because the contact radius, 'a', between probe and silicon depends not only on the tip radius of the probe but also, as will be shown, on its microscopic roughness, the method of preparing samples for measurement and the operating parameters of the equipment. The most uncontrollable of these is the probe tip shape. It has been found that any change in the tip nature necessitates recalibration of the equipment. The greatest problem encountered with the equipment used in this work results from accidental damage to the probe tip when the probes are not raised correctly before moving the sample.

In the light of the non-ideality of the relationship between spreading resistance and resistivity, Mazur and Dickey (1966) proposed an empirical relationship :

$$R_{s(actual)} = \frac{k\rho}{4a} \qquad A3$$

where the factor k depended on ρ . This factor included the zero-bias barrier which could be present and which has already been mentioned (A2.1). Calibration curves obtained for n-type and p-type material are different, a result which supports the view of a zero-bias barrier since the work function difference between probe and n-type or p-type silicon is not expected to be the same. A more useful way of describing the R_s: ρ relationship is :

$$R_{s(measured)} = R_B + R_{s(actual)}$$

$$= R_B + \frac{\rho}{4a} \qquad A4$$

where R_B depends on ρ and other factors which are discussed by Fonash (1974). For the purposes of this work, where the spreading resistance is only being used as a comparative technique, detailed knowledge of R_B is not necessary since the calibration curve includes it in relating $R_{s(measured)}$ to ρ . The necessity of distinguishing between the actual spreading resistance and R_B is important when making measurements of rapidly varying resistivity-depth profiles (see A2.5).

Figure A3 shows the calibration curves obtained for the equipment used here. They were constructed using a series of n-type and p-type (111) silicon wafers with resistivities in the range 1m Ωcm - 100 Ωcm. The resistivities were measured by the standard four-point-probe method in the centres of the wafers. An average of five readings close together yields a sheet resistance, R_{SH}, in ohms per square. The wafer thicknesses, t, were measured with a micrometer ($\pm$ 2μm in 300μm) and the resistivity was obtained using the relationship $= R_{SH} \cdot t$ (Grove, 1967). The samples were all large enough to avoid any error in R_{SH} due to geometrical effects (these can arise if the area of a

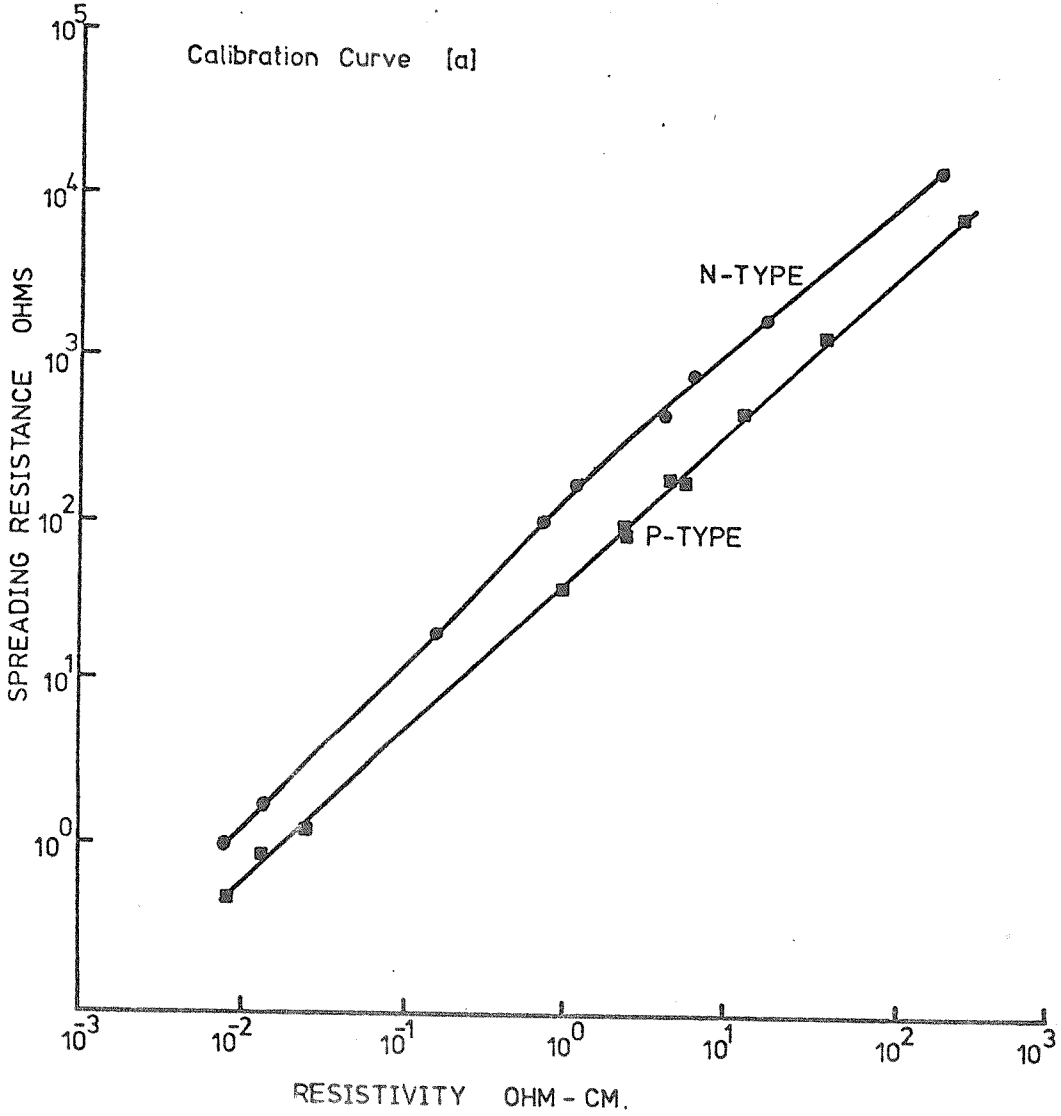


FIG. A3(a): Calibration curves for the spreading resistance apparatus using large probes.

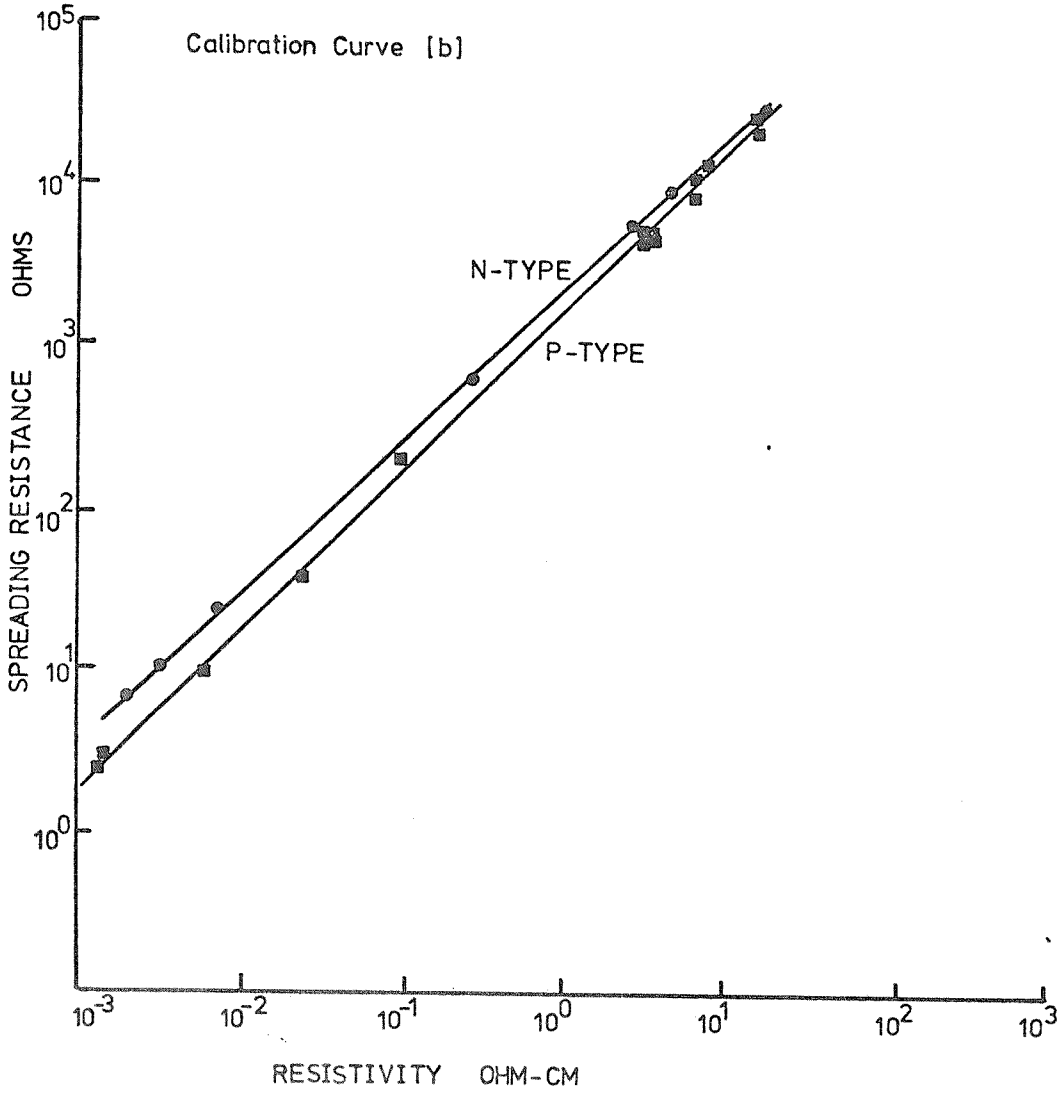


FIG. A3(b): Calibration curves for the spreading resistance apparatus using small probes.

wafer is comparable to or smaller than the area over which the four point probe measurement is averaged - Smits, 1958). Since this resistivity was a weighted average over the centre portion of the wafer, a number of spreading resistance measurements, over the same area, were taken to yield a similarly averaged value. The short term reproducibility of the spreading resistance 'average' is within 2% on a given wafer although readings taken close together suggest an absolute reproducibility better than this. The reproducibility of the four point probe measurements is $\pm$ 4%. The ultimate accuracy of the $R_s \rightarrow \rho$ conversion is limited by the resistivity measurement of the calibration samples.

Figures A3 (a) and (b) are two calibration curves obtained with two different probes. The first was a probe used in the condition in which it was received from the manufacturer - i.e. a tungsten carbide needle with a ground tip of 25 m radius. The calibration curve for n-type silicon is not linear. Any long term shift in calibration (due to probe damage etc.) necessitated a complete, time consuming recalibration of the apparatus. The second diagram is a calibration curve for a probe with a much smaller tip radius (see A5). For this probe both n-type and p-type silicon gave straight line relationships between R_s and ρ. In this case it was found that long term shifts in calibration were always parallel to the 'original' curve within the accuracy of the measurements. This fortunate result enables recalibration to be carried out with a minimum number of measurements. Changes in calibration are usually distinct rather than gradual and may generally be attributed to a specific event such as moving the silicon specimen without raising the probes. Occasional checking of the calibration with one of the test wafers allows changes to be identified and the necessary shift in the calibration curve to be made.

A2.5 Spreading Resistance Measurements on Non-Uniformly Doped Silicon

The calibration procedure described above is carried out on nominally uniformly doped samples. If the sample resistivity varies significantly within the probe "sampling volume" (as occurs when measuring shallow

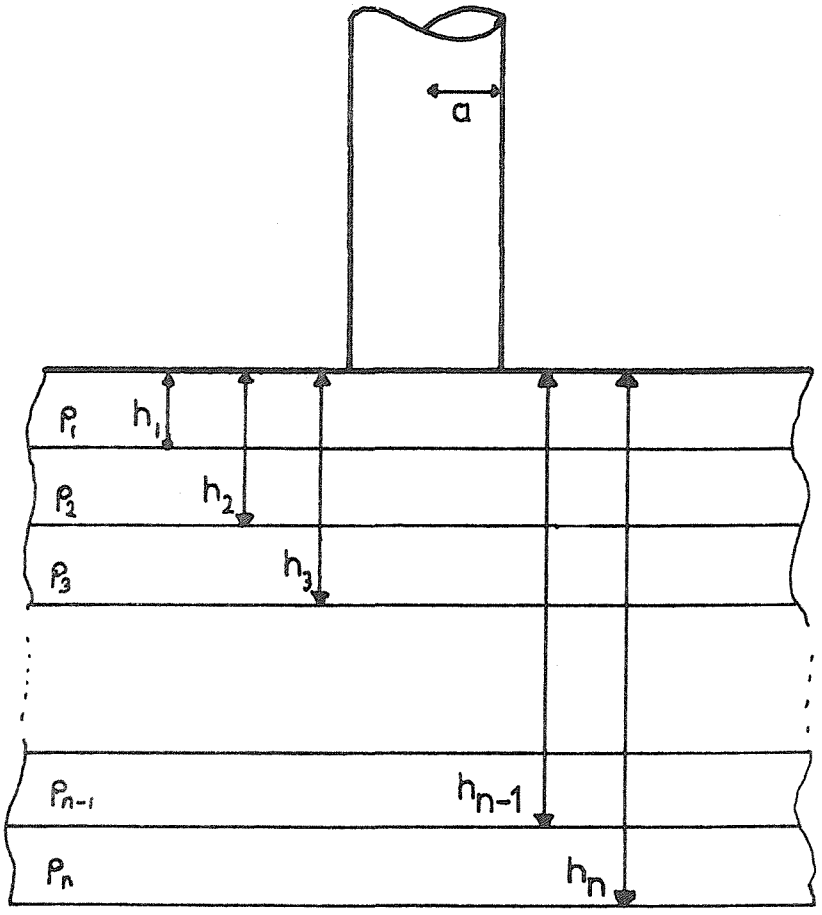


FIG. A4: Multi-layer scheme for modeling spreading resistance measurements on a non-uniform layer.

dopant profiles) a direct conversion of spreading resistance to resistivity, using the calibration curve, may not be effected; the measured spreading resistance is a weighted average over the material present in the sampling volume. Various models for correcting spreading resistance values for the variation of resistivity within the sampling volume have been proposed and a number of them appear in the proceedings of the symposium already referenced ; others are due to : Schumann and Gardner, 1969, i and ii ; Yeh and Kokhani, 1969 ; Hu, 1972 ; Gutai and Vicsek, 1975 ; Choo, Leong and Kuan, 1976 ; Kudoh, Uda, Ikushima and Kamoshida, 1976 ; Hendrickson, 1975.

The problem involved in producing any correction routine is one of the exact modelling of the nature of the probe contact with the silicon surface. Most of the treatments assume that the flat circular disc model is correct. The potential distribution is then determined by considering the non-uniform profile of interest to be a multi-layered structure as indicated in figure A4, the layer thicknesses are determined by the depth increments between successive measurements. The resultant routines for carrying out corrections to the raw spreading resistance data are very complex and usually involve large amounts of computer C.P.U. time.

For the profiles presented in this work, it has not been found necessary to use one of the correction routines. The measurements of prime interest are in regions of silicon samples where the resistivity is generally a slowly varying quantity compared to the sampling depth of the probe ; see chapter 8. For this reason an accurate correction routine was not considered necessary for the gold profiles obtained in the bulk of the silicon wafers. In all cases where the gold distribution close to the boundary formed by the wafer back surface was required, the problem of having to correct the data for the proximity of the back surface was circumvented by bevelling and measuring two pieces of each sample, one from the front and one from the back. In this way none of the measurements were taken within a 'probe sampling depth' of an electrical boundary. The nearest allowable point of approach to a boundary may easily be found by carrying out depth profiles of uniformly doped control samples ; this measurement also gives an indication of the actual sampling depth of a probe.

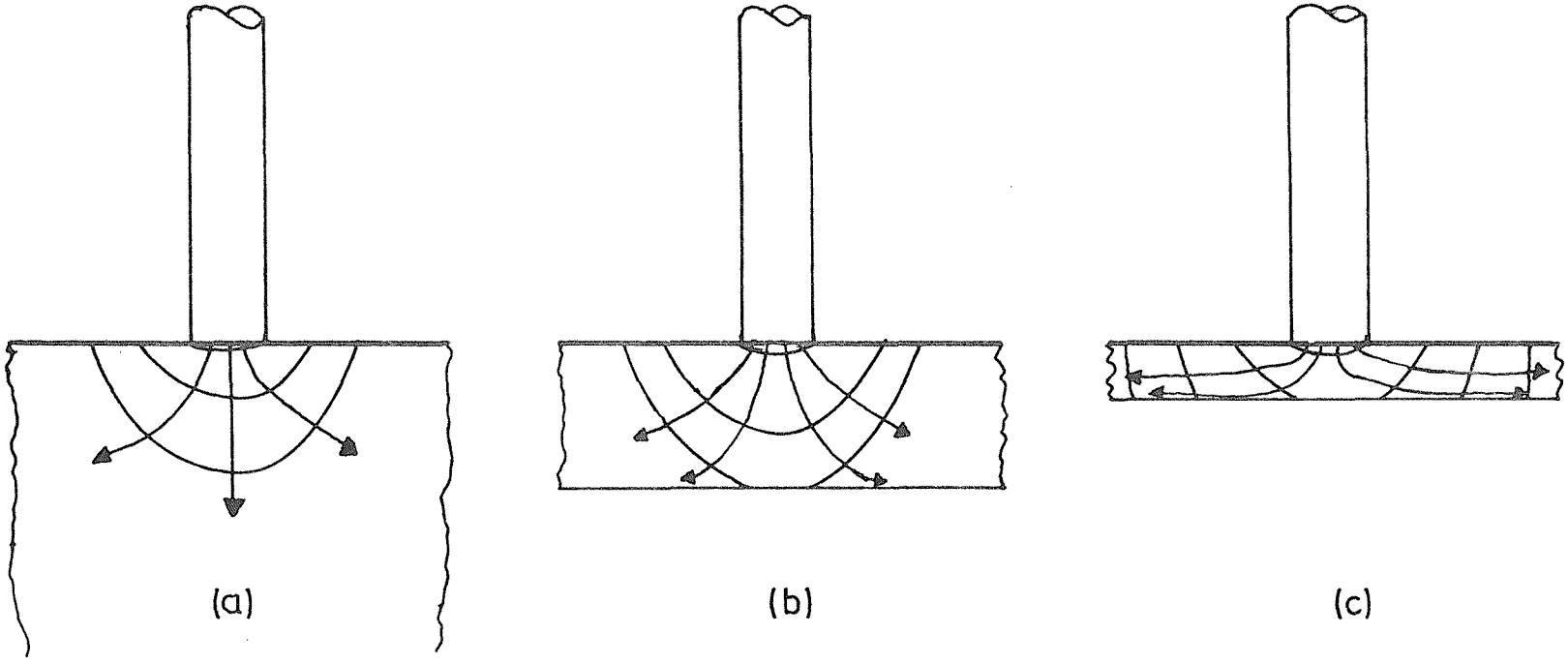


FIG. A5: Potential distribution and current flow from a spreading resistance probe:

(a) Semi infinite sample.

(b) Sample with a barrier within a sampling depth.

(c) Very thin layer.

For the measurements of very shallow phosphorus diffused layers (i.e. much shallower than the probe sampling depth) an approximation suggested by a qualitative consideration of the current flow from the probe in such a shallow layer was used . This approximation, as will be shown, allowed surprisingly accurate profiles to be obtained from raw spreading resistance data when compared to profiles of similar layers made by the differential sheet resistance method. The advantages of realising the phosphorus profiles from the spreading resistance data, without lengthy correction procedures were two-fold; first a considerable time saving is achieved since the incremental sheet resistance method is very slow and secondly a much better resolution of the phosphorus profile (or for that matter any shallow profile) was obtained through taking more measurements closer together.

