JOURNAL OF MATERIALS SCIENCE 3 (1968) 110–119

The Controlling Factors in Semiconductor Large Area Alloying Technology

F. M. ROBERTS, E. L. G. WILKINSON
Solid State Physics Group, Nelson Research Laboratories, English Electric Co Ltd, Beaconhill, Stafford, UK

Received 9 October 1967

Alloying is widely used in semiconductor device manufacture to form both ohmic contacts to, and p-n junctions in, the semiconductor body. Most of the literature on alloying deals with devices in which the alloyed area is small, while the present work refers to junction diameters of 19 to 25 mm.

In the alloying process investigated, a metal foil preform is held in contact with the surface of a silicon disc during a thermal cycle, while arrangements are made to measure the temperature gradient across the slice and relate this to the metallography of the junction. It is shown that the initial formation and spreading of the liquid phase in a uniform manner is dependent on the achievement of a number of specific conditions. In particular, the temperature of the liquid should be higher than that of the solid during both heating and cooling stages. Jigs must therefore be designed as much for their thermal properties as for component location.

Assemblies containing two metal/silicon alloy systems that have experienced the same thermal cycle, but opposite thermal gradient, illustrate how re-growth layer structure is affected by such gradients, and enable the effects of different dopants and amounts of dopant to be compared.

Measured depths of alloying into silicon agree with those calculated from phase diagrams, except where support plate material is involved in the process.

1. Introduction

1.1. General

Alloying is widely used in semiconductor device manufacture to form both ohmic contacts to, and p-n junctions in, the semiconductor body. Much of the literature on alloying deals with devices or experimental test pieces in which the alloyed area is small. Typically, spheres of metal or alloy 250 to 300 μm in diameter have been employed [1-3], giving alloyed areas up to 0.75 mm in diameter, while in the experiments to be described the alloyed area diameters are 19 to 25 mm.

Precise control of the depth of alloying, of the planarity of the base of the semiconductor re-growth region and of its parallelism with already existing junctions, is necessary, to produce in the device the required electrical characteristics and to maintain their radial uniformity. Non-uniformities, small in relation to the full thickness of the semiconductor body, may amount to an appreciable proportion of the thickness of diffused or epitaxially grown layers, and affect device operating parameters [4, 5] to a correspondingly greater degree. The depths of such layers are limited by technological feasibility to magnitudes comparable with the thicknesses of alloying material metallurgically necessary. Control must therefore be especially rigorous when alloying into such relatively thin regions. The occurrence of irregularities is likely to grow with increasing area, so that large area device processing is even more critical, and the comparatively great thermal masses involved place closer limits on the technology practicable.

1.2. The Alloying Process

In the alloying process the metal or alloy, often a foil disc or other preform, is held in contact with the surface of the semiconductor body (frequently a silicon disc) during a thermal cycle. During heating, the metal and silicon interdiffuse until the eutectic composition is reached. (Alternatively a metal/silicon eutectic alloy foil is used as the starting material.) Melting, wetting, and further dissolution follow as the temperature rises. During cooling, silicon-rich alloy rejected from the melt re-grows on the undissolved portion of the original crystal, and the remainder of the alloy finally solidifies above the re-growth layer. The structure that results in practice will be shown to depend significantly on the thermal flow conditions existing during the firing process; in particular, on the temperature gradients across the metal/silicon interfaces. It is further shown that these gradients can be controlled by appropriate jig design.

2. Experimental Procedure

2.1. Semiconductor Surface Preparation and Furnace Atmospheres

The (111) orientation is chosen for the alloying faces of silicon discs because the conditions for planar alloying on it are less stringent than those for other orientations [3, 6-8]. Although it is known [9] that special techniques of surface preparation promote wetting of silicon by metals, they were not used because the primary interest in these experiments was in the effect of thermal factors on alloying to conventionally-prepared surfaces. However, since oxide films, inimical to wetting [10], are quickly re-grown on silicon surfaces after etching, an oxide-removing etch was used immediately before alloying.

Wetting is also affected by furnace atmosphere [1, 10]. Alloying has been carried out in air, inert and reducing atmospheres, and in vacuum [11]. Since vacuum firing was used exclusively in the present experiments, the effects of furnace atmosphere are not discussed further.

2.2. Experimental Assemblies

A typical assembly (fig. 1a) comprised a gold-alloy foil annulus and/or disc, a 25 mm diameter silicon disc, either not previously specially processed, or with internal diffused structure and/or external epitaxial layer, an aluminium/silicon eutectic alloy foil disc, and a molybdenum disc. The components were prepared by the usual chemical techniques.