A2.5.1 Simplified correction procedure for spreading resistance measurements on shallow layers

Figure A5 shows the potential distribution and current flow from the spreading resistance probe in three situations. In (a), measurement of a 'semi-infinite' sample, such as used for the calibration procedure, is illustrated. For this case no correction of the spreading resistance is required. In (b), an insulating layer is present within the probe sampling depth, the flow of current from the probe is modified somewhat and some correction of the spreading resistance data is necessary. For situation (c), where a very shallow layer (compared to the sampling depth) is being measured most of the current flow is constrained to be lateral and practically all of the potential drop is in a lateral direction. The resistance measured in this situation is clearly related more closely to the sheet resistance of the layer than to a spreading resistance. To a reasonable approximation, if a very accurate profile is not required, the resistance may be related directly to the sheet resistance, R_{SH} :

$$R_S = kR_{SH} \qquad \text{A5}$$

The constant k may be obtained empirically by measuring the surface of the sample (i.e. the whole layer) by both methods and equating the results. The value of k usually comes out to be very close to or equal to unity. Figure A6 shows two phosphorus profiles obtained

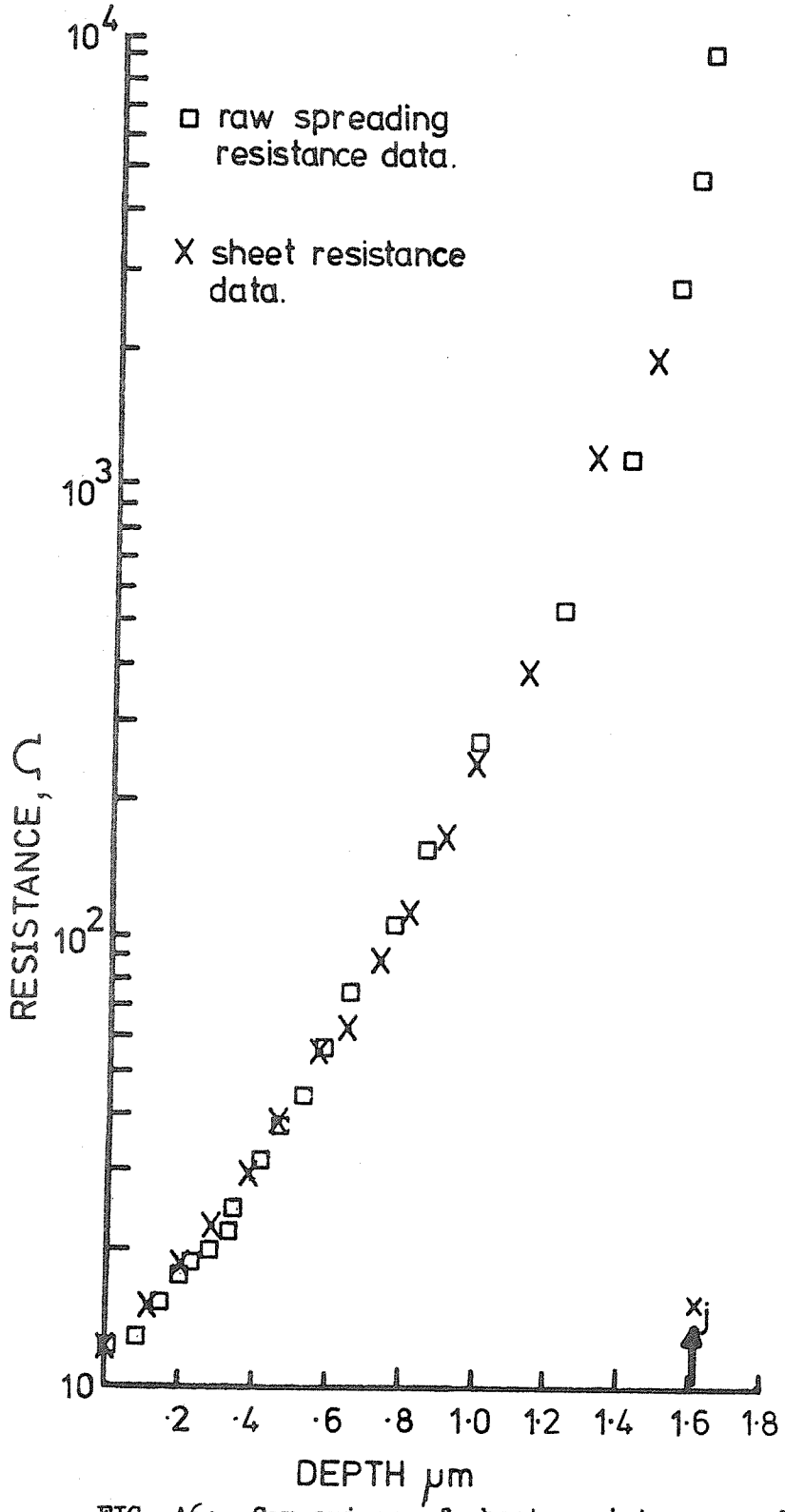


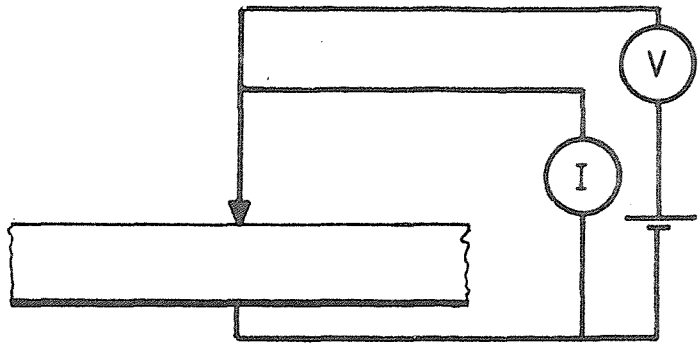
FIG. A6: Comparison of sheet resistance vs depth and uncorrected spreading resistance vs depth measurements on the same shallow layer.

by spreading resistance (on a bevel) and incremental sheet resistance (see Appendix B) for the same sample. The constant k was assumed to be unity and a common resistance axis is used. The closeness of the results obtained by the two methods is an excellent indicator of just how good the original approximation is.

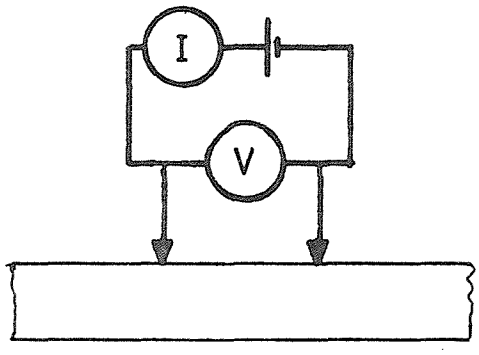
Another distinct advantage of spreading resistance profiling measurements is also apparent on this example. Sheet resistance measurements above 1k Ω per square are difficult to obtain and often inaccurate due to junction leakage (Runyan 1975, p.79). The "sheet resistances" inferred from the spreading resistance data in this example are reliable to almost 10k per square. It will be shown later in this appendix, that by improving the probe resolution it is possible to make spreading resistance measurements of over 10^6 Ω.

A3. Practical Probe Configurations

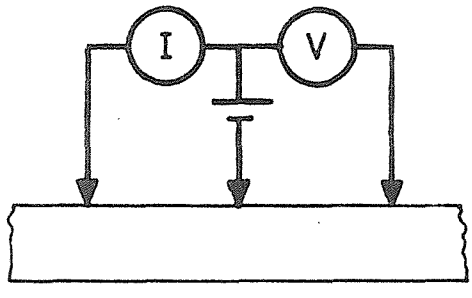
Several probe arrangements which are commonly used for making spreading resistance measurements are illustrated in figure A7. The first arrangement - a single probe and a broad area contact for the current return path - is not very versatile since the technique is not self-contained. The second arrangement, the so-called 'two-probe' system is used by most workers since it is employed in the commercially available automatic spreading resistance apparatus designed by the originator, Mazur. The V/I quotient is the sum of the spreading resistances due to each probe as they are effectively in series. This raises a major problem since not only are two reproducible contacts required but it is also impossible to separate the contributions of each to the total measured resistance. If the contact areas are identical and the silicon resistivity under each is exactly the same, then the actual spreading resistance under one of the probes is half of the total - this is the assumption which has to be made although it is extremely unlikely that either of these conditions is met. Thus the two-probe set-up has a built in uncertainty in the validity of measurements taken with it. This problem is particularly acute when rapidly varying diffusion profiles on a bevelled structure are to be measured. If the spreading resistances due to each probe are to be equal the probes must be at exactly the same depth below the sample



(a)



(b)



(c)

FIG. A7: Three practical probe configurations.

surface on the bevel. Complicated procedures for lining the probes up parallel to the bevel edge have been reported (e.g. Mayer et al, 1974).

These problems are eliminated by using the three probe arrangement which is also illustrated in figure A7. The third probe is used to measure the potential difference between the silicon wafer and one of the other probes designated as the 'critical' or measuring probe. The current flow through the critical probe is also measured, as shown, and the V/I quotient is the spreading resistance of the critical probe only. The properties of the non-critical probes are not important since one provides a means of measuring the wafer potential and the other provides a return path for the current. When measurements are made on a bevel accurate lining-up of the probes parallel to the bevel edge is not necessary. In practice it has been found that rough alignment by eye is adequate. A difference between profiles made on the same sample has only been observed when gross misalignments relative to the bevel edge have occurred.

In the experimental apparatus used here, the power supply is a simple 10mV constant voltage source. The current is measured with a Keithley 610C electrometer and the voltage is measured with a Solartron A200 digital voltmeter with an input impedance greater than $10^{10}\Omega$; this ensures that the current drawn by the voltage measuring probe is negligible.

The equipment is manually operated and has received a number of modifications as a result of experiments carried out during this work. With the small diameter critical probe already mentioned, spreading resistances of greater than $10^{6}\Omega$ may be measured. To avoid effects due to photo-generated carriers in high resistivity samples the measurements are carried out in darkness.

A4. Factors Affecting the Resolution and Reproducibility of Spreading Resistance Measurements

A4.1 The Probe Contact Area

If valid comparisons between spreading resistance measurements on

experimental and calibration samples are to be made, the contact made by the probes must be identical in all cases. Maintaining the reproducibility of the contact area, as already mentioned, has been a major problem. Other workers have experienced similar difficulties and the generally accepted way of overcoming them is to subject the probes to a "run-in" procedure. This involves carrying out a large number (>1000) of measurement cycles on a lapped silicon surface (Dickey, 1974). It is believed that such a treatment applied to different probes wears them down to a similar geometry. This type of treatment is, of course, a necessity for spreading resistance equipment employed as an in-line parameter checking facility where speed of measurement can be weighed against the ultimate resolution of the equipment. In this work, probes produced by the "run-in" were rejected for two reasons. First, the run-in period on the manually operated equipment was prohibitively time-consuming ($\simeq$16 hours of continuous operation!). Secondly, and more importantly, the "run-in" results in probes which are certainly similar to each other but at the expense of the tip-radius (figure A8) which becomes blunter. Since the high spatial resolution which results from small tipped probes was desired, "run-in" probes were not used. Careful operation of the apparatus on polished (rather than lapped) samples has been found to give a reproducible contact pattern for long periods.

The true nature of the probe-silicon contact was investigated by observing the damage left on the silicon surface after measurement. These observations were made by optical microscopy using Nomarski interference contrast. A short (5 second) SIRTL etch (Sirtl et al, 1961), made the probe damage, or footprint, more easily visible. Figure A9 shows examples of footprints produced by the probes supplied with the apparatus. The contact pattern is more a cluster of small penetrations of the silicon surface than a 'flat circular disc'. The reproducibility of the contact may be seen by comparing successive footprints of a given probe; the large differences which may occur between two different probes are also apparent.

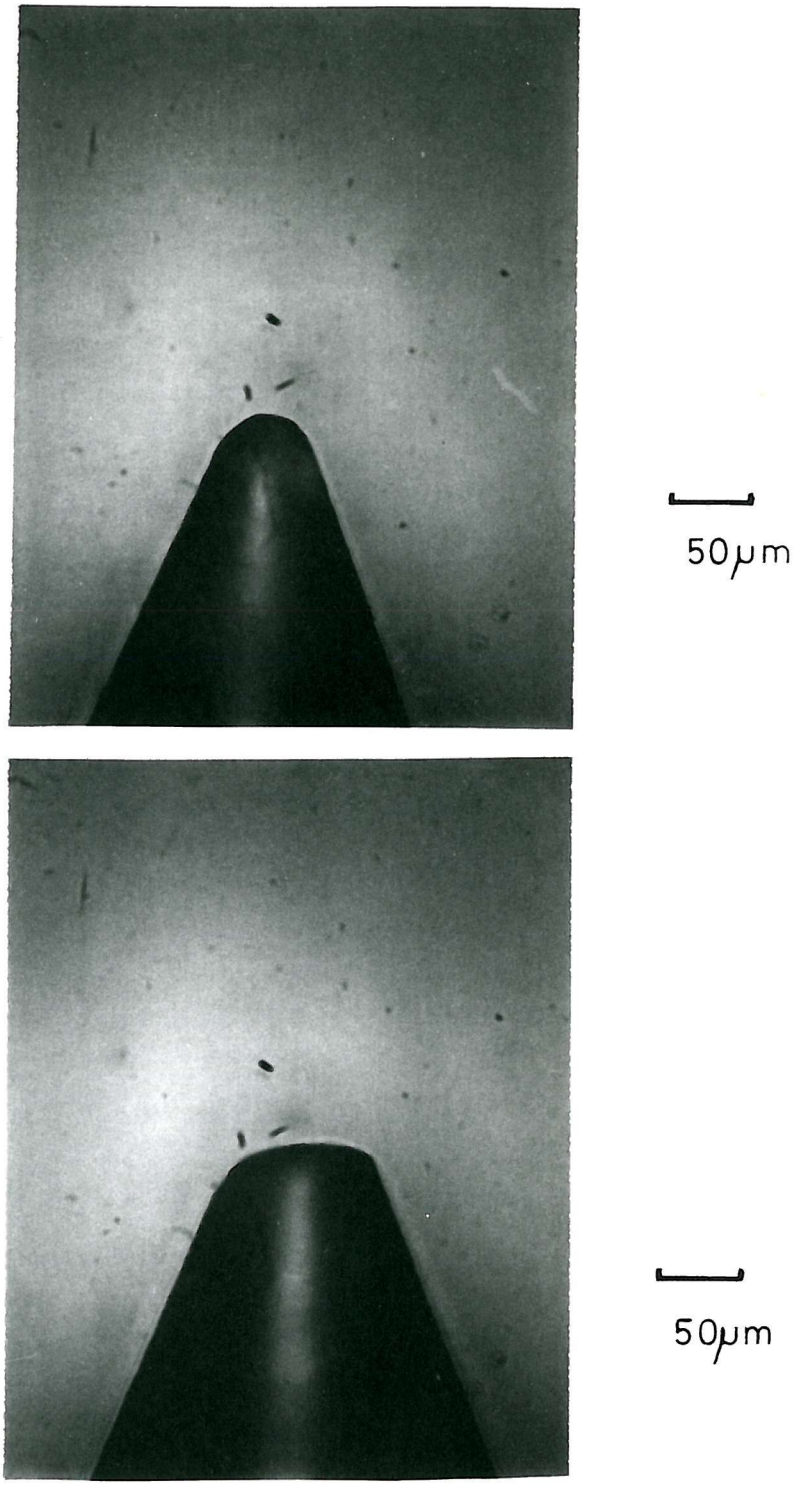
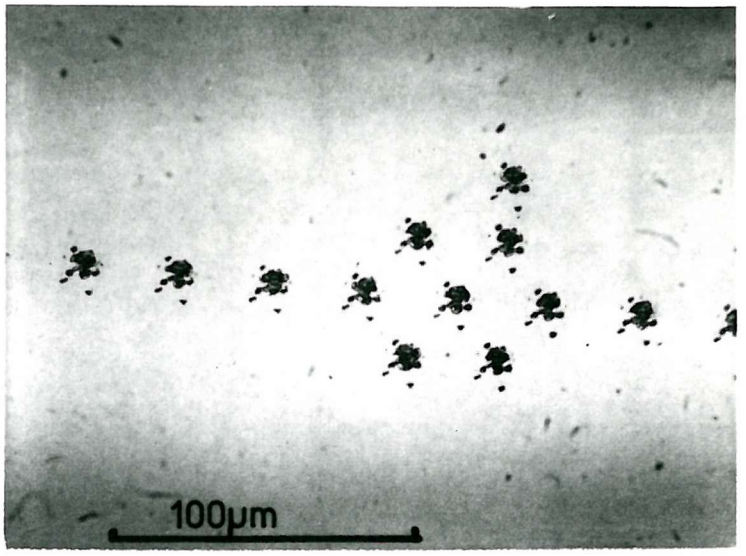
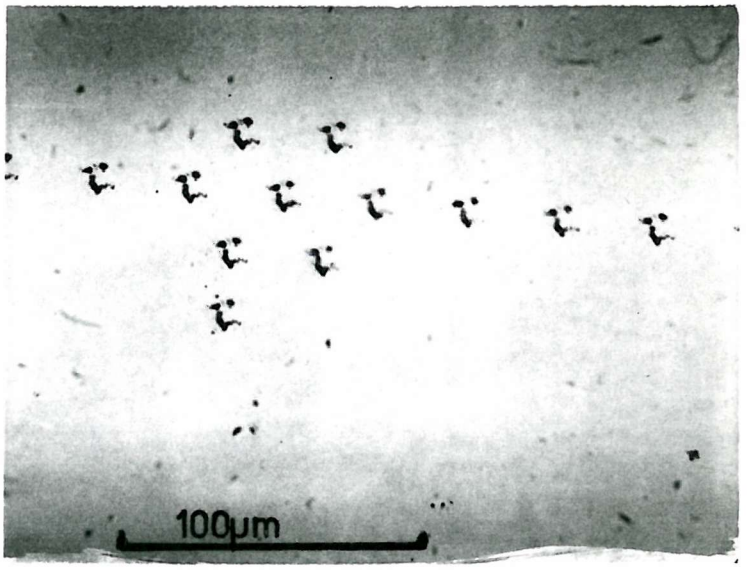


FIG. A8: Comparison of a new probe with a 'conditioned' or run-in probe.





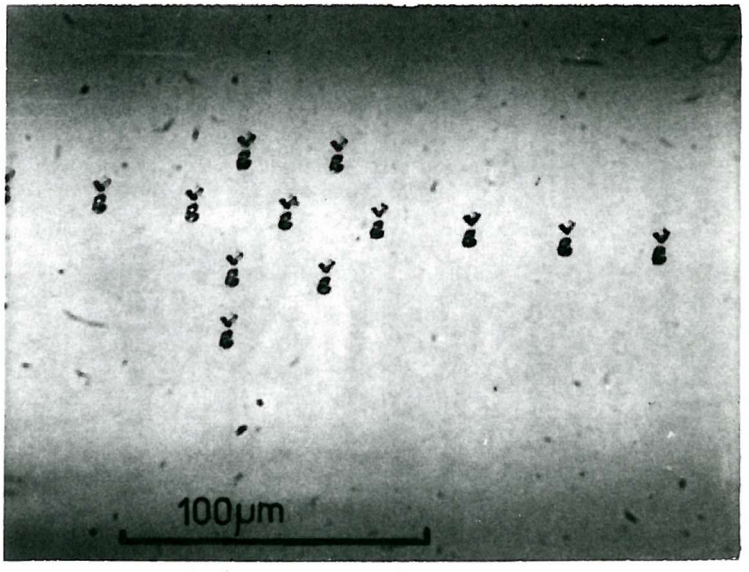


Fig A 9 Spreading Resistance Probe Footprints.

Reproducibility of the actual spreading resistance measurement could not be assessed directly since the damage made by lowering the probe once changes the surface topograhy enough to affect the reading if the probe is relowered over the same area. Reproducibility was checked by making measurements as close together as possible (~20μm) - some tolerance in the results must be allowed for small spatial variations in the sample resistivity.

Factors which can affect the reproducibility of the contact area and hence the measurements, were investigated in order to optimise operation of the equipment. These factors are :

Sample preparation prior to measurement
Probe loading
Probe descent rate
Mechanical rigidity of the equipment

Prior to the publication of the spreading resistance symposium proceedings, it appeared that the generally accepted method of producing a bevel for spreading resistance vs depth profiles was to lap the silicon surface and subsequently etch it briefly in an HNO_3:HF solution (e.g. Mazur and Dickey, 1966 ; Brotherton and Rogers, 1972). Calibration samples were prepared in the same way. Initial studies using such a surface preparation gave results which suffered from random scatter. Readings taken close together could differ by up to 20% without observable trends, suggesting that the scatter was 'noise' rather than a true variation in sample resistivity. A polishing technique was developed to overcome this problem. The experiments and results are described in a publication which is included as the next section of this appendix.

Experiments on probe loading and descent rate were carried out on the first set of large tipped probes (20μm tip radius, tungsten carbide needles). When the smaller tipped probes were installed (see A5), results inferred from the behaviour of the larger probes, appropriately scaled down, were used to optimise the operation of the equipment once more. The comprehensive results obtained with the larger probes are

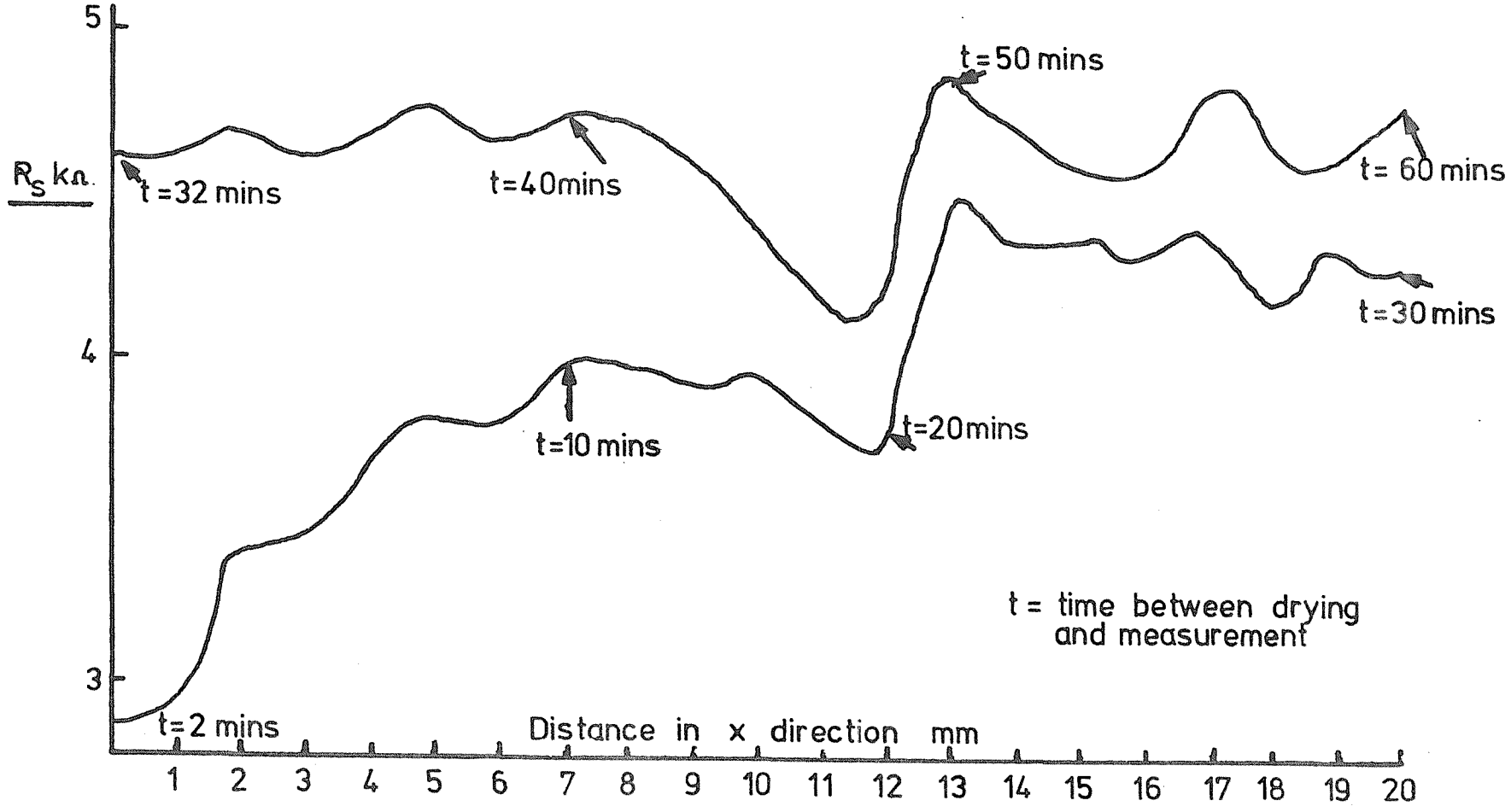


FIG. A10: Drift with time of spreading resistance measurements on a polished sample immediately after cleaning.

summarised in section A4.3 and a brief description of modifications necessitated by the use of the smaller probes is then given.

A4.2 Surface Preparation

(See publication printed on subsequent pages).

A.4.2 Additional comments on Surface Preparation

The ageing effect which is described in the paper is shown in figure A10. In this example an attempt was being made to obtain a resistivity profile as a function of position across the wafer. Two tracks, 100μm apart, were made ; the trends of the resistivity fluctuations are the same in both but an overall rise due to the ageing effect may also be seen. Changes in spreading resistance values of 100% or more have been observed.

Although the bake at 150°C did not entirely eliminate the ageing for the larger tipped probes, it has been found that the smaller tipped probes give stable measurements for many hours. At the time that the probes were changed, the oven used for the bake was also changed; it is not clear if the increased period of stability was due to the different probes or was connected with the actual temperature control in the oven during the bake. It is now only necessary to carry out the bake procedure for 15 - 20 minutes to ensure a surface which remains stable for at least one day.

A 4.2

A new surface preparation and bevelling technique for spreading resistance measurements on N- and P-type silicon

by T. F. Unter and D. R. Lamb†

The use of the spreading resistance technique to obtain quantitative measurements of the local resistivity of semiconductors, particularly silicon, has developed rapidly during the past ten years. The technique is based on the measurement of the constriction resistance[1] of a small area (~ 50 μm²) metal-to-semiconductor pressure contact.

Samples of unknown resistivity are evaluated by comparing spreading resistance measurements taken on them with calibration curves produced by making spreading resistance measurements on "calibration samples" of known resistivity.

This technique and the problems associated with its use have been discussed thoroughly at a Symposium on Spreading Resistance held by the National Bureau of Standards in June, 1974[2]. Some of the major problems arise in the preparation of samples for measurement and the reproducibility of the technique used. Dopant concentration versus depth profiles in silicon are obtained by bevelling diffused samples at a shallow angle and probing down the bevel with spreading resistance equipment. This requires that the spreading resistance reading for a given resistivity on the calibration sample is the same as the spreading resistance reading for the same resistivity on the prepared bevel. This can only be achieved if it is ensured that the condition of the silicon surfaces for both the calibration sample and the bevelled sample are the same and reproducible so that the probes make an identical reproducible contact pattern on both.

1. Experimental results and discussion

Experiments have been carried out with various methods of bevel preparation on both N- and P-type silicon over a range of resistivities 0·01 Ωcm to 100 Ωcm. Spreading resistance readings were obtained with an "Epiprobe" three probe apparatus‡ using tungsten carbide probes loaded with 30 grams; measurements of voltage and current were made with a Keithley 610C electrometer and a Keithley 602 electrometer respectively.

Initially, polished slices (as received from various manufacturers) were used, however difficulties were encountered in obtaining a comparable surface when the samples were bevelled and polished with various grades of diamond paste. This difference was shown by a change of up to 20% in the spreading resistance reading when probing on the top surface and then on the diamond polished bevel of a uniformly doped silicon slice. (This effect has been noted by other workers[3].) A further undesirable effect with the polished samples was a steady increase in the value of spreading resistance with time, starting within a few minutes of spinning the sample dry after cleaning and continuing for a period of forty-eight hours or so. This "ageing" process has also been noted by other workers[2]. see fig A.10.

As a result of this, experiments were carried out on lapped silicon surfaces as described by other authors[4,5]. Lapping was carried out on uniformly (10% of mean resistivity) doped silicon samples with 40 μm particle size carborundum, 13 μm alumina and ½ μm alumina. All samples were cleaned ultrasonically to remove lapping debris. Spreading resistance measurements exhibited ±30% scatter and were not very reproducible with the 40 μm and 13 μm lapping materials. The ½ μm alumina lapped surfaces gave reduced scatter but proved to be extremely difficult to reproduce, meaning that from slice to slice and even from one area of a slice to another the probe contact area was varying.

Figure 1 shows a spreading resistance profile across a nominally 10-20 Ωcm silicon slice lapped with 13 μm alumina, the scatter is ± 30% of the mean value.

Talysurf plots of the 13 μm alumina lapped samples were made to obtain a clearer picture of the nature of

†Electronics Department, Southampton University.

‡Manufactured by Jandell Engineering Limited, England.

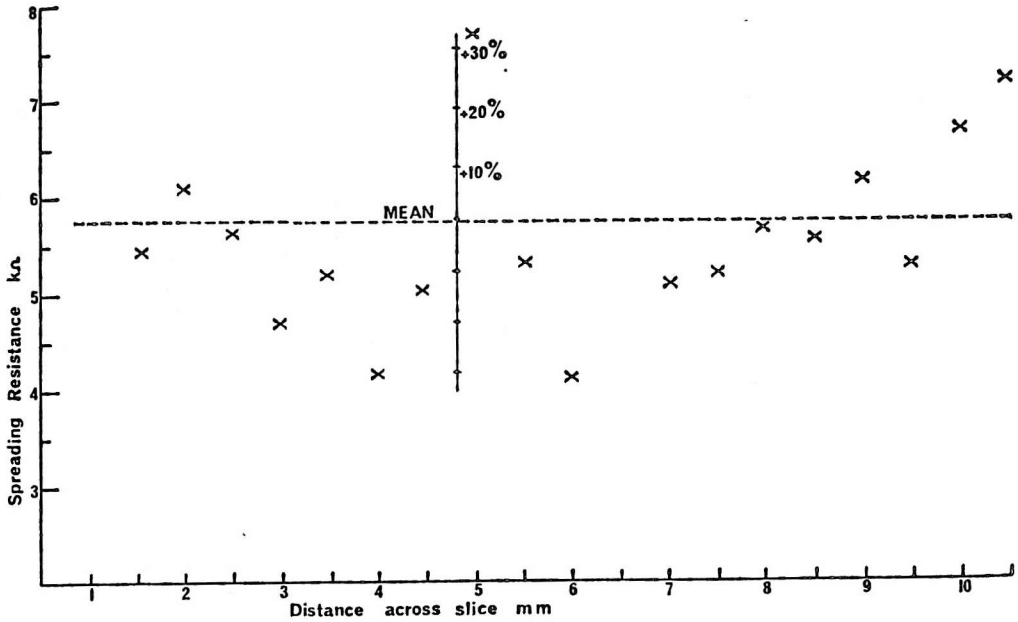


Fig. 1 Spreading resistance profile across a silicon slice lapped with 13 μm particle size alumina.

the surface, these are shown in Fig. 2. As may be seen, many of the pits made by the alumina are comparable in size with the probe tip (the area of contact has a radius of about 10 μm) and so on a microscopic scale the probes are literally "climbing up and down hills", this makes the probe contact area variable and the depth of a particular measurement uncertain.

An investigation of the reproducibility and scatter on the polished surfaces showed that the scatter in the readings would be well within the doping fluctuations expected for the silicon samples if the ageing process could be eliminated. Thus, subsequent experiments were directed to the elimination of this ageing and improvement of the polishing process used to bevel the diffused samples.

Various cleaning procedures were tried with N- and P-type silicon slices to see if the ageing was associated with them. The slices were boiled for 15 minutes in

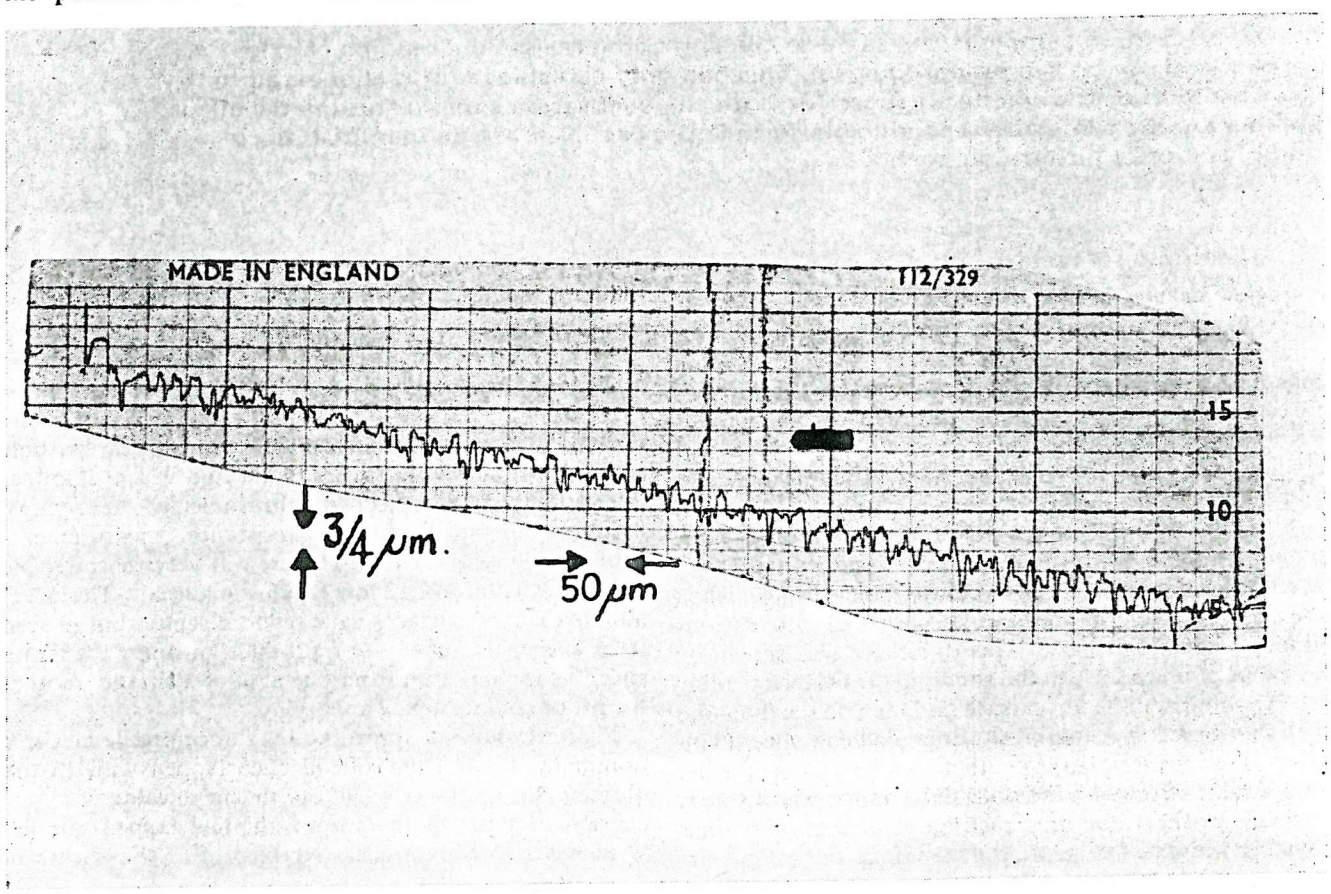


Fig. 2 Talysurf profile of silicon slice prepared as in Fig. 1.

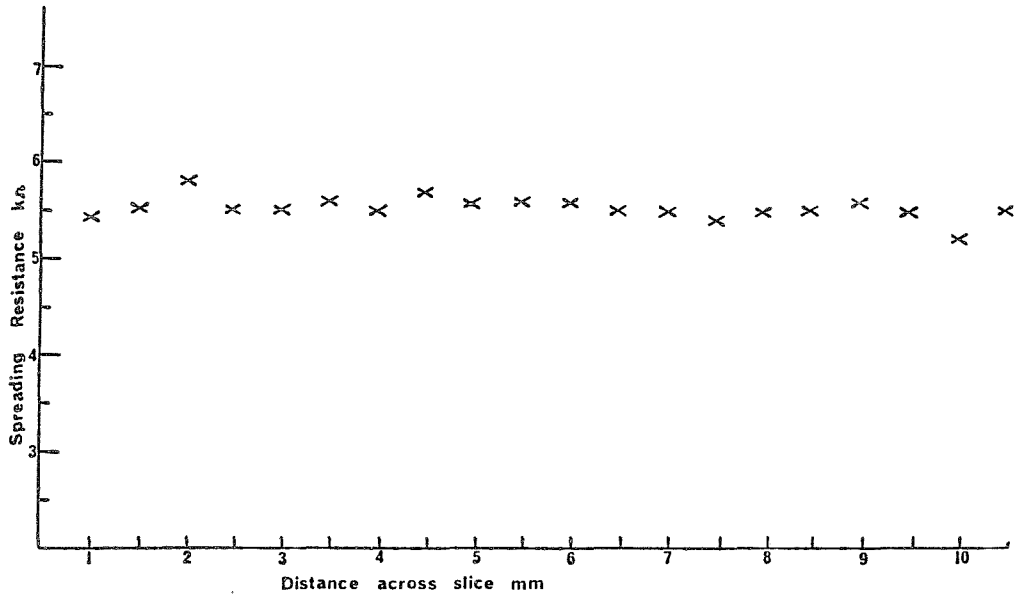


Fig. 3 Spreading resistance profile across a polished silicon slice.

concentrated nitric acid, rinsed in double distilled water and spun dry, and measurements were then made back and forth across the slice surface over a period of several hours. (No two measurements can be made on exactly the same spot as the probes damage the silicon surface.) The sample was then dipped in buffered HF to remove any oxide, rinsed, dried and measured in the same way. The third procedure tried was a sulphuric acid: hydrogen peroxide "bomb" treatment for ten minutes after a buffered HF dip, followed by rinsing and spinning dry. The ageing was most rapid on the slice which had been dipped in buffered HF, and slowest, but still too rapid, on the slice cleaned with the "bomb" treatment.

The rapidity of the ageing process was also seen to be very dependent on the relative humidity of the ambient air; it was slower during dry weather. For this reason it was decided to try "drying" the silicon slices by baking them after the "bomb" clean. Subsequent to a one hour bake at 150°C in air the samples showed no ageing for a period of 35 minutes in ambients of average relative humidity (50%). The spreading resistance equipment was then housed in a box with a desiccant to keep it dry and protect it from draughts and sudden fluctuations in temperature and relative humidity; this extended the period of reliable stability to over one hour. After this period a reclean and rebake was found to bring the measured spreading resistance back to within 1% of its original value. Figure 3 shows a spreading resistance profile after this preparation across a 10-20 Ωcm slice from the same batch as that in Fig. 1. On this sample, the fluctuations seen are reproducible to within 1% if measurements are repeated close to the original ones; this indicates that the fluctuations are due to real fluctuations in resistivity rather than to scatter in the measurements.

Having reduced the ageing of the polished surfaces, to an acceptable degree, the variation in spreading resistance in going from an already polished top surface to a polished bevel was investigated. Careful examination of the diamond polished silicon surface, especially after a short etch, showed that the surface was covered with fine scratches. To overcome this a chem-mechanical polish such as that used by silicon manufacturers to

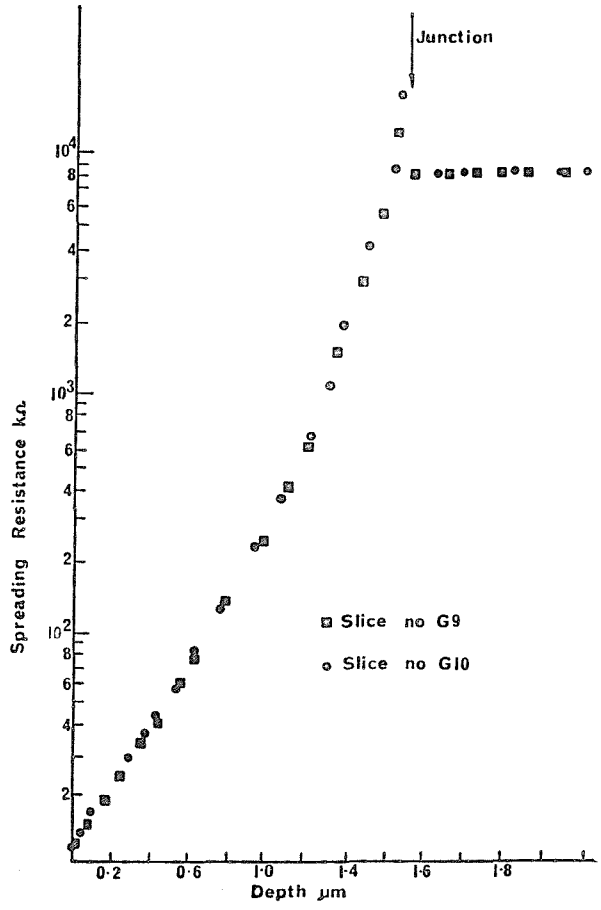


Fig. 4 Spreading resistance profiles of two phosphorus diffused silicon slices (1050°C, $POCl_3$ source, 5 mins deposition, 35 mins drive-in in nitrogen).

polish the silicon top surfaces was used. The samples were mounted on a simple bevelling jig made of stainless steel to the required angle (1°). The bevel was first ground with 40 μm particle size carborundum suspended in water, ultrasonically cleaned to remove lapping debris, partly polished with ½ μm alumina suspended in water and ultrasonically cleaned again. It was then polished in a bowl polisher on a Corfam pad with a steady flow of Lustrox 1000 polish washing over it*. The result after fifteen minutes of polishing was a scratch free polished bevel, visually indistinguishable from the top surface of the silicon. After washing, cleaning and baking as described above, uniformly doped slices returned the same spreading resistance values on the polished bevel as on the original top surface to within 1% (i.e., well within the doping fluctuations expected in the slices).