Assembled on the jig base, the components were temporarily held together concentrically by the surface tension effects of n-butyl or isopropyl alcohol while 5 μm graphite powder was compressed over them. Alternatively, the powder was omitted and the components were surmounted by a graphite or alumina disc.

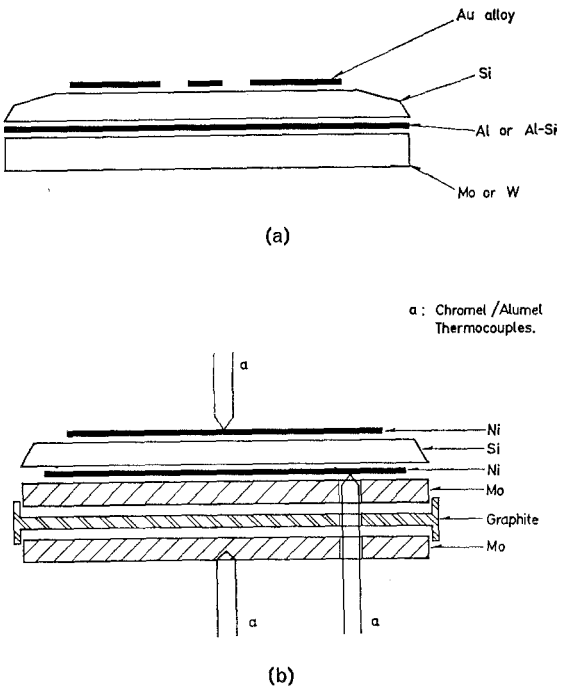


Figure 1 (a) Assembly of specimen components. (b) Thermal geometry calibrations assembly.

2.3. Temperature Measurement

The assembly shown in fig. 1b was used to measure the temperature gradient across a silicon disc. Nickel foils, to which thermocouples were brazed, simulated the alloy foils in contact with each side of the silicon disc. The relationship between the nickel foil temperatures and that of the lower molybdenum disc, to which a thermocouple was attached also, was established, and the particular jig being investigated was thus thermally characterised. The signal from the thermocouple in the lower molybdenum disc was used to control the furnace cycle during both calibration and actual firings. Temperatures were recorded on multi-point instruments.

Since the nickel and silicon in the calibration

assembly do not melt at the temperatures reached in the alloying cycles used, and nickel has thermal parameters different from those of the alloys used in actual firings, the temperature distributions and thermal flow patterns recorded during calibration are obviously not identical with those in a real assembly. However, that the actual situation was simulated with sufficient accuracy to suit the main requirements of the experiments is borne out by metallographic examination of the fused assemblies, and by comparison of actual and calculated depths of alloying.

2.4. Furnace Cycles

Assemblies were fired in RF-heated vertical vacuum furnaces. Heating rates varied from 25 to 250° C/min in powder jigs, and up to 350° C/min in solid component arrangements. Cooling was natural, the initial rate being about 30° C/min. In the single and two-stage firings used, maximum temperatures recorded by the control thermocouples ranged from 450 to 883° C.

2.5. Jig and Thermal Process Development

The temperature gradient across the components assembled in a simple jig (fig. 2) is not necessarily always in the same direction as the external

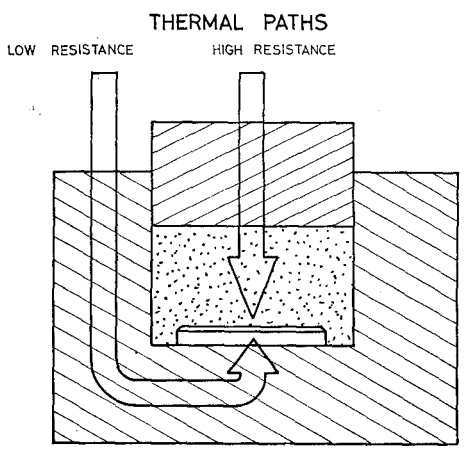


Figure 2 Simple jig.

temperature gradient in which the jig may be placed, but can be modified unfavourably by conduction through the jig body. In certain jig geometries (fig. 3) a gradient in one direction during fast heating could become inverted at slower rates, during hold periods, or while