The requirements of a stable surface for the calibration samples and the ability to reproduce this surface on shallow angle bevels have thus been achieved for both N- and P-type silicon over the range 0.01 Ωcm to 100 Ωcm, and accurately reproducible diffusion profiles have been obtained.

Figure 4 shows uncorrected spreading resistance versus depth profiles for two phosphorus diffused silicon samples which were diffused together but bevelled and profiled separately. The depth measurements were obtained by interference microscopy.

*Lustrox is a chem-mechanical silicon polish and is a trademark of the Tizon Chemical Corporation, Flemington, N.J., U.S.A.

2. Conclusion

By an investigation of polishing, cleaning and baking processes it has been possible to significantly reduce two of the most serious problems of spreading resistance measurements on silicon. These are the production of a reproducible bevel which is indistinguishable electrically and visually from the calibration slice top surfaces and the "ageing" of the results. This has enabled a reproducibility of ± 1% to be achieved when making such measurements on both N- and P-type silicon in the range 0·01 Ωcm to 100 Ωcm.

3. References

[1] Holm, R., "Electric Contacts", 4th ed., Springer-Verlag, New York, 1967.

[2] Proc. Spreading Resistance Symposium, held at the National Bureau of Standards, Gaithersburg, Maryland, U.S.A., 13th-14th June, 1974.

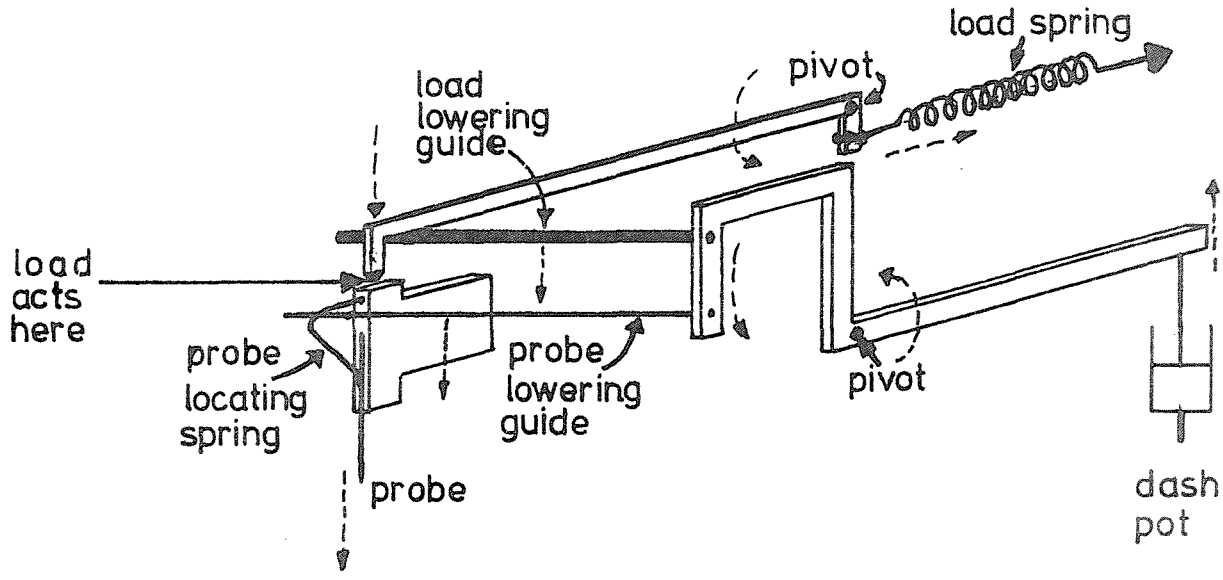
[3] Severin, P. J., Bulle, H., Poodt, G., and Wasscher, J. D., "On the Technique and Evaluation of Angle-Bevelling Silicon Epitaxial Layers", *J. Electrochem. Soc.*, **122**, pp. 440-443, 1975.

[4] Mazur, R. G. and Dickey, D. H., "A Spreading Resistance Technique for Resistivity Measurements on Silicon", *J. Electrochem. Soc.*, **113**, pp. 255-259, 1966.

[5] Severin, P. J., "Measurement of Resistivity of Silicon by the Spreading Resistance Method", *Solid-State Electronics*, **14**, pp. 247-255, 1971.

4. Acknowledgements

This work was carried out with financial support from the Science Research Council.



Spreading Resistance Apparatus.

FIG. A11: **Lowering and loading system for spreading resistance apparatus.**

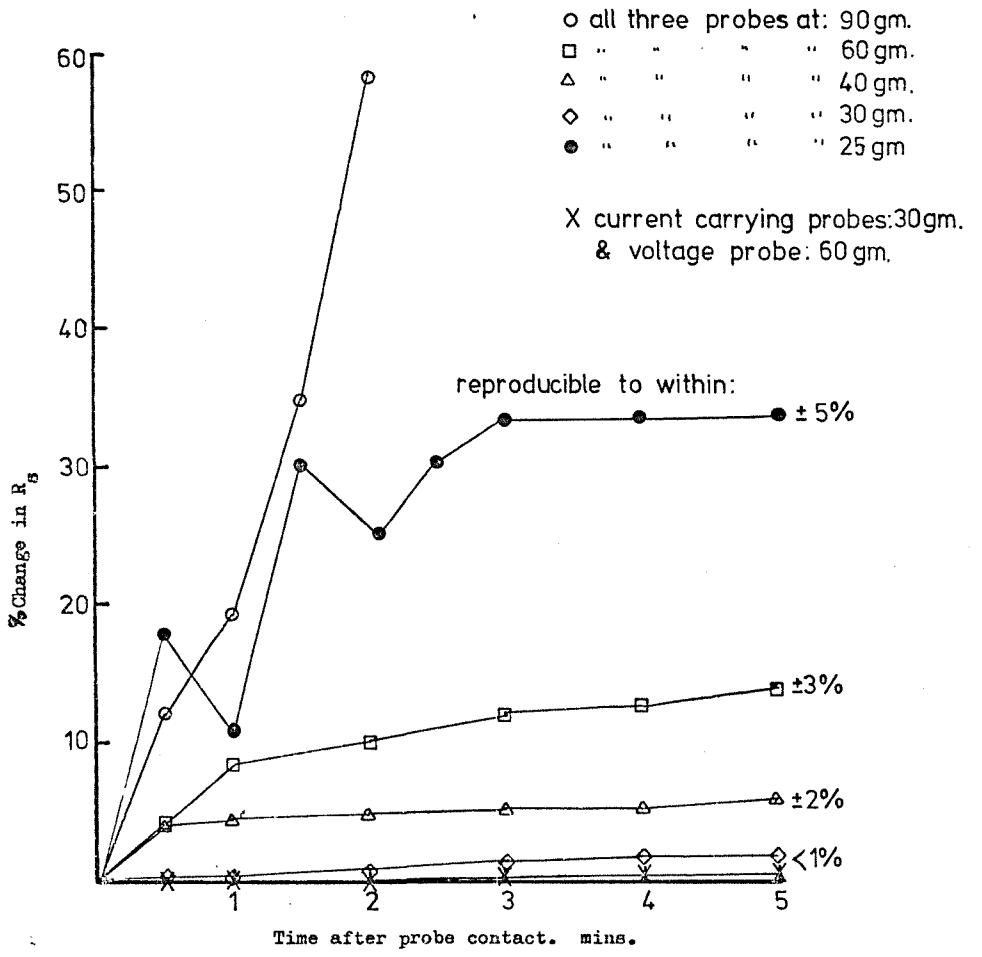


FIG. A12: Drift in spreading resistance measurements for various probe loadings. Polished specimen.
Probe lowering time = 15 seconds (approx.).

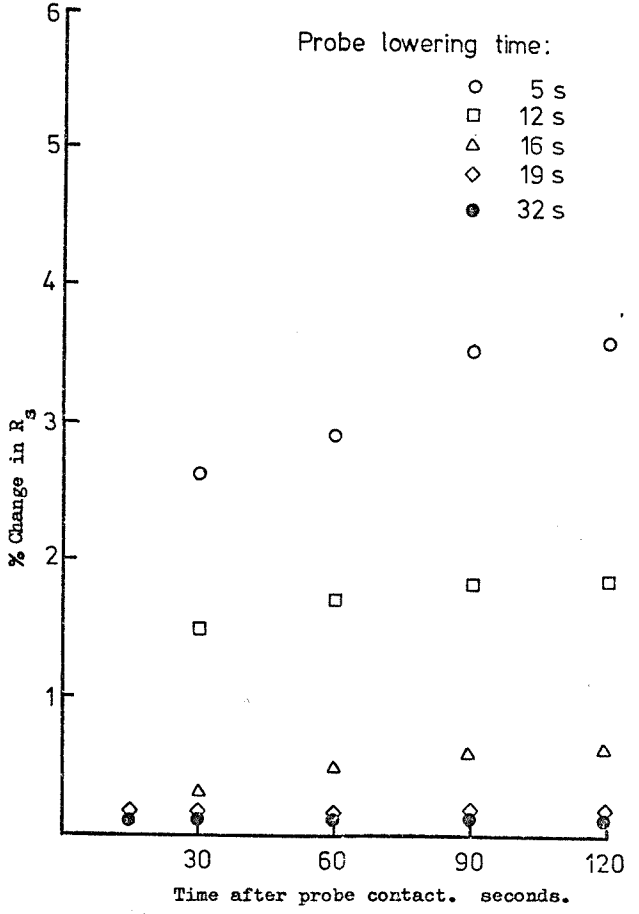


FIG. A13: Drift in spreading resistance measurements for various probe descent times (Loadings are optimum values from figure 12). Polished specimen.

A4.3 Probe Loading and Descent Rate Effects

On the apparatus used (the "Epi-probe" supplied by Jandel Engineering Ltd. UK) the loading arrangement for each probe comprises a calibrated spring operating through a lever arrangement. This is sketched in figure A11. Also shown is the adjustable air dash-pot which controls the lowering rate. The system has been arranged so that the probes lower and make contact with the silicon surface under their own weight plus that of the probe support ($\sim 1^2/3$ grams) ; after contact has been made with the sample surface the loading arms are brought to bear on the tops of the probes as shown. The initial probe load settings were 60 grams (recommended by T. L. Rogers - see Brotherton and Rogers, 1972). Current and voltage measurements were made and it was found that there was appreciable drift of both over a period after lowering the probes. This drift was measured as a function of time for various probe loadings ; results are shown in figure A12. These experiments were carried out on n-type and p-type silicon wafers in the resistivity range 1-20 Ω cm). Also shown is the reproducibility of results after the drift had stabilised. In this figure, probe 1 is the critical measuring probe, probe 2 is the current return path and probe 3 is the voltage measuring probe. (The lines joining the experimental points have no significance).

Similar experiments on drift as a function of probe descent time are shown in figure A13 for the optimum probe loadings.

The operating parameters which gave the least drift and best reproducibility were : probe 1 = 30 g, probe 2=30g, probe 3 = 60g, lowering time ≈ 20 seconds, readings taken after 25 secs.

It is suggested that the drift effects noted when changing the loadings and descent rates, may be due to microcracking of the silicon surface beneath the probes. Runyan (1965) shows that silicon with a chipped surface (a possible result of a high descent rate) is much more likely to crack under a given stress than silicon with a perfect surface.

Tilt = 70°

10 microns

FIG. A14: Electron micrographs of spreading resistance probe "footprints".

The reproducibility of the probe contact patterns on a microscopic scale is clearly evident in figure A14 which shows two different foot prints on different samples from the same probe. These photographs were taken in the Scanning electron microscope.

A4.4 Mechanical Vibration

During all measurements the equipment has been found to be very sensitive to mechanical vibrations. To minimise this the apparatus has been mounted on impact absorbing rubber on a sturdy bench. It is only necessary, however, for someone to slam a door in the viscinity to upset a reading which may be in progress.

A5. Production of Smaller Probes

The spatial resolution of the spreading resistance measurements is most dependent on the radius of contact between probe and silicon. In order to increase the resolution of the equipment used here some probes were modified to give a much smaller tip radius. This should not only enhance the spatial resolution but should also allow measurements to be made much closer together without the probe footprints overlapping.

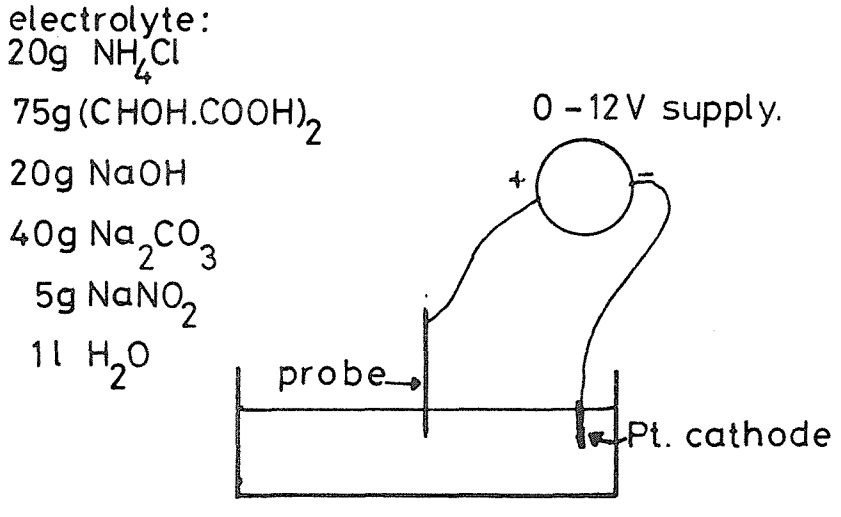


FIG. A 15: Probe sharpening by electrolysis.

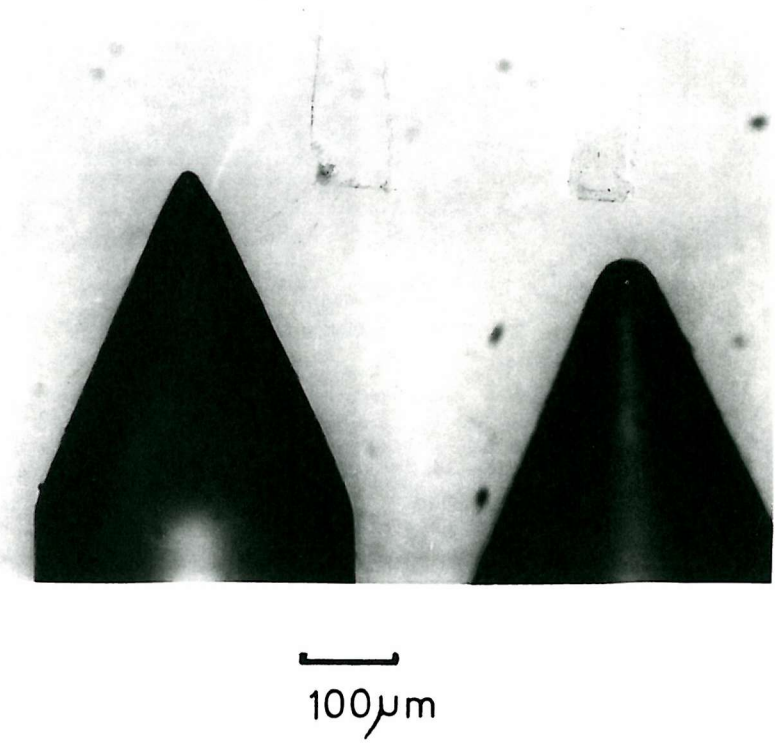


FIG. A16: Comparison of new 'sharpened' probe with as-received probe.

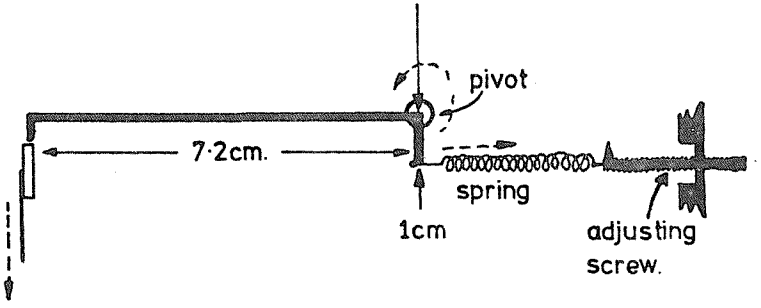
Sharpening of the probes was achieved by electrolysis, as shown in figure A15. The electrolyte was based on a solution used by Lehtinen (1968) to thin tungsten carbide to electron transparency for use in the TEM. A ten second electrolysis resulted in the probe tip shown, compared to a new 25μm tip radius probe, in figure A16. The tip radius is less than 5μm. Results from this sharpening procedure were erratic and of four probes, two became much sharper and two became much blunter. The two sharp probes are still in operation in the equipment and although the initial tips appeared - from measured values of spreading resistance - to wear somewhat, their performance after several months of use has become very stable and reproducible.

The smaller probes are much more sensitive to variations in loading and lowering and several modifications to the equipment have been necessary.

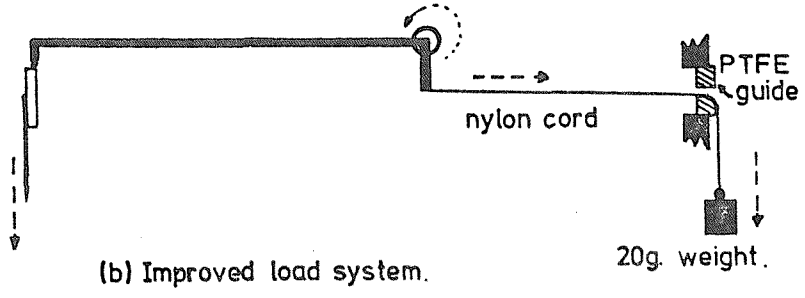
A5.1 Loading of Smaller Probes

A scaling down of the optimum loading on the large probes indicated that the small probes should be loaded with about 5 grams. Several unforeseen problems were encountered with the loading and probe guidance systems of the apparatus. A 'hi-fi' stylus balance was used to measure the loading effects accurately to within 0.1 grams.

The first problem resulted from the spring and lever arrangement used to load the probes. This is shown again in figure A17 (a) with some of the important dimensions. It can be seen that lowering and raising the probes actually results in a small extension and contraction of the spring. At high spring loads this was insignificant but at low loads the difference in probe height between the top and the bottom of a bevel resulted in a 2.5% change in loading with a consequent variation of 10% or more in the probe calibration. This problem was solved by replacing the spring with a piece of nylon cord and a brass weight as shown in A17 (b). This system gives a load which is independent of probe height and provides a force of 4½ grams in total at the probe tip.



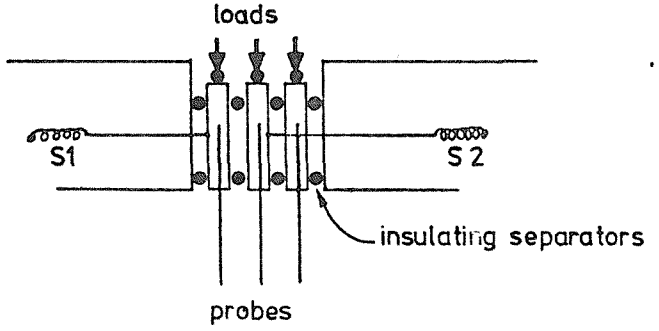
(a) Probe loading system



(b) Improved load system.

FIG. A.17

FIG. A17: Probe loading systems.



S1 & S2 are springs which keep probes laterally constrained.

FIG. A18: Schematic diagram of system used to constrain and insulate the spreading resistance probes.

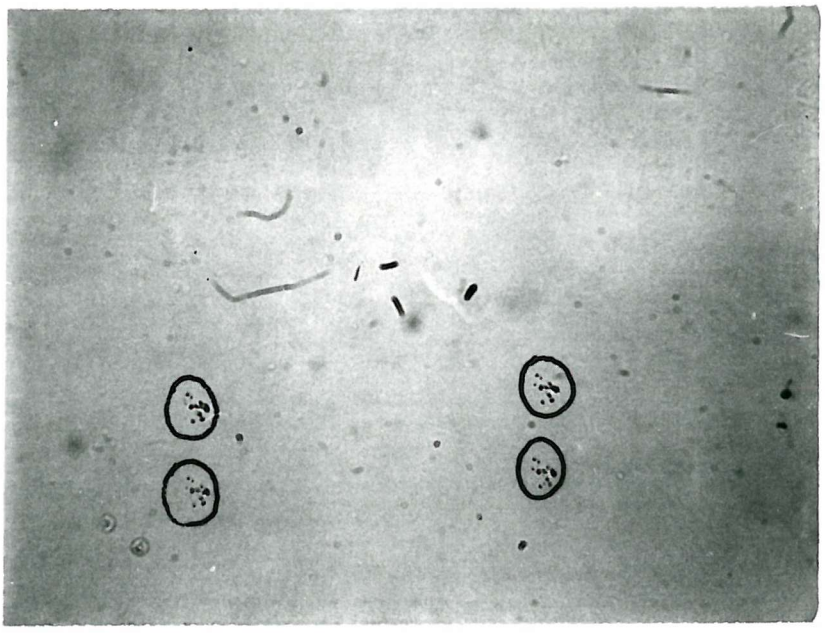


FIG. A19: Probe footprints obtained with smaller probes.

The second problem involved the detailed design of the system used to constrain the probe carrier and guidance frames. This is shown schematically in figure A18. The centre probe was designated by the manufacturer as the critical probe. The two springs S1 and S2 which intentionally pull laterally and in opposite directions also provide a small upward force when the probes are lowered. The magnitude of this upward component was found to be surprisingly large ($\simeq$ 3 grams) when the probes were fully lowered. The right-hand probe is constrained by the lateral force on it exerted by S2 via the central support frame. Measurements showed that even when the central probe was raised relative to the right hand probe, friction was too small to move the right hand probe or reduce its loading. The right hand probe position was chosen for the small critical probe and the other two probes were loaded enough to overcome the upward forces. The loadings now in use are : 60 grams for the voltage probe, 10 grams (nominal with the spring) for the centre probe and 4½ grams for the spreading resistance critical probe.

Figure A19 illustrates the probe footprints from the critical probe. The sample has been lightly etched to make the mark visible.

A5.2 Results with the Smaller Probe

The unexpected advantage of linearisation of the calibration curve has already been mentioned. The increase in resolution is indicated by the ability to approach to within 10 μm of a totally insulating barrier on a nominally uniformly doped wafer before any effect on the readings is noted. This suggests a contact radius of about 2 μm and a consequent sampling depth of 8 μm or so.

On very shallow bevels through n-p junctions, it is possible to make reproducible measurements of the order $10^6\,\Omega$. This implies that the smaller probes are able to make useful measurements of silicon with resistivity greater than 1k Ω cm. In comparison with the four point probe this is a significant advantage since difficulty is almost always experienced with sheet resistance measurements greater than 1kΩ per square, implying a resistivity limit of 500 Ω cm.

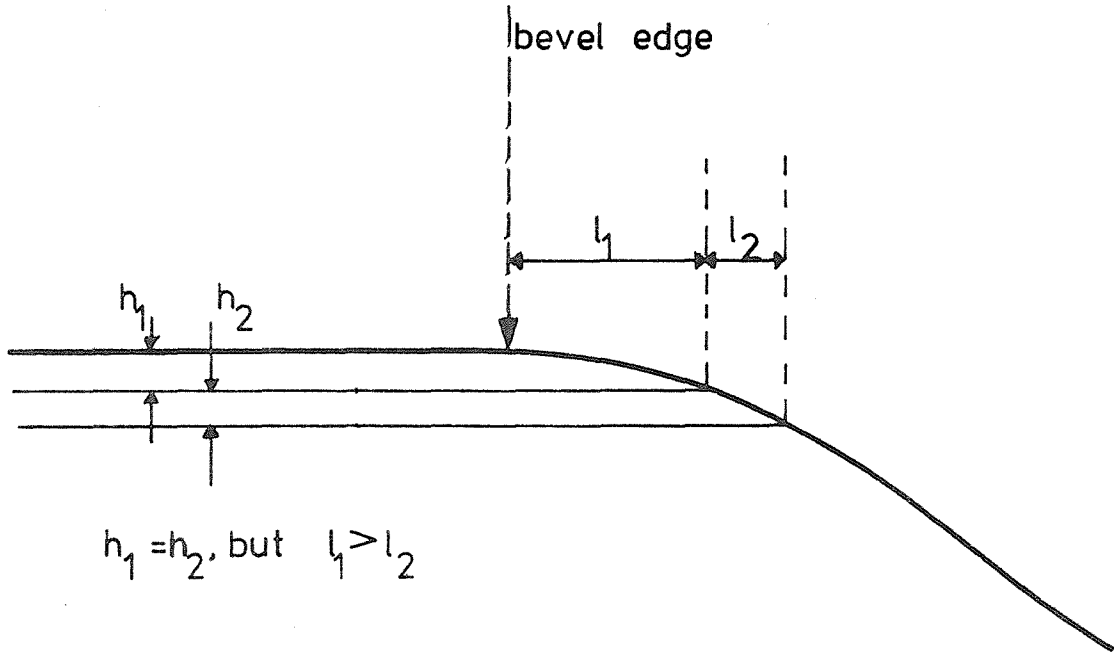


FIG. A20: Cross section through a typical polished bevel.

The increased resolution coupled with the high reproducibility that results from the bevelling technique described in A4.2 is shown, in chapter 8 of this thesis, to allow observation of features of gold profiles that are either lost in the 'noise' due to lapping damage or that vary spatially within the greater sampling depth of the larger probe.

A6. Bevelling and Depth Measurement Technique

The bevelling technique developed for this work has already been described in the publication presented in section A4.2. Figure A20 shows a cross-section through a typical bevel obtained by this method. The bevel does not make a sharp angle with the sample top surface but curves away until it reaches the required angle. In this work bevels of 1^{o} and 4^{o} have been used for gold bulk profiles. For shallow diffused layer profiles a shallow slope would normally be employed, however the rounding of the bevel top has the useful side-effect of magnifying the area in which the shallow diffused layers are situated. The rounded bevel edge would present problems if used with a two-probe spreading resistance apparatus because of difficulties in lining the probes up parallel to an ill-defined edge.

Depth measurements were obtained by two methods. For accurate determination of the depth of each reading, the probe marks on the bevel were observed directly in a microscope employing a beam splitting, Watson interferometer and a sodium lamp source. The interference fringes obtained with sodium light are spaced at .2940μm intervals. It was found that accurate measurements to within a quarter of a fringe or better could be made.

The second depth measurement technique involved the use of the Talysurf equipment which was employed to assess the surface quality of lapped silicon in section A4.2. A profile of the bevel was obtained and the depths of the spreading resistance measurements inferred by superimposing the horizontal increments of the spreading resistance measurements on the horizontal scale of the Talysurf plot. One difficulty associated with this method is the lack of a fiducial mark on

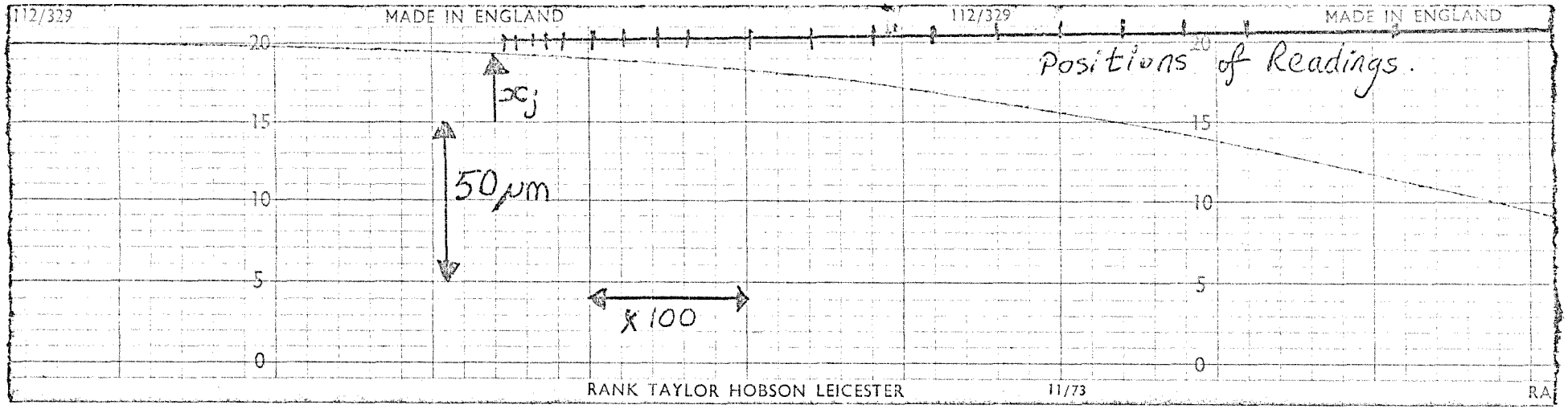


FIG A21 Talysurf profile of bevel

both Talysurf profile and spreading resistance horizontal scale. For samples with shallow junctions, the position of the junction on the bevel was obtained by staining the p-type side with a solution containing dilute hydrofluoric acid and copper sulphate and then measuring the junction depth with the interferometer. The spreading resistance measurement closest to the junction was then lined up with this depth value on the Talysurf profile. For other samples the depths of a few spreading resistance footprints close to the top of the bevel were obtained with the interferometer; the Talysurf profile and horizontal spreading resistance scale were then lined up as before. Checking of one or two profiles with the more lengthy interference fringes method indicates that the depths on bulk profiles are accurate to within about 5μm.

Figure A21 illustrates a typical Talysurf profile of a bevel with the measurement depths marked on it.

A7. Conclusion

Development of a new sample preparation and bevelling technique, coupled with optimisation of the operation of the spreading resistance apparatus, has enabled measurements to be made with high reproducibility. The resolution of the technique has been greatly improved by implementing measurements with smaller probes at lower pressures. Examples of measurements obtained are given in chapter 8. A very simple "correction" routine for spreading resistance measurements on shallow diffused layers has been described.

This technique has also been used successfully in assessment of variations in resistivity across the surface of silicon wafers. Large fluctuations of resistivity on a microscopic scale present problems in the manufacture of large area power devices with deeply diffused junctions. Figure A22 shows two examples. One is a normal Czochralski grown phosphorus doped silicon wafer which exhibits large periodic variations in resistivity, these can result in ragged junctions. The other profile is of a wafer from a crystal grown by the float zone technique with a very low doping level. Doping is carried out by subjecting the silicon crystal to a neutron flux which trans-mutates some of the

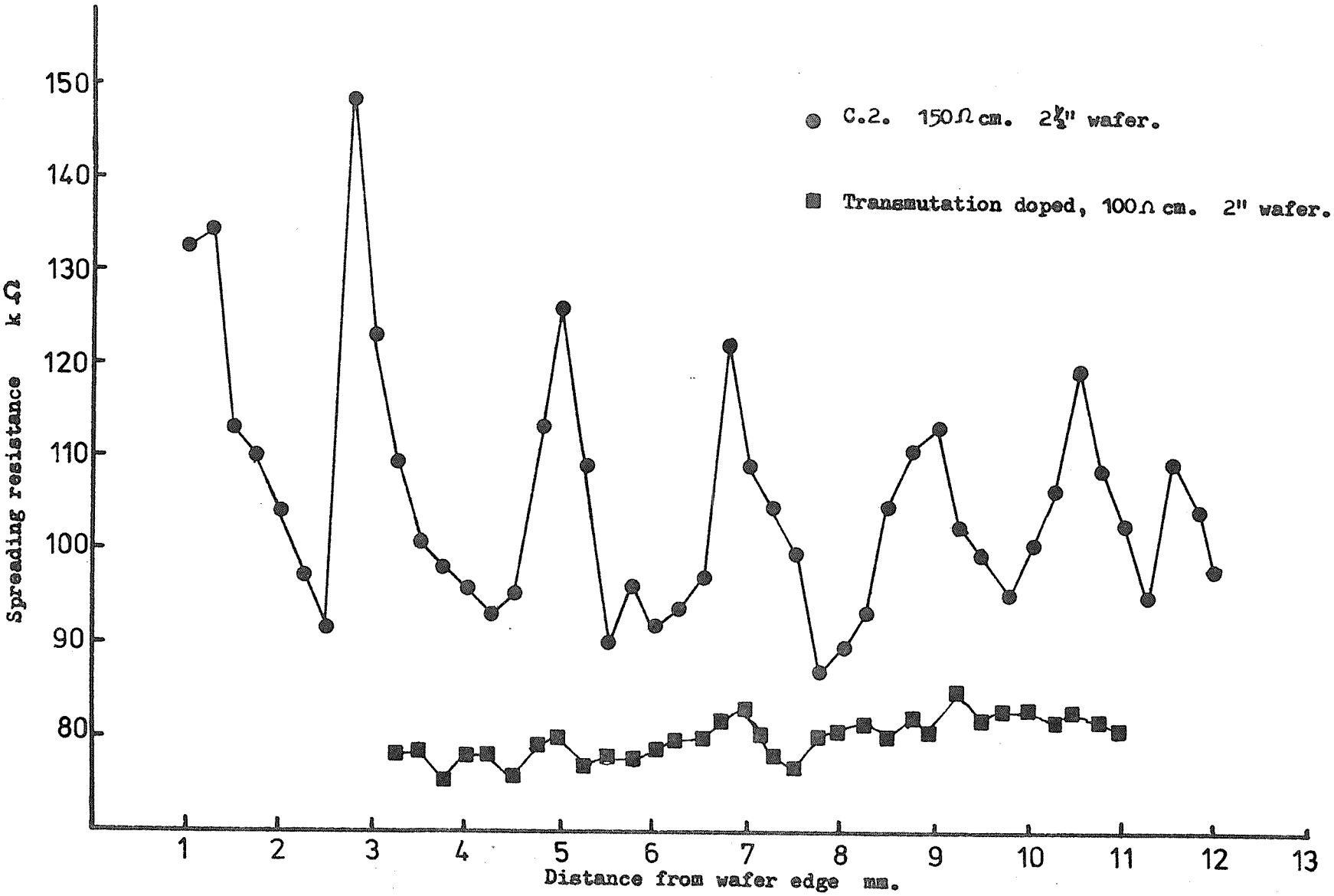


FIG. A22: Spreading resistance profiles illustrating cyclic resistivity variations across silicon wafers.

silicon atoms so that they assume the nature of shallow donors. The result is a very uniform dopant distribution across the wafers cut from the crystal. Neutron-transmutation doping will almost certainly become a major technique in the production of homogeneously doped silicon crystal since, with the advent of LSI and VLSI circuits, the need for such crystals will become universal. Consequently a high resolution technique for their evaluation will also be required.

APPENDIX B

Electrically Active Dopant Profiles by Incremental Sheet Resistance Measurements

A few profiles for this work were obtained using the incremental sheet resistance technique. This method is well documented in the literature and is only described very briefly here since the same analysis was used to convert spreading resistance versus depth measurements on shallow layers to doping concentration versus depth.

As already mentioned in Appendix A, the sheet resistance measurement is carried out using a four-point-probe (Valdes, 1954; Smits, 1958). The sheet resistance of the layer is related to the average conductivity, $\bar{\sigma}$, of the layer by :

$$\bar{\sigma} = \frac{1}{R_{SH}\, x_j} \qquad B1$$

where x_j is the depth below the surface of the insulating barrier (junction).

For a diffused layer, where the actual conductivity varies with depth, the average conductivity of the layer may be written

$$\bar{\sigma} = \frac{1}{x_j} \int_0^{x_j} q\mu c \, dx \qquad B2$$

where q is the electronic charge, μ is a carrier mobility typical of the layer and c is the density of carriers. From equation B1 therefore :

$$\bar{\sigma} x_j = \sigma_{SH} = \int_0^{x_j} q\mu c \, dx = \int_0^{x_j} \sigma(x) \, dx \qquad B3$$

where σ_{SH} = the sheet conductivity of the layer ($= \frac{1}{R_{SH}}$) and $q\mu c$ is the conductivity $\sigma(x)$ of an incremental layer at depth x.

If thin layers are successively removed from the sample surface, and measurements of the sheet resistance are made - or if measurements of R_{SH} down a bevel through the layer are made, the general form of equation B3 is :

$$\sigma_{SH}(x) = \int_{x}^{x_j} \sigma(x)\, dx \qquad B4$$

where x is the depth of the 'new' surface below the original surface. From this equation by differentiation :

$$\frac{d\sigma_{SH}(x)}{dx} = \sigma(x_j) - \sigma(x) \qquad B5$$

at the junction ($x = x_j$) the conductivity is zero since the junction is an insulator, therefore

$$\frac{d\sigma_{SH}(x)}{dx} = -\sigma(x)$$

or

$$\frac{d(R_{SH}(x))}{dx} = -\rho(x) \qquad B6$$

Thus the slope at x of a plot of R_{SH} (the measured quantity) versus depth, yields the resistivity at the depth x.

Generally, the data obtained ranges over several orders of magnitude (typically 10 to 1000 ohms per square) and it is more convenient to plot the values logarithmically. In this case a relationship between resistivity and the slope of log (R_{SH}) at x is desired.

Since

$$\frac{d \log_a(x)}{dx} = \frac{1}{x \ln a} \qquad B7$$

and

$$\frac{d}{dx} \log_a \sigma_{SH} = \frac{d}{dx} (\sigma_{SH}(x)) \cdot \frac{1}{\sigma_{SH}(x) \ln a} \qquad B8$$

$$\frac{d}{dx} \ln \sigma_{SH}(x) = \frac{d}{dx} (\sigma_{SH}(x)) \frac{1}{\sigma_{SH}(x)} \qquad B9$$

Substituting in B6

$$\sigma(x) = -\sigma_{SH}(x) \frac{d}{dx} (\ln \sigma_{SH}(x)) \qquad B10$$

The desired relationship is between R_{SH} and $\rho(x)$. Since

$$\sigma_{SH}(x) = \frac{1}{R_{SH}(x)}$$

then :

$$(\rho(x))^{-1} = -\frac{1}{R_{SH}(x)} \cdot \frac{d}{dx}\left(\ln \frac{1}{R_{SH}(x)}\right) \qquad B11$$

Now,

$$\ln\left(\frac{1}{R_{SH}(x)}\right) = \ln(1) - \ln R_{SH}(x)$$

$$= -\ln R_{SH}(x)$$

$$\therefore \quad (\rho(x))^{-1} = \frac{1}{R_{SH}(x)} \cdot \frac{d}{dx} \ln (R_S(x)) \qquad B12$$

Converting to logarithms of base 10,

$$\rho(x) = \frac{R_{SH}(x) \log_{10} e}{\frac{d}{dx} \log_{10}(R_{SH}(x))} \qquad B13$$

The resistivity $\rho(x)$ may be related to dopant density with Irvin's empirical curves of resistivity versus impurity concentration (Irvin, 1961).