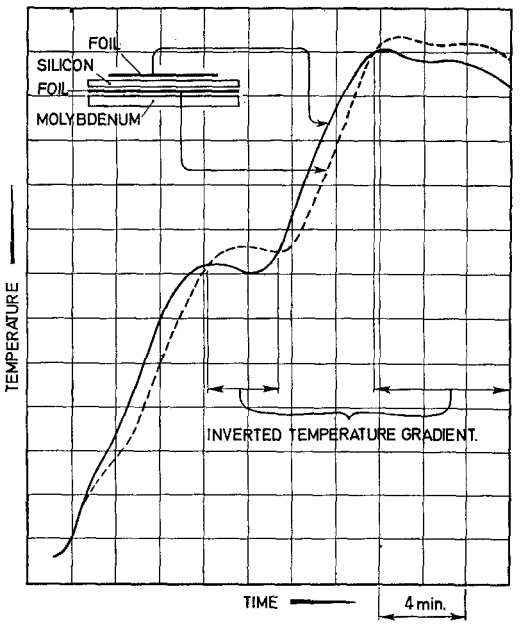


Figure 3 Temperature inversion during thermal cycle.

cooling. These effects were eliminated by developing jigs (fig. 4) in which radiant heat from the graphite susceptor was transferred via a graphite cylinder to the top surface of the device component assembly. Incorporated in this design was a thermal insulator to reduce heat flow through the jig body to the base of the component assembly. Thus throughout the thermal cycle there was maintained through the device a positive temperature gradient, i.e. the melt on the upper surface of the silicon was hotter than the solid silicon below it.

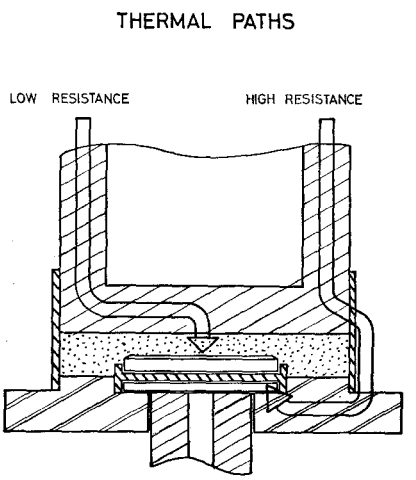


Figure 4 Thermal gradient jig.

In single-stage firing the maximum temperature used was intermediate between those optimum for the gold/silicon and the silicon/aluminium/molybdenum systems. In the two stage process the Mo/Al/Si joint was formed first at a higher temperature than that of the second stage, when the temperature was adjusted to suit the Au/Si system.

3. Experimental Results

3.1. Wetting Processes

The whole area of silicon under the foil preform must be wet virtually simultaneously if dissolution of the silicon is to commence uniformly. Wetting of a semiconductor body by an alloying metal is initiated as the temperature rises in areas of greatest contact pressure between the two adjacent faces by interdiffusion until the eutectic composition is reached, when a melt is formed. Fig. 5a shows areas of the (111) surface of a silicon disc in the initial stages of wetting by 1% antimony/gold foil. Contact between metal and silicon was poor. To obtain this micrograph the thermal cycle was interrupted immediately after wetting commenced.

In fig. 5b alloying is deepest in an arc under the periphery of a gold-alloy preform. In this example the foil disc used was allowed to retain the burred edge produced by the operation of punching it out from strip; no flattening treatment whatsoever was applied to it. The assembly was jigged with the upturned edge of the foil facing the silicon. When pressure was applied to the jig, metal/silicon contact was most intimate around the foil edge. Alloying therefore commenced, and subsequent penetration was deepest, in this area. The remainder of the alloyed area has been lapped off to reveal the deep peripheral penetration.

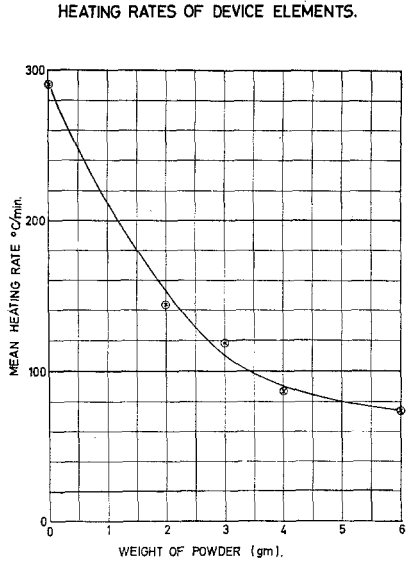
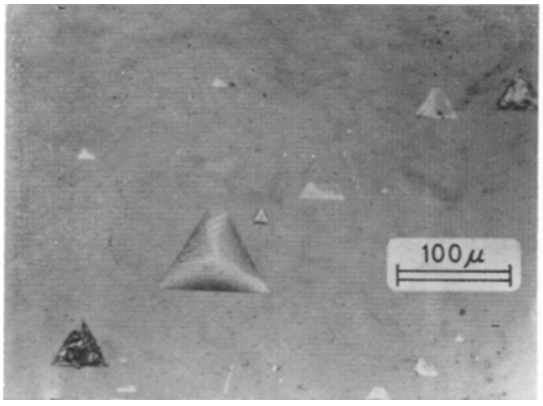