The sheet resistance versus depth distributions were determined with a special four probe head (Fell Manufacturing Company) after removal of successive layers of silicon. The Fell head has special light (60 gram) probe loadings to minimise damage caused to the silicon surface.

The silicon layer removal was effected with a silicon etch containing 99% HNO_3 and 1% HF. Depths were measured by masking one edge of the sample with Apiezon wax before each etch so that a series of steps were formed. Depths were measured by the interference microscopy technique described in A6.

When profiles of R_{SH} versus depth were inferred from spreading resistance data, they were also analysed using equation B3.

The slopes were measured by eye from the plots of log R_{SH} versus depth. A computer program was developed to convert the raw sheet resistance against depth data to impurity concentration profiles, however it was found to be quicker to do the analysis by hand !

APPENDIX C. Statistics of Partially Compensated Gold Doped Silicon

The experimental determinations of the gold energy levels by Collins et al. (1957), Bruckner (1971) and Thurber et al. (1973) which were described in chapter 4, all involve achieving conditions in the silicon which ensure that the dominant temperature dependent factor in the change balance equation (7.1) is an exponential which involves the energy of only a single impurity level. A single example here will serve to illustrate how the energy levels were determined. It is necessary to carry the experiments out at low enough temperatures to ensure that the gold energy levels behave independently ; i.e. they are many kT apart. In n-type silicon, therefore, all of the gold centres may be considered as acceptors unless $N_{Au} >> N_d$.

Consider the situation where the sample contains a density N_{Au} of gold acceptors and a density N_d of shallow donors such that $N_{Au} > N_d$ and, therefore, the material is partially compensated. This arises because electrons from the donor 'drop' into the acceptor levels and only $N_d - N_{Au}^-$ electrons are available for distribution between the donor level and the conduction band. It is assumed that all of the shallow donors are ionised.

The electron concentration (using nondegenerate approximations) is given by equation 7.2

$$n = N_c \exp - \left[\frac{E_c - E_F}{kT}\right] \qquad C1$$

and the density of ionised gold acceptors is given by the Fermi probability function (equation 4.11)

$$N_{Au}^- = \frac{N_{Au}}{1 + g_A \exp\left[\frac{E_A - E_F}{kT}\right]} \qquad C2$$

An expression for the density of neutral gold acceptor centres may also be written :

$$N_{Au}^x = \frac{N_{Au}}{1 + \frac{1}{g} \exp\left[-\frac{E_A - E_F}{kT}\right]} \qquad C3$$

where $N_{Au} = N_{Au}^{-} + N_{Au}^{x}$ C4

Electron summation gives :

$$n + N_A^{-} = N_d \qquad C5$$

since in this situation all of the shallow donors will be ionised - see chapter 7.

A little algebraic manipulation of these expressions may be used to obtain the desired relationship linking the carrier concentration to the position of the gold acceptor level.

The ratio $\frac{nN_{Au}^{o}}{N_A^{-}}$ is given by :

$$\frac{n.N_{Au}^{o}}{N_A^{-}} = \frac{n.N_{Au}\left[1 + g \exp\left[\frac{E_A - E_F}{kT}\right]\right]}{\left[1 + \frac{1}{g}\exp\left[-\frac{E_A - E_F}{kT}\right]\right].N_{AU}} \qquad C6$$

dividing throughout by $\frac{1}{g}.\exp.\left[-\frac{E_A - E_F}{kT}\right]$, gives

$$\frac{n.N_{AU}^{o}}{N_A^{-}} = \frac{n\left[1 + g \exp\left[-\frac{E_A - E_F}{kT}\right]\right]\left[g \exp\left[\frac{E_A - E_F}{kT}\right]\right]}{1 + g \exp - \frac{E_A - E_F}{kT}} \qquad C7$$

$$= n.g \exp\left[\frac{E_A - E_F}{kT}\right] \qquad C8$$

Substituting for n from equation C1 :

$$\frac{nN_A^{o}}{N_A^{-}} = g.N_c \exp\left[-\frac{E_c - E_A}{kT}\right] \qquad C9$$

using equations C4 and C5 :

$$\frac{n(N_{Au} - n - N_d)}{n + N_d} = g\,N_c \exp\left[-\frac{E_c - E_A}{kT}\right] \qquad C10$$

The compensation conditions and low temperature are chosen so that $n \ll N_d$ and $n \ll N_{Au} - N_d$, which yields the desired relationship

$$n = \frac{N_d}{N_{Au} - N_d} \; g \, N_c \exp\left[- \frac{E_c - E_A}{kT} \right] \qquad C11$$

Similar derivations may be carried out for compensation conditions in which the other energy levels of interest are dominant. The expressions used by the authors cited above may all be obtained in this way.

APPENDIX D : Equilibrium Statistics of Recombination and Generation through a Deep Trap

The transitions of carriers between the conduction and valence bands via an intermediate 'trap' energy level were depicted in figure 4.12. The theory of recombination-generation processes via such traps has been dealt with in detail by Hall (1952) and Shockley and Read (1952). The basic equations governing the emission processes via such traps were quoted in Chapter 4 (equations 4.50 and 4.51) and are derived here.

With reference to figure 4.12, the four processes are considered individually. The rate of the electron capture process (a) should be proportional to the concentration, n, of free electrons in the conduction band and also to the concentration of traps which are not already occupied by electrons. If the concentration of traps is N_t per m^3, then the number which are unoccupied is given by equation 4.3 as : $N_t(1 - F(E))$; where F(E) is the Fermi probability function (equation 4.9). If E_t is the trap energy level and E_F is the Fermi level; then the rate of process (a) is given by :

$$r_a \alpha n N_t (1 - F(E)) \qquad D1$$

The constant of proportionality, which must have dimensions of m^3/second, is defined as the product $V_{th}\sigma_n$, where V_{th} is the thermal velocity of the carriers ($= 3\sqrt{kT/m}$) and σ_n is the capture cross section of the trap for electrons. This quantity may be interpreted as a measure of how close the electron must come to the trap in order to be captured. As one might expect, the magnitude of σ_n is generally of the order of atomic dimensions ($10^{-19} m^2$).

D1 may be written :

$$r_a = V_{th}\ \sigma_n\ n\ N_t\ (1 - F(E)) \qquad D2$$

The emission process, (b), may be considered much more simply since the emission rate will be proportional to the concentration of centres which are occupied by electrons: $N_t.F(E)$. The constant of proportionality is defined as the emission probability e_n so that:

$$r_b = e_n\ N_t.F(E) \qquad D3$$

One would expect e_n to depend on the density of unoccupied states in the conduction band and on the location of the trap level since the closer it is to E_c the higher the probability of emission; it will be shown below, that this is indeed correct.

By analogy to process (a), the capture rate of holes, process (c) is given by :

$$r_c = V_{th}\sigma_p \; p \; N_t.F(E) \qquad D4$$

and by analogy with (b), the rate of process (d) is given by:

$$r_d = e_p \; N_t \; (1 - F(E)) \qquad D5$$

Once again, it is expected that e_p is related in some way to the density of centres not occupied by holes in the valence band and the location of the trap relative to the valence band.

In equilibrium - i.e. with no external generation of carriers - the rates of the two processes describing transitions from a particular band must be equal, i.e. $r_a = r_b$.

Equating D2 and D3 and using the relationship for the electron concentration in equilibrium (equation 7.2) :

$$n = N_c \exp - \left[\frac{E_c - E_F}{kT}\right] \qquad D6$$

and equation 4.1 for F(E), the following relationship is obtained :

$$e_n = g_t \; V_{th} \; \sigma_n \; N_c \exp - \left[\frac{E_c - E_t}{kT}\right] \qquad D7$$

It is seen that, as expected, e_n is related to the density of states N_c and the trap level relative to E_c; g_t is the degeneracy factor of the trap.

One may similarly obtain an expression for e_p by equating r_c and r_d:

$$e_p = \frac{1}{g_t} \cdot V_{th} \; \sigma_p \; N_v \exp - \left[\frac{E_t - E_v}{kT}\right] \qquad D8$$

Equations D7 and D8 are those used in chapter 4.

APPENDIX E. Statistics of Occupancy for the Gold Energy Levels

In situations where the gold energy levels interact (e.g. in Chapter 7) the simple occupancies of the levels predicted by the Fermi probability function (equation 4.1) must be modified to take into account the possibility of a given impurity centre being occupied in more than one way. (i.e. gold can be an acceptor or a donor). Shockley and Last (1957) have developed a general theory of the relative occupancies of the states of an impurity which gives rise to a number of energy levels in a semiconductor energy band gap. The relative probabilities of different charge conditions on the impurity are obtained by finding the charge distribution that maximises the probability. The resulting relationship for the relative concentrations of two adjacent charge states (i.e. differing by one increment of charge) is given by :

$$\frac{N_{j-1}}{N_j} = \frac{g_{j-1}}{g_j} \exp\left[\frac{E_F - E}{kT}\right] \qquad E1$$

where N_j = density of impurities in charge state j
N_{j-1} = density of impurities in charge state j-1
g_j = degeneracy of the jth charge state
E_F = the Fermi energy
E = the energy required to convert the impurity from the jth state to the j-1th state of charge

Substitutional gold may exist in any of three charge states, +1, 0 and -1 electronic charges. The energy required to go from the acceptor ionised state to the neutral state is E_A and that required to go from the neutral state to the donor ionised state is E_D. If there are N^+ gold acceptors, N^x neutral gold atoms and N^- gold donors, then, taking the neutral situation as the energy zero :

$$N^+ : N^x : N^- = g^+ \exp\left(\frac{E_D - E_F}{kT}\right) : g^x : g^- \left(\frac{E_F - E_A}{kT}\right) \qquad E2$$

where g^+, g^x and g^- are the degeneracies of the appropriate charge states. From this expression the following relationship is obtained:

$$\frac{N^+}{g_D \exp\left[\frac{E_D - E_F}{kT}\right]} = \frac{N^-}{g_A^{-1} \exp\left[\frac{E_F - E_A}{kT}\right]} = N^x \qquad E3$$

where $g_A = \frac{g^x}{g^-}$ and $g_D = \frac{g^+}{g^x}$ (see chapter 4).

The total gold concentration is made up of these three charge states giving :

$$N^+ + N^x + N^- = N_{Au} \qquad E4$$

To obtain the proportion of total gold which is in the positive charge state the above relationships are used as follows:

$$N^- = N_{Au} - N^+ - N^x \qquad E5$$

therefore, from E3 substituting in terms of N^+ for N^- and N^x we obtain:

$$\frac{N^+}{g_D \exp\left[\frac{E_D - E_F}{kT}\right]} \cdot \left[g_A^{-1} \exp\left[\frac{E_F - E_A}{kT}\right]\right] =$$

$$N_{Au} - N^+ - \frac{N^+}{g_D \exp\left[\frac{E_D - E_F}{kT}\right]} \qquad E6$$

rearranging :

$$N_{Au} = N^+ \left[\frac{g_A^{-1} \exp\left[\frac{E_F - E_A}{kT}\right]}{g_D \exp\left[\frac{E_D - E_F}{kT}\right]} + 1 + \frac{1}{g_D \exp\left[\frac{E_D - E_F}{kT}\right]} \right] \qquad E7$$

therefore

$$N^+ = N_{Au} \left[1 + \left[g_D^{-1} \exp\left[\frac{E_F - E_D}{kT}\right]\right] \left[1 + g_A^{-1} \exp\left[\frac{E_F - E_A}{kT}\right]\right] \right]^{-1} \qquad E8$$

This expression, and a similar one for N^-, were used in the solution of the charge balance equation in Chapter 7.

APPENDIX F : Determination of Thermal and Optical Emission Rates of Deep Level Impurities

Determinations of the emission and capture rates of the gold acceptor and donor levels have been reviewed in Chapter 4 for cases where their temperature dependences have been used to obtain the activation energies of the gold levels. In this appendix a brief summary of the background to the techniques is given. This appendix is based on work described by C. T. Sah in numerous papers. A good review of all of the techniques pioneered by the group working with Sah may be found in Sah (1976).

The basic S-R-H emission and capture processes have been discussed in Appendix D and Chapter 4. The carrier generation-recombination-trapping processes at imperfection centres, such as gold, can be characterized by six parameters:

c_n the electron capture probability
e_n the electron emission probability
E_n the electron energy change during the transition.

A similar set of three parameters, c_p, e_p and E_p characterize the hole transitions. The six parameters were characterized in appendix D.

A very simple consideration of these processes will illustrate the basis of the techniques. If, by some means, a proportion of the traps are filled with electrons (process (a)) and the material subsequently depleted of free carriers, the only occurrence observed will be the emptying of the traps via process (b). The rate of this emptying process will depend on, n_T, the number of traps filled and e_n, the emission rate (see appendix D). A rate equation may be written :

$$\frac{dn_T}{dt} = n_T e_n \qquad \text{F1}$$

The solution of this indicates that the rate of emission of electrons from the traps will be exponential with a time constant $1/e_n$. If the conditions described above can be set up, the measurement of this time constant allows the emission rate to be obtained without the necessity of knowing the trap concentration. These conditions are easily arranged in p-n junctions, Schottky barrier diodes and MIS capacitors. A number of techniques based on this simple effect have been described by Sah et al. (1970).

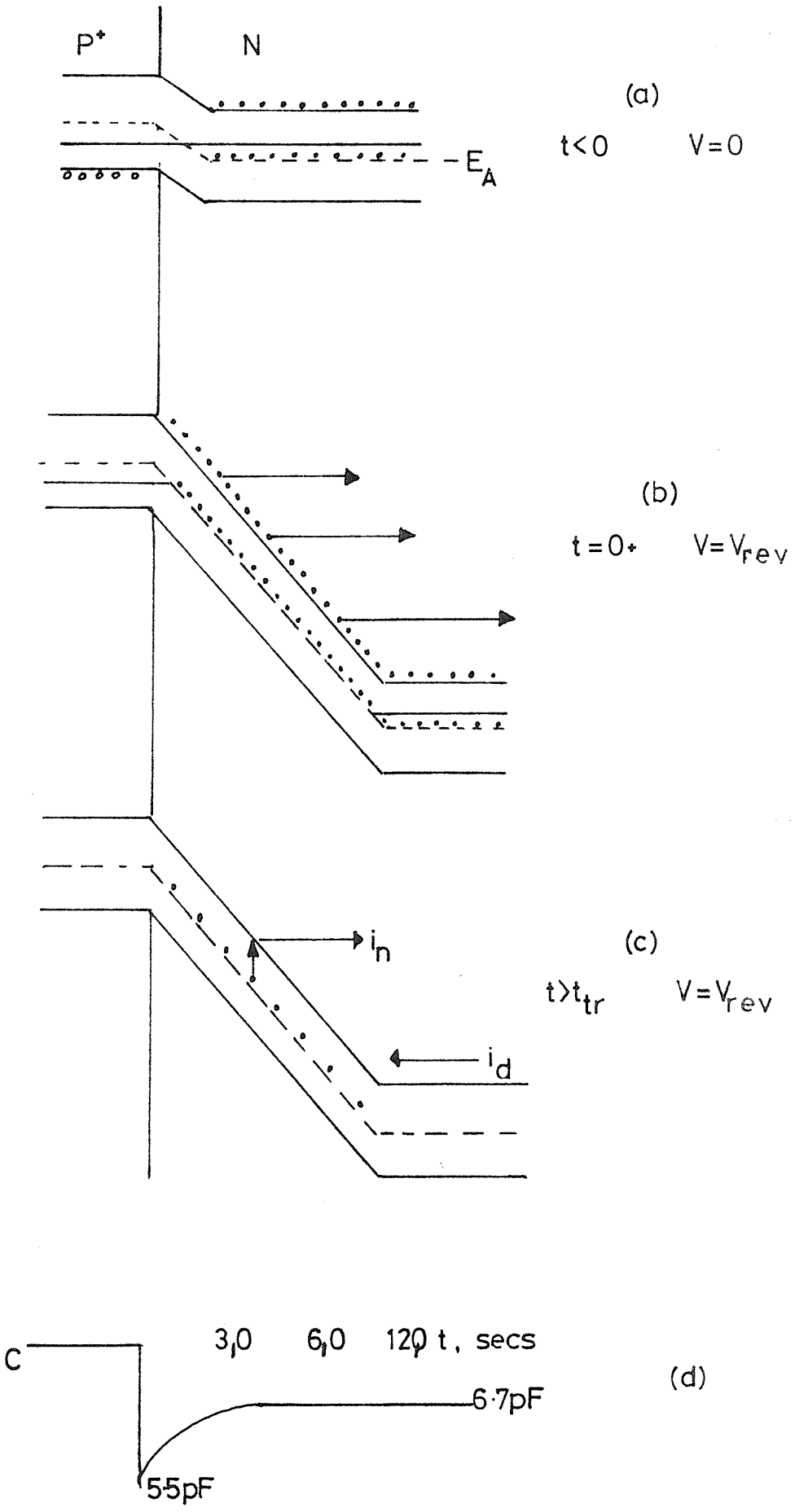


FIG. F1: Emission of electrons trapped at gold acceptor in depleted region of a reverse biased p-n junction. T= -65°C (Sah, 1976)

The chosen device structure, which is doped with the impurity of interest, is operated in such a way as to fill the traps with the carrier of which the emission rate is required. The structure is then depleted of free carriers by reverse biasing in the case of junctions or switching into depletion in the case of MIS structures. The transient caused by the emission process is measured by monitoring the high frequency capacitance of the structure (which is proportional to the depletion width) or by monitoring the transient current flow. One of the techniques is described briefly below in order to illustrate the modus operandi.

F1. Dark Capacitance Transient

The sequence of events in the dark capacitance transient experiment is shown in figure F1. A Schottky barrier diode (on n-type silicon) or a p^+-n junction diode is illustrated. In the first stage (F1(a)) the diode is zero biased and the generation-recombination-trapping centres are filled with electrons. The diode is then switched into reverse bias (V_R). The free carriers (electrons) on the 'n' side of the junction are immediately swept away by the electric field to leave a depletion region (F1 (b)). The transit time for this process is very small : 10^{-12} seconds. The situation described earlier has now been achieved, i.e. there are no free carriers. Trapped electrons are now emitted from the g-r-t centres with a time constant defined as above. This rate may range from a few thousand per second to a few per hour depending on the temperature and thermal activation energy. When the trapped electrons are excited into the conduction band they are immediately swept out of the depletion region (fig. F1 (c)). (If the trap is one which can trap both electrons and holes, then holes would be released and swept out of the depletion region in the opposite direction, for simplicity it is assumed that the electron trapping and emission process is dominant). If an electron is released and swept out of the space charge region in this manner, the net space charge within the region is increased. This increase reduces the width of the space charge layer and hence increases the space charge high frequency capacitance. The time constant of the capacitance increase will, therefore, yield the time constant of the emission process which is the reciprocal of the electron thermal emission rate. A typical capacitance transient is shown in figure F1 (d).

This measurement yields the kinetics of trapped majority carriers at traps whose energy levels are located in the majority carrier half of the band gap. Care must be taken in such measurements to ensure that high electric fields in the depletion region are not enhancing the emission rate although deliberate application of high fields does allow the emission rate field dependence to be observed.

The other techniques used by various authors mentioned in Chapter 4 are all based on this method.

APPENDIX G. Solubility Enhancement Caused by Excess Vacancies

The assumption embodied in equation 6.58, that any solubility enhancement caused by excess vacancies affects all gold atoms, appears to contradict the previous statement that, at the solubility limit, the neutral gold concentration, Au^x, is invariant (section 6.2.1). Taking this into account, however, has a negligible effect on the final result.

The model proposed in chapter 6 - that all of this 'extra' enhancement is in the form of $(Au^-.P^+)$pairs - implies that all of the additional gold atoms are negatively charged. The total amount of gold may be written as :

$$Au(total) = Au^x + Au^+ + (Au^-)\chi \qquad G1$$

where Au^-includes the Fermi-level enhancement and ion-pairing equilibrium effects. In very heavily-doped n-type silicon most of the gold is negatively charged (the Fermi level is close to E_c) and any contributions to the total concentration from Au^x and Au^+ will be negligible, i.e. in the absence of the vacancy enhancement effect :

$$Au^- >> Au^x + Au^+ \simeq Au^{TOT}(intrinsic).\left[\begin{matrix}\text{Fermi}\\ \text{Level}\\ \text{Effect}\end{matrix} + \begin{matrix}\text{Ion}\\ \text{Pairing}\\ \text{Effect}\end{matrix}\right] \qquad G2$$

and in the presence of the excess vacancy effect, we may write (as in equation 6.58)

Total gold concentration = $(Au^-)\chi =$

$$= \left[\begin{matrix}\text{Fermi}\\ (\text{Level}\\ \text{Effect}\end{matrix} + \begin{matrix}\text{Ion}\\ \text{Pairing}).Au^{TOT}(intrinsic)\\ \text{Effect}\end{matrix}\right]\chi \qquad G3$$

The only problem is knowing whether Au^- ions do occupy all of the excess vacancies - this is established experimentally by the 'push-out' experiments described in chapter 8 and is supported by the success of this theory (see table 6.3).

APPENDIX H

```
C00831  JOB REQUEST  18241 MJ05
C00831  JOB REQUEST  18241 MJ05

FORTRAN COMPILATION BY WATFOR IV 2.  DATE  15/06/77  TIME  12/10/56

0001                  LIST$LP)
0002                 LIBRARY(SUBROUTSRSP)
0003                  PROGRAM(TERRY)
0004                 INPUT=CRD
0005                 OUTPUT1=LP)
0006                 OUTPUT3=TAPE7(OUT2)
0007                  TRACE 1
0008                 COMPACT DATA
0009                  END



0010           C      FILE NAME.....DATPR31
0011           C      THIS VERSION ONLY CALCULATES FOR ONE SET OF DEGENERACIES.
0012           C
0013           C
0014                  MASTER TERRY
0015                  INTEGER X,Y,4
0016                  REAL K,NOBI,NOIL,NOHCE,NOUCH,ME,MH
0017                 DIMENSION YAXIS(39),XAXIS(39),YNAU(40),RES(2,40,8,12),
0018                1GA(3),GD(3),GAA(3),GGA(3),GDA(3),SDONOR(12),SACC(12),
0019                2GOL)(14,17),FINRES(14,17),12(14)
0020                 COMMON YAXIS,XAXIS,YNAU,I,J,L,RES,IJ,GOLD,FINRES,
0021                1MZ/BLOCK1/EG,E1,E2,E3,E4,E5,PHIA,PHID,PHINAU,PHIAU,PHIAAU
0022           C
0023           C      THE NEXT INSTRUCTIONS INITIALISE THE MAG-TAPE FOR THE GRAPH PLOTS
0024           C
0025                  CALL PLOTS(0,0,3)
0026                  CALL JNAME
0027                  CALL PLOT(25.,2,3,-3)
0028                  DO 300 I=1,2
0029                   K=3,NZC=5
0030           C      J AND L ARE COUNTERS USED TO LOAD THE ARRAYS OF RESULTS.
0031                   J=0
0032                   L=0
0033           C
0034           C      READ THE VALUES OF GOLD CONCN INTO THE ARRAYS
0035           C      UP TO 40 VALUES OF GOLD CONCN MAY BE INPUT, .
0036           C


0037                  READ(2,10) YNAU(I)
0038           10      FORMAT(E8.2)
0039           C
0040           C      IF ALL FORTY ELEMENTS OF THE ARRAY ARE NOT USED THEN THE LAST
0041           C      VALUE READ IN IS 1.11E11.
0042           C
0043                  IF(YNAU(I).EQ.1.11E 11)GOTO71
0044           70     CONTINUE
0045           C
0046           C      THE ENERGY VALUES FOR THE BAND-GAP,SHALLOW ACCEPTOR LEVEL, SHALLOW
0047           C      DONOR LEVEL, GOLD DONOR LEVEL, GOLD ACCEPTOR LEVEL AND GOLD-COUPLED
0048           C      ACCEPTOR LEVEL ARE READ IN,
0049           71     READ(2,11)EG,E1,E2,E3,E4,E5
0050                  CALL ETOPHI
0051           C
0052           C      E1 = SHALLOW ACCEPTOR ENERGY LEVEL RELATIVE TO THE VALENCE BAND,
0053           C      E2 = SHALLOW DONOR ENERGY LEVEL REL CONDUCTION BAND,
0054           C      E3 = GOLD DONOR LEVEL REL VAL BAND,
0055           C      E4 = GOLD ACCEPTOR LEVEL REL CON BAND,
0056           C      E5 = GOLD COUPLED SHALLOW ACCEPTOR ENERGY LEVEL REL VAL BAND,
0057           C
0058           11      FORMAT(6F0.0)
0059                   READ(2,12) XNI
0060           C
0061           C      XNI = SQRT N,P
0062           C
0063           12      FORMAT(E12.5)
0064           C
0065           C      T = ABS TEMP,
0066           C
0067                   READ(2,13) T
0068           13        FORMAT(F0.0)
0069           C
0070           C      GA = DEGEN SHALLOW ACC
0071           C      GD = DEGEN SHALLOW DONOR
0072           C      GAA = DEGEN GOLD ACC
0073           C      GDA = DEGEN GOLD DONOR
0074           C      GAA = DEGEN GOLD-COUP SHALLOW ACC
0075           C
0076                  IJ=1
0077                   READ(2,14) GA(IJ),GD(IJ),GAA(IJ),GDA(IJ),GGA(IJ)
0078           14      FORMAT(5F0.0)
0079           C
0080           C      READ IN FIRST VALUE OF SHALLOW DONOR CONCN AND SHALLOW ACC CONCN,
0081           C
0082           02      READ(2,15) YND,YNA
0083           15     FORMAT(2E12.5)
```

```
0084            C
0085            C       CHECK IF LAST VALUES REACHED,,,,,1,0E25 IS THE TERMINATOR;
0086            C
0087                    IF(YND,EQ,1,0E25,OR,YNA,EQ,1,0E25) GOTO 50
0088                    IF(Y(J)),91,9
0089            C
0090            C       COUNT NO,, VALUES OF YND (SALLOW DONOR CONCN)
0091            C
0092                    J=J+1
0093                    GOTO54
0094            C
0095            C       COUNT NO,, VALUES OF YNA,
0096            C
0097            91      L=L+1
0098            92      CONTINUE
0099            C
0100                    DO 33 I=1,40
0101                    IF(YNAU(I),EQ,1,11E11)GOTO500
0102                    YNB=(((YNAU(I))/(1,7E17))**3,0)*(4,5E15)
0103                    IF(YND))3,60,9
0104            C
0105            C       SETS THE LIMITS OF ITERATION IF THE MATERIAL IS N-TYPE
0106            C
0107                    AA=1,5
0108                    BB=-0,0
0109                    PHIF=0,25
0110                    GOTO 61
0111            C
0112            C       SETS ITERATION LIMITS FOR P-TYPE MATL,
0113            C
0114            60      AA=0,0
0115                    BB=0,0
0116                    PHIF=0,25
0117            61      CONTINUE
0118            C
0119            C
0120            C       CALCULATE THE NEGATIVE SIDE OF THE C,B,EQUN,
0121            C
0122                    EFI=-((K*T)*(ALOG((2,51/1,10)**0,75)))
0123            58      CBA=((EFI-PHIF)/(K*T))
0124                    CBA=XII*EXP(CBA)
0125                    CBC=(PHIF-PHIA)/(K*T)
0126                    CBD=GA(IJ)*EXP(CBC)
0127                    CBE=YIA/(1,0+CBD)
0128                    CBF=(PHIF-PHIAU)/(K*T)
0129                    CBG=(PHIF-PHINAU)/(K*T)
0130                    CBI=GNA(IJ)*EXP(CBF)


0131                    CBI=1,0+GDA(IJ)*EXP(CBG)
0132                    CBJ=(YNAU(I)*0,9)/((1,0+CBH*CBI)*1,0)
0133                    CBK=(PHIF-PHINAU)/(K*T)
0134                    CBL=GDA(IJ)*EXP(CBK)
0135                    CBM=YNG/(1,0+CBL)
0136                    A=CBB*CBE*CBJ*CBH
0137            C
0138            C       CALCULATE POSITIVE SIDE OF C,B, EQUN
0139            C
0140                    CBN=(PHIF-EFI)/(K*T)
0141                    CBO=AII*EXP(CBN)
0142                    CBP=(PHID-PHIF)/(K*T)
0143                    CBQ=(1,0/GD(IJ))*EXP(CBP)
0144                    CBR=Y.D/(1,0+CBQ)
0145                    CBS=(PHINAU-PHIF)/(K*T)
0146                    CBT=(PHIPAU-PHIF)/(K*T)
0147                    CBU=(1,0/GDA(IJ))*EXP(CBS)
0148                    CBV=1,0+(1,0/GAA(IJ))*EXP(CBT)
0149                    CBW=(YNAU(I)*0,9)/((1,0+CBU*CBV)*1,0)
0150                    B=CBO*CBR*CBW
0151            C
0152            C       ACCURACY OF SOLUTION CHECK POINT,
0153            C
0154                     CHARS(A-B)/ABS(A)
0155                    IF(C,LE,1,0E-5)GOTO55
0156            C
0157            C       IF THE C,B, EQUN IS NOT SOLVED TO REQUIRED ACCURACY THE
0158            C       POSITION OF THE FERMI LEVEL IS CHANGED TO A NEW VALUE WHICH
0159            C       DEPENDS ON THE RESULT OF THE PREVIOUS ITERATION,
0160            C
0161                    IF(B,GT,A)GOTO56
0162                    BB=PHIF
0163            57      PHIF=(BB+AA)/4
0164            C
0165            C       SOLVE WITH A NEW VALUE OF THE FERMI LEVEL,
0166            C
0167                     GOTO 58
0168            56      AA=PHIF
0169            59      PHIF=(AA+BB)/4
0170                    GOTO58
0171            55      CONTINUE
0172            C
0173            C       CALCULATE THE RESISTIVITY
0174            C
0175                    IF (YND)3,00,9
0176                    SUMOR(J)=YND
0177                    SCAT1=CBE*CBJ*CBM*CBR*CBW
```

```
0178                CAR1=C88
0179                CAR2=C89
0180                ML=1,47
0181                MH=0,38
0182          C
0183          C     CALC MOBILITY FOR ELECTRONS
0184          C
0185                MOBL=1,350E+3
0186                MOBI=XMOB(1E,>CAT1,CAR1)
0187                MOBCE=1,0/$(1,0/MOBL)+(1,0/MOBI))
0188          C
0189          C     CALC MOBILITY FOR HOLES
0190          C
0191                MOBL=480,0
0192                MOBI=XMOB(MH,>CAT1,CAR2)
0193                MOBCH=1,0/$(1,0/MOBL)+(1,0/MOBI))
0194          C
0195          C     CALCULATE THE RESISTIVITY FOR N-TYPE MATERIAL
0196          C
0197                RES(1,I,IJ,J)=1,0/((1,602E-19)*((CAR1*MOBCE)+(CAR2*MOBCH)))
0198                GOTO 61
0199          C
0200          C     NOW CALC RESISTIVITY IF THE MATERIAL IS P-TYPE
0201          C
0202          60    SACC(L)=YYA
0203                SCAT2=C8E*CRJ*C8H*CHR*C8W
0204                CAR3=C88
0205                CAR4=C89
0206                ME=0,27
0207                MH=1,38
0208          C
0209          C     MOBILITY FOR ELECTRONS
0210          C
0211                MOBL=1,350E+3
0212                MOBI=XMOB(1E,>CAT2,CAR3)
0213                MOBCE=1,0/$(1,0/MOBL)+(1,0/MOBI))
0214          C
0215          C     MOBILITY FOR HOLES
0216          C
0217                MOBL=481,0
0218                MOBI=XMOB(MH,>CAT2,CAR4)
0219                MOBCH=1,0/$(1,0/MOBL)+(1,0/MOBI))
0220          C
0221          C     CALCULATE THE RESISTIVITY IN P-TYPE MATERIAL.
0222          C
0223                RES(2,I,IJ,L)=1,0/((1,602E-19)*((CAR3*MOBCE)+(CAR4*MOBCH)))
0224          61     CONTINUE



0225          53    CONTINUE
0226          500   CONTINUE
0227                GOTO62
0228          C
0229          C
0230          C
0231          C     THE REST OF THE MASTER SEGMENT IS CONCERNED WITH PRINTING OUT THE
0232          C     THE RESULTS OF THE CALCULATIONS.
0233          C
0234          C
0235          C
0236          50    I=I-1
0237          63    CONTINUE
0238                WRITE(1,900)
0239          900   FORMAT(1H ,'PARAMETERS USED IN CALCULATIONS,')
0240                WRITE(1,901)T,XJ,E1,E2,E3,E4,E5
0241          901   FORMAT(1H0,'TEMPERATURE=',F9,1,'   DEGREES ABS',
0242               1//1X,'INTRINSIC CONCN=',1PE9,3,'   CARRIERS PER CC',
0243               2//1X,'        PHIA=',0PF9,4,'   EV',//1X,'        PHID=',F9,4,
0244               3'   EV',//1X,'      PHIAAU=',F9,4,'   EV',//1X,'      PHIPAU=',
0245               4F9,4,'   EV',//1X,'      PHIAAU=',F9,4,'   EV')
0246                IF(J,EQ,0)GOTO 531
0247          C     WRITE THE TITLES,
0248                WRITE(1,200)
0249          200   FORMAT(1H1,5X,' GA  ',5X,' GD  ',5X,' GAA ',5X,' GDA ',5X,' GGA ',
0250               15X,' N-DOPING',5X,'AU-DOPING',5X,'RESIS=',//,1X,57X,'LEVEL',9X,
0251               1'LEVEL',7X,'TIVITY',///)
0252                JJ=1
0253          201   IJ=1
0254                DO 203 IK=1,I
0255                WRITE(1,204)G=(IJ),GD(IJ),GAA(IJ),GDA(IJ),GGA(IJ),SDONOR(JJ),
0256               1YYA(IK),RES(1,IK,IJ,JJ)
0257          204   FORMAT(1X,5X,F6,3,4X,F6,3,4X,F6,3,4X,F6,3,4X,F6,3,4X,1PE9,3,
0258               15X,E9,3,5X,E9,3)
0259          203   CONTINUE
0260          202   CONTINUE
0261                IF(JJ,EQ,J)GOTO205
0262                JJ=JJ+1
0263                GOTO201
0264          205   CONTINUE
0265          531   IF(L,EQ,0)GOTO 532
0266          C     START THE ANSWERS FOR P-TYPE
0267                WRITE(1,211)
0268          211   FORMAT(1H1,5X,' GA  ',5X,' GD  ',5X,' GAA ',5X,' GDA ',
0269               15X,' GGA ',5X,' P-DOPING',5X,'AU-DOPING',5X,'RESIS=',//,1X,
0270               157X,'TIVITY',7X,'LEVEL',7X,'LEVEL',///)
0271                LL=1
```

```
0272          406   IJ=1
0273                DO 209 IL=1,I
0274                WRITE(1,204)GA(IJ),GB(IJ),GAA(IJ),GBA(IJ),GBA(IJ)/VACC(LL),
0275               1YMAT(IL),RES(4,IL,IJ,LL)
0276          208   CONTINUE
0277          407   CONTINUE
0278                IF(LL.EQ.L)GOTO210
0279                LL=LL+1
0280                GOTO206
0281          210   CONTINUE
0282          032   CONTINUE
0283                WRITE(1,401)
0284          401   FORMAT(/////////////////)
0285                DO 500 I1=1,14
0286                DO 501 I2=1,10
0287                READ(2,502)GOLD(M1,I2),FIGRES(M1,M2)
0288          502   FORMAT(E5.2,E10.2)
0289                IF(.OLD(M1,I2).EQ.0.00E00)M2(M1)=M2-1
0290                IF(GOLD(M1,I2).EQ.0.00E00)GOTO300
0291          501   CONTINUE
0292          500   CONTINUE
0293          C
0294          C     THE CALCOMP GRAPH PLOTTING ROUTINES ARE IN A SUBROUTINE CALLED BY THE
0295          C     NEXT STATEMENT,
0296          C
0297                CALL GRAPHPLOT
0298          600   CONTINUE
0299                CALL PLOT(0.,0.,999)
0300                STOP
0301                END
```

END OF SEGMENT, LENGTH 3807 NAME TERRY

```
0302                FUNCTION XMOB(E1,SCAT,CARI)
0303          C
0304          C     THIS FUNCTION SEGMENT CALCULATES THE MOBILITY,
0305          C
0306                A=((1.595E+20)*EM)/(CARI*1.0E+6)
0307                B=ALOG(1.0+A)
0308                C=1.+(A/(1.3+A))
0309                D=((E1**0.3)*(SCAT*1.0E+6))*C
0310                XMOB=(2.457E+23/D)*1.0E+4
0311                RETURN
0312                END
```

END OF SEGMENT, LENGTH 62, NAME XMOB

```
0313                SUBROUTINE GRAPHPLOT
0314          C
0315          C     THE GRAPHS ARE PLOTTED ONTO A MAG-TAPE,
0316          C
0317                DIMENSION XARR(17),YARR(17)
0318                COMMON YAXIS(39),XAXIS(39),YMAU(40),I,J,L,RES(2,40,8,12),IJ,
0319               1GOLD(14,17),FIGRES(14,17),M2(14)
0320          C     FOR THE XAXIS ARRAY
0321                M11=1
0322                DO 120 N=2,I
0323                XAXIS(M11)=YMAU(N)
0324                M11=M11+1
0325          120   CONTINUE
0326                CALL SCALG(XAXIS,16,,37,1)
0327          C     SCALE THE NUMBER RESULTS ARRAYS,
0328                DO 303 M1=1,14
0329                GOLD(M1,(M2(M1)+1))=XAXIS(38)
0330                GOLD(M1,(M2(M1)+2))=XAXIS(39)
0331                FIGRES(M1,(12(M1)+1))=1.00E-3
0332          303   FIGRES(M1,(12(M1)+2))=0.375
0333          C     N TYPE DATA IN GOLD(X,Y),X=1,5
0334          C     PTYPE DATA IN GOLD (X,Y),X=7,14
0335                CALL LGAXS(3,.0,,12HAU CONC A/CC,-12,10.,.0,,XAXIS(38),XAXIS(39))
0336          C     DRAW YAXIS
0337                CALL LGAXS(3,.0,,19HRESISTIVITY OHM CM,,19.24,,90.,
0338               11.00E-3,3.0,3.75)
0339                IF(J.EQ.0)GOTO 835
0340          C     N-TYPE GRAPHS,
0341                IJJ=1
0342                DO 124 III=1,J
0343                DO 123 NNN=2,I
0344          123   YAXIS(NNN-1)=RES(1,NNN,IJJ,III)
0345                YAXIS(38)=1.00E-3
0346                YAXIS(39)=0.375
0347                WRITE(1,836)
0348          836   FORMAT(1X,'THE EXECUTION OF LGLIN ABOUT TO BE CARRIED OUT??')
0349                CALL LGLIN(XAXIS,YAXIS,37,1,0,0,0)
0350          122   CONTINUE
0351          121   CONTINUE
0352          833   CONTINUE
0353                DO 304 M1=1,0
0354                DO 303 M2=1,M2(M1)+2
0355                XARR(M2)=GOLD(M1,M2)
```

```
0336            305   YARR(I2)=FINRES(N1,N2)
0337            304   CALL LGLIN(XARR,YARR,N2(N1),1,-1,N1,0)
0338            C     REASSIGN THE ORIGIN.
0339                  CALL PLOT(30.,0.,-3)
0340                  CALL LGAXS(0.,0.,12HAU CONC A/CC,-12,10.,0.,XAXIS(38),XAXIS(39))
0341                  CALL LGAXS(0.,0.,19HRESISTIVITY OHM CM,,19,24.,90.,
0342                 11,1.0E-3,0.,573)
0343                  IF(L.EQ.0)GOTO 634
0344                  IJJ=1
0345                  DO 125 III=1,4
0346                  DO 126 JJJ=2,1
0347            126   YAXIS(JJJ-1)=RES(2,JJJ,IJJ,III)
0348                  YAXIS(53)=1.0E-3
0349                  YAXIS(52)=0.573
0350                  WRITE(1,603)
0351                  CALL LGLIN(XAXIS,YAXIS,37,1,0,0,0)
0352            125   CONTINUE
0353            124   CONTINUE
0354            634   CONTINUE
0355                  DO 306 N1=7,14
0356                  DO 307 I2=1,N2(N1)+2
0357                  XARR(I2)=GOL(N1,N2)
0358            307   YARR(I2)=FINRES(N1,N2)
0359            306   CALL LGLIN(XARR,YARR,N2(N1),1,-1,N1,0)
0360                  CALL PLOT(30.,0.,-3)
0361                  RETURN
0362                  END
```

END OF SEGMENT, LENGTH 408, NAME GRAPHPLOT

```
0363                  SUBROUTINE ETOPHI
0364            C
0365            C     THIS ROUTINE CONVERTS THE ENERGY LEVELS OF THE VARIOUS DOPANTS
0366            C     FROM VALUES RELATIVE TO THE BAND EDGES TO VALUES RELATIVE TO THE
0367            C     CENTRE OF THE BAND-GAP.
0368            C
0369                  COMMON/BLOCK1/EG,E1,E2,E3,E4,E5,PHIA,PHID,PHINAU,PHIPAU,PHIAAU
0370                  PHIA=(EG/2.0)-E1
0371                  PHID=(-EG/2.0)+E2
0372                  PHINAU=(EG/2.0)-E3
0373                  PHIPAU=(-EG/2.0)+E4
0374                  PHIAAU=(EG/2.0)-E5
0375                  RETURN
0376                  END
```

REFERENCES

ADAMIC J.W., McNAMARA J.E., (1964). Paper presented at Electrochem.Soc. Meeting. New York, Sept., 1963.