Figure 6 Effect of powder quantity on heating rate.

For greatest uniformity of contact pressure between component faces, hydrostatic pressure applied to the foil has obvious advantages. An approximation to this situation can be made by applying pressure via an inert powder, which also locates the components in position on the silicon. However, the powder method has disadvantages: while increase of pressure on the components promotes wetting, it may, by causing extrusion of molten material during the thermal cycle, adversely affect the metallurgy of joints.



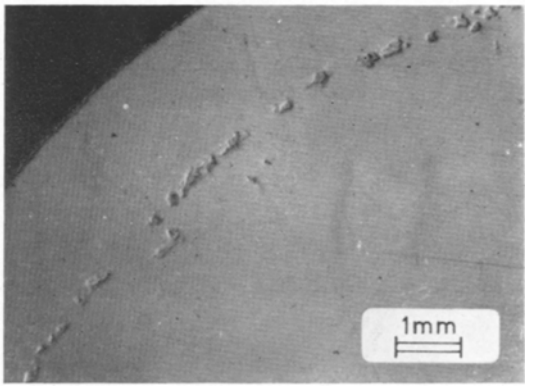


Figure 5 (a) Initiation of wetting of silicon by gold-alloy. (b) Deep penetration into silicon at edge of alloy foil disc.

Again, the low thermal conductivity of powder limits the heating rates attainable in a given jig arrangement and can, as shown in fig. 3, contribute towards negative temperature gradients. Fig. 6 shows heating rates, as measured by the control thermocouple, of different powder quantities in otherwise similar jig arrangements. For some device configurations jigging without powder may be preferable.

The spreading of wet areas across the silicon surface is affected by the solubility of silicon in the liquid metal [12]. Too low a solubility inhibits wetting; too high promotes local erosion rather than lateral spreading. Silicon is more soluble in gold than in aluminium in the temperature range suitable for alloying; e.g. at 700° C the respective solubilities are ~32% and ~20%. Hence any single-stage firing of a structure containing both Au/Si and Al/Si alloys must be at a compromise temperature that is optimum for neither system; also, conditions to promote spreading of wet areas and prevent local erosion have to be more stringently observed for gold/silicon than for the more tractable aluminium/silicon system.

Metallography indicates that the solubility required to produce the most effective wetting probably lies between ~25% and ~32% silicon in metal, corresponding to temperature ranges of about 750 to 850° C for aluminium, and 550 to 700° C for gold. Two-stage firing of devices containing both Au/Si and Al/Si systems is therefore advantageous.

The surface tension of most liquids decreases almost linearly as temperature rises [1]. Wetting is therefore encouraged by increasing the maximum temperature reached during the thermal cycle. In practice high maximum temperatures may result from fast heating rates (the effects of which are discussed below) associated with jigs of large thermal mass. The effects of higher temperatures on wetting are not all beneficial: solubility and its consequence, a greater tendency to local penetration, also increase with temperature. These factors must therefore be balanced by the choice of appropriate thermal cycles.

3.2. Dissolution Processes

3.2.1. Temperature Gradient and Heating Rate Effects

Thermal conductivity across the alloy/silicon interface in wet areas is greater than elsewhere [6]. Whatever the direction of thermal flow during heating, the solid adjacent to the melts becomes hotter and dissolves next. If the heating rate is too slow, the silicon surface unsuitably prepared, the lateral temperature distribution inhomogeneous, or the temperature gradient normal to the surface negative (fig. 7a), (i.e. the

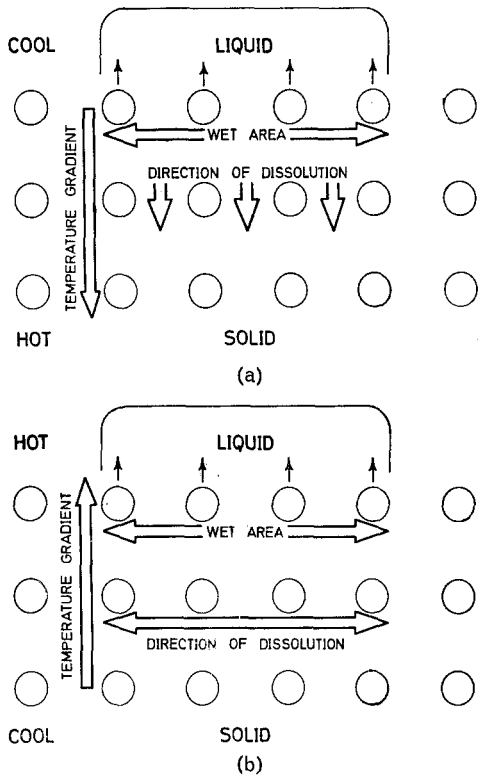


Figure 7 (a) Negative temperature gradient. (b) Positive temperature gradient.

silicon below the molten areas hotter than the liquid above them), dissolution proceeds locally into the solid, since increasingly energetic, more readily dissociated atoms lie in this direction. Local erosion from initially wet areas (fig. 8) is therefore encouraged and an irregular alloy front produced.

These micrographs show gold-alloy/silicon/gold-alloy sandwiches fired in a temperature gradient, the upper surface being hotter than the lower. Non-uniform wetting and irregular penetration has occurred in the case of the negative temperature gradient across the lower surface.

If the temperature gradient is positive (fig. 7b) dissolution proceeds laterally across the solid surface region, since now the most readily

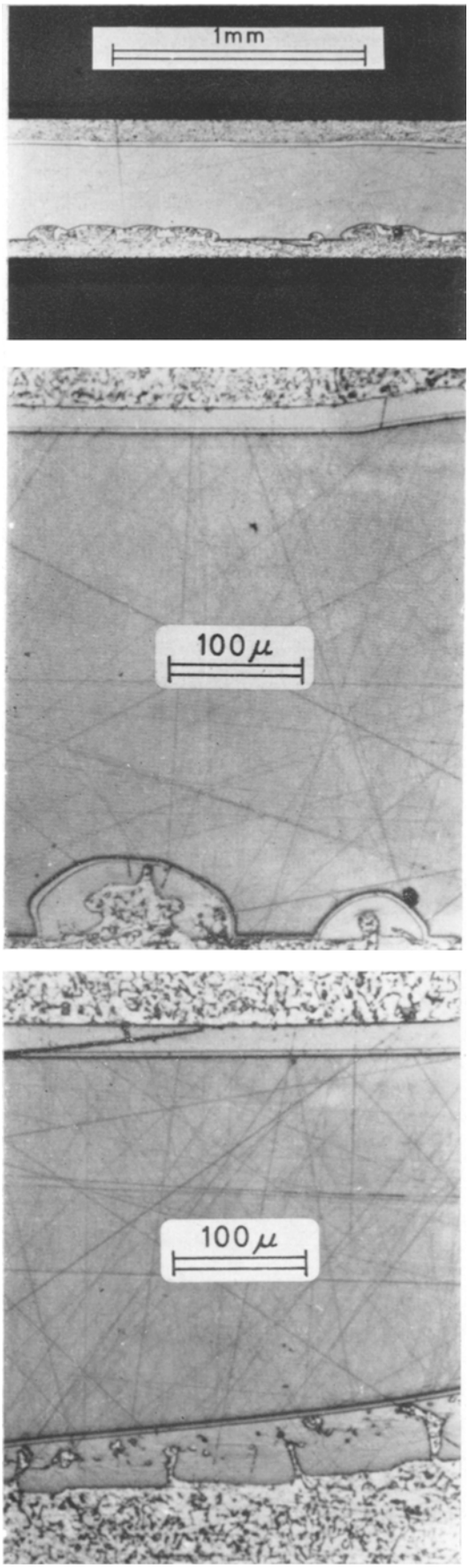


Figure 8 Regular/irregular alloy fronts in positive/negative temperature gradients.

dissociated atoms are at the surface. Thus the alloy front tends to spread over the whole area covered by the preform, and becomes and remains planar during further penetration into the silicon as the temperature rises. This mechanism is illustrated in fig. 9a, where two laterally spreading wet areas have approached each other. In fig. 9b two such areas with different penetration depths have joined.