ADAMIC J.W., McNAMARA J.E., (1964). Paper presented at Electrochem. Soc. Meeting, Washington D.C., Oct., 1964.

ANDERSON C.L., BARON R., (1976) Rev. Sci. Instrum. 47, 1366.

BADALOV A.Z. (1970) Sov.Phys.Semicond. 3, 1435.

BARBER H.D. (1967) Solid State Electron., 10, 1039.

BEMSKI, G. (1958). Phys.Rev., 111, 1515.

BLAKEMORE J.S., (1962). "Semiconductor Statistics", Permamon Press, 1962.

BLUDAU W., ONTON A., HEINKE E. (1974) J. Appl.Phys., 45, 1846.

BOLTAKS B.I., KULIKOV G.S., MALKOVICH R. Sh., (1960). Sov.Phys. - Solid State, 2, 167.

BOLTAKS B.I., KULIKOV G.S., MALKOVICH R.Sh., (1961) Sov.Phys - Solid State, 2, 2134.

BONCH-BRUYEVICH V.L. (1966). "The Electronic Theory of Heavily Doped Silicon", Elsevier, New York, 1966.

BRAUN S., GRIMMEISS H.G. (1973). J.Appl.Phys., 44, 2789.

BRAUN S., GRIMMEISS H.G. (1974). J.Appl.Phys., 45, 2658.

BROOKS H. (1955) Phil. Mag., 46, 831.

BROTHERTON S.D., ROGERS T.L. (1972) Solid State Electron., 15, 853.

BROTHERTON S.D. (1976) Solid State Electron., 19, 341.

BROWN M., JONES C.L., WILLOUGHBY A.F.W. (1975) Solid State Electron., 18, 763.

BROWN M. (1976) Ph.D. Thesis, University of Southampton.

BRUCKNER B. (1971) Phys.Stat.Sol., (a), 4, 685.

BUCK T.M., WHEATLEY G.H. (1972) Surf.Sci. 33, 35.

BUCK T.M., PICKAR K.A., POATE J.M., HSIEH C.M. (1972). Appl.Phys.Lett.,21,485.

BUCK T.M., POATE J.M., PICKAR K.A., HSIEH C.M. (1973) Surf.Sci.35, 362.

BULLIS W.M. (1966) Solid State Electron., 9, 143.

BULLIS W.M., STRIETER F.J. (1968) J.Appl.Phys.,39, 314.

BURGER R.M., DONOVAN R.P. (1967) Eds. "Fundamentals of Silicon Integrated Device Technology", Vol.I, 1967, Prentice Hall Inc.,New Jersey.

CAGNINA S.F. (1969) J.Electrochem. Soc., 116, 498.

CARLSON R.O. (1956) Phys.Rev., 104, 937.

CHOO S.C., LEONG M.S., KUAN K.L. (1976) Solid State Electron., 19, 561.

CHOU S.L., GIBBONS J.F. (1975) J.Appl.Phys., 46, 1197.

COLLINS C.B., CARLSON R.O., GALLAGHER C. (1957) Phys.Rev.105, 1168.

CONWELL E.H., WEISSKOPF V.F. (1950), Phys.Rev., 77, 388.

CULLIS A.G., (1974) International Conference on Electron Microscopy.

DASH W.C. (1960) J.Appl.Phys., 31, 2275.

DAVIS W.D. (1959) Phys.Rev., 114, 1006.

DECLERCK G.J., De MEYER K., JANSSENS E., LAES E., Van Der SPIEGEL J. (1976), I.E.E.E. Trans. ED. 23, 297.

DEKKER A.J. (1952) "Solid State Physics", Macmillan and Co. Ltd., 1952.

De KOCK A.J.R. (1971) J.Electrochem. Soc., 118, 1851.

De KOCK A.J.R., ROKSNOER P.J., BOONEN P.G.T., (1975) J.Crystal Growth,30, 279.

De KOCK A.J.R., FERRIS S.D., KIMERLING L.C., LEAMY H.J. (1977), J.Appl.Phys., 48, 301.

DORWARD R.C., KIRKALDY J.S. (1968) Trans.Met.Soc.A.I.M.E., 242, 2055.

DORWARD R.C., KIRKALDY J.S. (1969) J.Electrochem.Soc., 116, 1284.

DRESSELHAUS G. (1956) J.Phys.Chem.Solids, 1, 15.

DUNLAP W.C., BOHM H.V., MAHON H.P. (1954) Phys.Rev., 96, 822.

ENGSTROM O., GRIMMEISS H.G. (1974) Appl.Phys.Lett.,25, 413.

ENGSTROM O., GRIMMEISS H.G. (1975) J.Appl.Phys 46, 831.

FAIR R.B. (1977) "Semiconductor Silicon 1977", Eds.HUFF H.R.,SIRTL E. Publ. Electrochem.Soc.,Princeton,New Jersey, 1977.

FAIRFIELD J.M., GOKHALE B.V. (1965) Solid State Electron. 8, 685.

FAIRFIELD J.M., SCHWUTTKE G.H. (1966) J.Appl.Phys., 37, 1536.

FISTUL V.I. (1969) "Heavily Doped Semiconductors", Plenum Press,New York,1969.

FONASH S. (1974) In "Proc.Symposium on Spreading Resistance Measurements", N.B.S. Special Publication. 400-10, 1974.

FRANK F.C., TURNBULL C.D. (1956) Phys.Rev., 104, 617.

FRENKEL J. (1938) Phys.Rev., 54, 647.

GROVE A.S. (1967) "Physics and Technology of Semiconductor Devices", J.Wiley and Sons Inc., New York, 1967.

GUTAI L., VICSEK T. (1975) Solid State Electron. 18, 93.

HALDANE F.D.M., ANDERSON P.W. (1976) Phys.Rev. B. 13, 2553.

HALL R.N. (1952) Phys. Rev. 87, 387.

HALL R.N., RACETTE J.H. (1964) J.Appl. Phys. 35, 379.

HAYNES J.R., LAX M., FLOOD W.F. (1959) J.Phys.Chem.Sol., 8, 392.

HEIMAN F.P. (1967) I.E.E.E. ED., 14 781.

HENDRICKSON T.E. (1975) J.Electrochem. Soc., 122, 1539.

HERMAN J.M., SAH C.T. (1972) Phys.Stat.Sol.(a), 14, 405.

HERRING C. (1955) Bell.Syst.Tech.J., 34,237.

HERRMANN H., LIERZER H., SIRTL E. (1975) Festkorperprobleme, XV, 279.

HOLM R. (1958) Electrical Contacts Handbook. 3rd Ed. Springer, Berlin, 1958.

HU S.M., YEH. T.H. (1969) J.Appl.Phys., 40, 4615.

HU S.M. (1972) Solid State Electron. 15, 809.

HU S.M. (1974) J.Appl.Phys., 45, 1567.

HU S.M. (1975) Appl.Phys.Lett., 27, 165.

HU S.M. (1977) J.Vac.Sci.Technol., 14, 17.

HUNTLEY F.A., WILLOUGHBY A.F.W. (1970) Solid State Electron., 13, 1231.

HUNTLEY F.A., WILLOUGHBY A.F.W. (1971) Solid State Electron., 14, 641.

HUNTLEY F.A. (1972) Ph.D. Thesis, University of Southampton.

HUNTLEY F.A., WILLOUGHBY A.F.W. (1973i) J.Electrochem. Soc., 120, 414.

HUNTLEY F.A., WILLOUGHBY A.F.W. (1973ii) Phil.Mag.28, 1319.

HWANG C.J., BREWS J.R. (1971) J.Phys. Chem.Solids, 32, 837.

IRVIN J.C. (1962) Bell Syst.Tech.J. 41, 387.

JAIN R.K., OVERSTRAETEN R.J. Van. (1974) I.E.E.E. ED.21,155.

JONES C.L. (1974) Ph.D. Thesis, University of Southampton.

JONES C.L., WILLOUGHBY A.F.W. (1975) J.Electrochem.Soc., 122, 1531.

JONES C.L., WILLOUGHBY A.F.W. (1976) J.Electrochem.Soc., 123, 1531.

JOSHI M.L., DASH S. (1966) J.Appl.Phys., 37, 2453.

KAMM J.D., MULLER R. (1977) Solid State Electron., 20, 105.

KANE E.O. (1956) J.Phys.Chem.Sol. 1, 82.

KANE E.O. (1963) Phys Rev., 131, 79.

KASSING R., LENZ H. (1974 i). Phys.Stat.Sol. (a) 25, 131.

KASSING R., LENZ H. (1974 ii) Phys.Stat.Sol. (a) 26, 155.

KENDALL D.L., DE VRIES D.B. (1969) "Semiconductor Silicon 1969", Eds. HABERECHT R.R., KERN E.L. Publ.by Electrochem.Soc.Inc.,Princeton, New Jersey, 1969.

KERN W., PUOTINEN P.A. (1970) R.C.A. Rev., 3, 187.

KLEPPINGER D.D., LINDHOLM F.A. (1971) Solid State Electron., 14, 407.

KUDOH D., UDA K., IKUSHIMA Y., KAMESHIDA M. (1976) J.Electrochem,Soc., 123, 1751.

LAMBERT J.L., REESE M. (1968) Solid State Electron. 11, 1055.

LARK-HOROVITZ K. (1959) "Methods of Experimental Physics", vol. 6, Part B; Solid State Physics, Academic Press, 1959.

LAX M. (1960) Phys.Rev. 119, 1502.

LEE D.B., WILLOUGHBY A.F.W. (1972) J.Appl.Phys. 43, 245.

LEE D.B., (1974) Ph.D. Thesis, University of Southampton , and Philips Research Suppl. No. 5, 1974.

LESK I.A. (1973) "Semiconductor Silicon 1973" Eds. HUFF H.R., BURGESS R.R. Publ. by Electrochem. Soc.Inc., Princeton, New Jersey, 1973.

LIDIARD A.B. (1954) Phys.Rev., 94, 29.

LLEWELLYN-JONES (1957) "Physics of Electrical Contacts", Clarendon Press, 1957.

LOGAN R.M. (1976) Proc.3rd International Conference on The Technology and Applications of Charge Coupled Devices", University of Edinburgh, 1976.

LUDWIG G.W., WATTERS R.L. (1956) Phys.Rev., 101, 1699.

MACFARLANE G.G., McLEAN T.P., QUARRINGTON J.E., ROBERTS V. (1958) Phys.Rev.,111, 1245.

MALKOVICH R.Sh., (1968) Sov.Phys. - Solid State, 9, 1676.

MARTIN J., HAAS E., RAITHEL K. (1966) Solid State Electron. 9, 83.

MAYER A., SHWARTZMAN S. (1974) In "Proc.Symposium on Spreading Resistance Measurements" N.B.S. Special Publ. 400-10, 1974.

MAZUR R.G., DICKEY D.H. (1963) Paper presented at Spring Meeting Electrochem. Soc., Pittsburgh, 1963.

MAZUR R.G., DICKEY D.H. (1966) J.Electrochem. Soc., 113, 255.

McLEAN T.P. (1960) Progress in Semiconductors 5, 53.

McKELVEY J.P. (1966) "Solid State and semiconductor Physics", Harper Row, New York and London, 1966.

MEEK R.L. SEIDEL T.E., CULLIS A.G. (1975) J.Electrochem. Soc., 122, 786.

MEEK R.L., SEIDEL T.E. (1975) J.Phys.Chem.Solids, 36, 731.

MILLER M.D. (1976) I.E.E.E. ED. 23, 1279.

MORGAN T.W. (1965) Phys.Rev., 139, 343.

MURARKA S.P. (1976) J.Electrochem. Soc., 123, 765.

NAKAMURA M., KATO T., OI N. (1968) Jap.J.Appl.Phys., 7, 512.

NEWMAN R. (1954) Phys.Rev., 94, 1530.

NICHOLS K.G., VERNON E.V. (1966) "Transistor Physics", Chapman and Hall Ltd., London, 1966.

O'SHAUGHNESSEY T.A., BARBER H.D., THOMPSON D.A., HEASELL E.L. (1974) J.Electrochem. Soc., 121, 1350.

O'SHAUGHNESSEY T.A., BARBER H.D., HEASELL E.L. (1976) J.Electrochem.Soc., 123, 1560.

PALS J.A. (1974) Solid State Electron. 17, 1139.

PARKER W.C., FORBES L. (1975) I.E.E.E. ED. 22, 916.

PARRILLO L.C., JOHNSON W.C. (1972) Appl.Phys.Lett.,20, 104.

PEART R.F., NEWMAN R.C. (1972) In "Radiation Damage and Defects in Semiconductors' Reading Conf.Proc. 1972, Inst.Phys.Lond,

PETROFF P.M., De KOCK A.J.R. (1975) J.Crystal Growth, 30, 117.

PHILLIPS J.C. (1973) "Bonds and Bands in Semiconductors", Academic Press, New York and London, 1973.

PUTLEY E.H., MITCHELL E.W. (1958) Proc.Phys.Soc.Lond., A72, 193.

REISS H., FULLER C.S., MORIN F.J. (1956) Bell Syst.Tech.J., 35, 535.

ROGERS T.L. Private Communication (see Brotherton et al. 1972).

RUIZ H.J., VOLTMER F.W. (1974) In "Proc Symposium on Spreading Resistance Measurements", N.B.S. Special Publ. 400-10., 1974.

RUNYAN W.R. (1965) "Silicon Semiconductor Technology", McGraw-Hill, New York, 1965.

RUNYAN W.R. (1975). "Semiconductor Measurement and Instrumentation", McGraw-Hill, New York, 1975.

SAH C.T., FORBES L., ROSIER L.T., TASCH A.F., TOLE A.B., (1969) Appl.Phys.Lett., 15, 145.

SAH C.T., FORBES L., ROSIER L.T., TASCH A.F. (1970) Solid State Electron., 13, 759.

SAH C.T. (1976) Solid State Electron., 19, 975.

SANDERS I.R., DOBSON E.S. (1969) Phil. Mag.20, 881.

SCHUMANN P.A., GARDNER E.C. (1969) J.Electrochem.Soc., 116, 87.

SCHUMANN P.A., GARDNER E.C. (1969) Solid State Electron., 12, 371.

SCHWETTMAN F.N., KENDALL D.L. (1972) Appl.Phys.Lett., 21, 2.

SECCO D'ARAGONA F. (1972) J.Electrochem.Soc., 119, 948.

SEIDEL T.E., MEEK R.L. (1973) In "Ion Implantation in Semiconductors", Ed.CROWDER B.L. Plenum Press, New York, 1973.

SEIDEL T.E., MEEK R.L., CULLIS A.G. (1965) J.Appl.Phys., 46, 600.

SENECHAL R.R., BASINSKI J. (1968) J.Appl. Phys. 39, 3723.

SHAKLEE K.L., NAHORY R.E. (1970) Phys.Rev.Lett., 24, 942.

SHOCKLEY W., READ W.T. (1952) Phys.Rev., 87, 835.

SHOCKLEY W., LAST J.T. (1957) Phys.Rev., 7, 392.

SHOCKLEY W., MOLL J.L. (1960) Phys.Rev., 119, 1480.

SHTIVEL'MAN K.Ya. (1974) Sov.Phys.-Semicond.8, 528.

SIGMON T.W., CSEPREGI L., MAYER J.W. (1976) J.Electrochem.Soc., 123, 1116.

SIRTL E., ADLER A. (1961) Z.Metall.K., 52, 529.

SMITH R.A. (1964) "Semiconductors" Camb.Univ.Press, 1964.

SMITS F.M. (1958) Bell Syst.Tech.J. 37, 711.

SPROKEL G.J. (1965) J.Electrochem.Soc., 112, 807.

SPROKEL G.J., FAIRFIELD J.M. (1965) J.Electrochem.Soc., 112, 200.

SRINIVASAN G. (1969) Phys.Rev., 178, 1244.

STRUTHERS J.D. (1956) J.Appl.Phys.27, 1560.

STRUTHERS J.D. (1957) J.Appl.Phys. 28, 516.

STURGE M.D. (1958) Proc.Phys.Soc., 73, 297.

SZE S.M., IRVIN J.C. (1968) Solid State Electron.11, 599.

SZE S.M. (1969) "Physics of Semiconductor Devices", J.Wiley and Sons Inc., New York, 1969.

TAFT E.A., HORN F.H. (1954) Phys.Rev., 93, 64.

TANNENBAUM E. (1961) Solid State Electron., 2, 123.

TASCH A.F., SAH C.T. (1970) Phys.Rev., B, 1, 800.

TEITLER S., WALLIS R.F. (1960) J.Phys.Chem.Sol.16, 71.

THURBER W.R., LEWIS D.C., BULLIS M. (1973) N.B.S. Report No. A.F.C.R.L.-T.R.-73 - 0107.

TRUMBORE F.A. (1960) Bell Syst.Tech.J., 39, 205.

TWEET A.G. (1958) Phys.Rev., 111, 67.

UNTER T.F. (1977) "Studies of pre and post gate oxidation gettering in MOS structures" to be published.

UNTER T.F., ROBERTS P.C.T., LAMB D.R. (1977) Electron.Lett., 13, 94.

VALDES L. (1954) Proc.I.R.E., 42, 420.

VAN VECHTEN J.A., PHILLIPS J.C. (1970) Phys.Rev.,B, 2, 2160.

VAN VLACK L. (1970) "Material Science for Engineers", Reading, Massachusetts, (Addison-Wesley) 1970.

VARSHNI Y.P. (1967) Physica, 34, 149.

VIEWEG-GUTBERLET F.G. (1973) In "Semiconductor Silicon 1973", Eds. HUFF H.R., BURGESS R.R. Publ. by Electrochem.Soc.Inc., Princeton, New Jersey, 1973.

WATKINS G.D., CORBETT J.W. (1968) Phys.Rev., 173, 134.

WILEY J.D. (1971) J.Phys.Chem.Sol., 32, 2053.

WILCOX W.R., LA CHAPELLE T.J., FORBES D.H., (1964), J.Electrochem.Soc . 111, 1377.

WOLF H.F. (1969) "Silicon Semiconductor Data", Pergamon Press, New York, 1969.

WOLLEY E.D., STICKLER R. (1967) J.Electrochem.Soc., 114, 1287.

WONG D.C., PENCHINA C.M. (1975) Phys.Rev.B, 12, 5840.

YEH T.H., KHOKHANI K.H. (1969) J.Electrochem.Soc., 116, 1461.

YEOW Y.T. (1974) Ph.D. Thesis, University of Southampton.

YOSHIDA M. (1970) Jap.J.Appl.Phys.8, 1211.

YOSHIDA M., SAITO K. (1970) Jap.J.Appl.Phys. 9, 1217.

YOSHIDA M. (1973) Jap.J.Appl.Phys., 12, 1956.

YOSHIDA M., ARAI E., NAKAMURA H., TERUNUMA Y., (1974) J.Appl.Phys., 45, 1498.

ZERBST M. (1966) Z. Agnew Phys. 22, 30.