If the heating rate is fast enough, those parts of the solid surface that melt last do so before dissolution in the initially wet areas has had time to produce local penetration into the solid. Thus when all the requisite surface is wet virtually simultaneously, local inequalities of penetration are minimised and cannot readily be accentuated by negative temperature gradients. Dissolution then proceeds by a mechanism similar to that in temperature gradient zone melting and maintenance of solid/liquid interface planarity is governed by similar conditions.

Unlike aluminium/silicon, gold/silicon eutectic is too brittle to be rolled into foil, so that pure gold is used as a doping element carrier, and fast heating therefore assumes more importance than in the aluminium/silicon alloy system.

3.2.2. Effects of Doping Elements

The nature and amount of doping element included in the gold has an effect on the wetting and alloying characteristics. In fig. 10 alloying of 1% B/Au and 1% Ga/Au foils with silicon is compared, both having been fired under the same conditions, which were such as to produce a fairly uniform penetration on the side of the silicon disc in a positive temperature gradient in the case of boron-doped gold, but not in the gallium/gold alloy. On the negative gradient side of the silicon alloying was poor in both cases.

In fig. 11 the effects of different amounts of doping alloy are contrasted. The change in the boron content of the gold alloy from 0.1 to 1% increases the irregularity of alloying in both directions of temperature gradient.

3.3. Re-growth Processes

During cooling, silicon-rich alloy is deposited from the melt epitaxially on to the undissolved portion of the crystal. If the temperature gradient remains positive, re-growth continues by the mechanism of stable Czochralski crystal growth, the re-growth layer/melt interface remains planar (fig. 12a) and the final re-growth

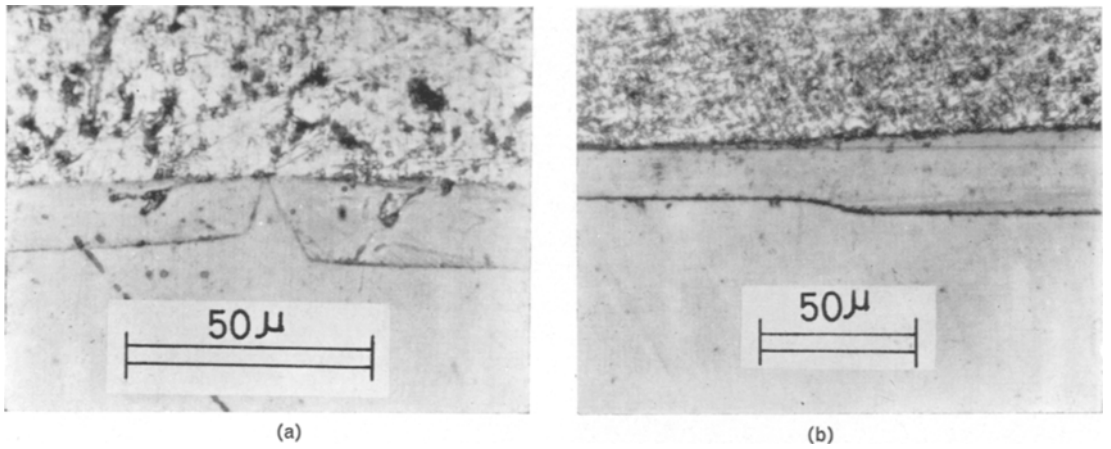


Figure 9 (a) Laterally spreading wet areas. (b) Laterally spreading wet areas of different penetration depths, overlapping.

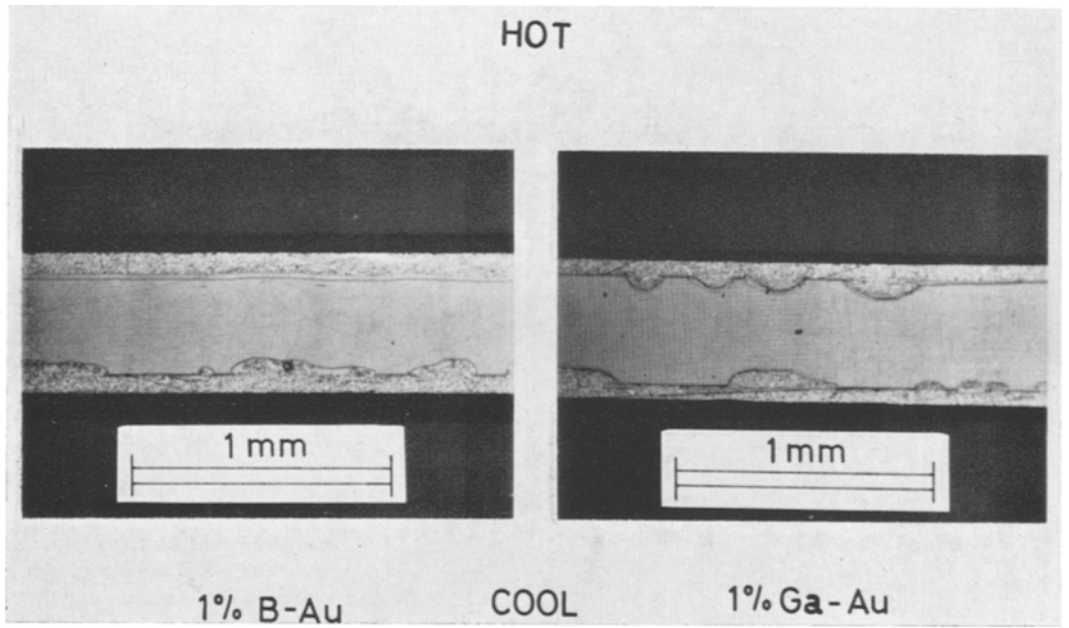


Figure 10 Alloying of 1% B/Au and 1% Ga/Au in a temperature gradient.

thickness is uniform. If the temperature gradient is negative during cooling, re-growth layers of irregular shape and thickness are produced (fig. 8).

It is found that a number of structural features recur in re-growth layers. Some of these can be attributed to changes in composition where a silicon-rich re-growth layer is partially re-dissolved by a laterally-spreading melt of lower silicon content (fig. 12b). Lamellar re-growth (fig. 13a) may extend across the whole cross-section or occur only locally. Fig. 13b shows two deeper re-growth regions representing melts which have been spreading laterally, but have not met in the section, separated by a shallow re-growth layer. The surface in the latter area was evidently not wet until late in the thermal cycle, so that shallower penetration has resulted.

3.4. Control of the Depth of Alloying

The depth of penetration of the alloy melt into the silicon is determined in equilibrium conditions by the silicon-rich liquidus line of the appropriate phase diagram, since there is invariably more

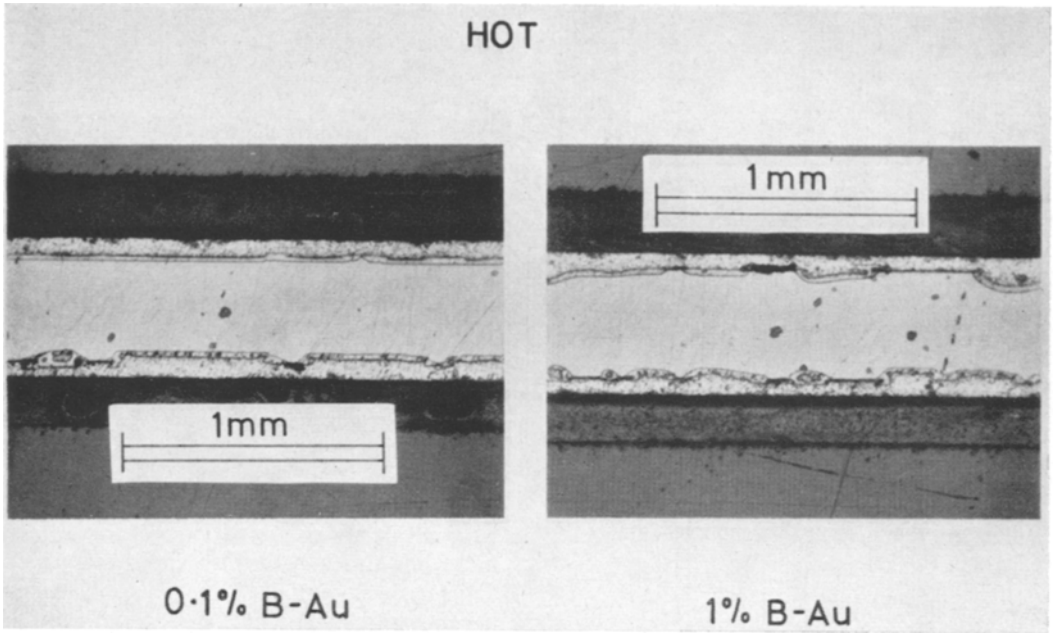


Figure 11 Alloying of 0.1% B/Au and 1% B/Au in a temperature gradient.

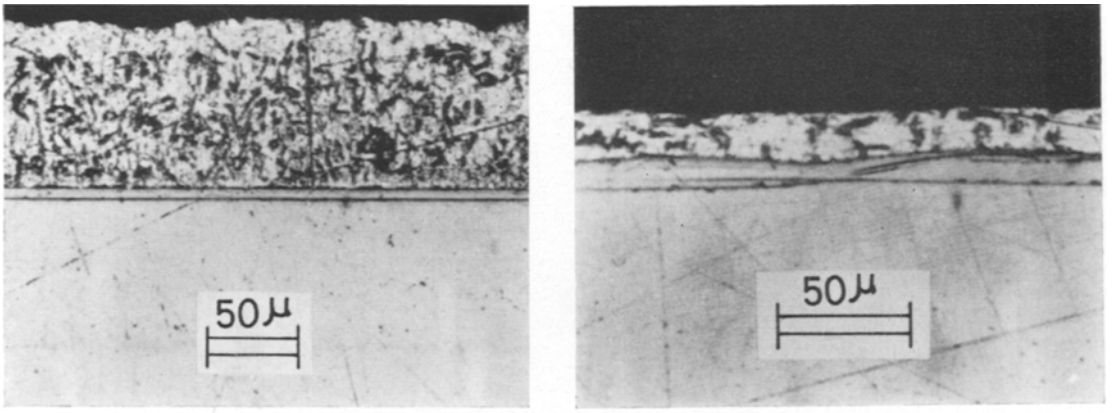


Figure 12 (a) Planar silicon re-growth layer. (b) Partial re-dissolving of re-growth layer by laterally spreading melt.

silicon than alloying metal available to feed the melt. The amount of silicon dissolved is therefore determined by the quantity of metal available, the maximum temperature reached in the thermal cycle, and the extent to which the process approaches equilibrium. The quantity of silicon dissolved is obtained from the following relationship:

$$\frac{\text{vol. silicon dissolved}}{\text{vol. alloying metal}} = \frac{\text{at. \% silicon in the melt}}{\text{at. \% metal in the melt}} \times \frac{\text{density of metal}}{\text{density of silicon}} \times \frac{\text{at. wt of silicon}}{\text{at. wt of metal}}$$

Both metal and alloyed layer approximate in shape to that of a flat cylinder, so that the volume of each is proportional to its thickness. Deviations away from the cylindrical can be neglected as insignificant in relation to the diameters and thicknesses involved. In practice, therefore, in a given thermal cycle the depth of alloying is determined by the foil thickness. Fig. 14 shows the depth of penetration of gold and aluminium into silicon at maximum temperatures from 500 to 900° C as a percentage of the foil thickness, calculated from the phase diagrams.

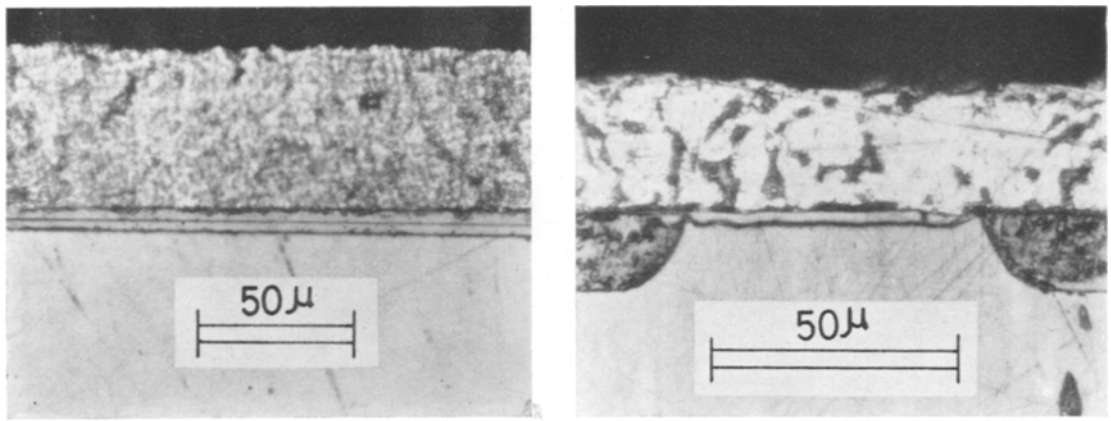


Figure 13 (a) Lamellar re-growth layer. (b) Shallow late re-growth layer.

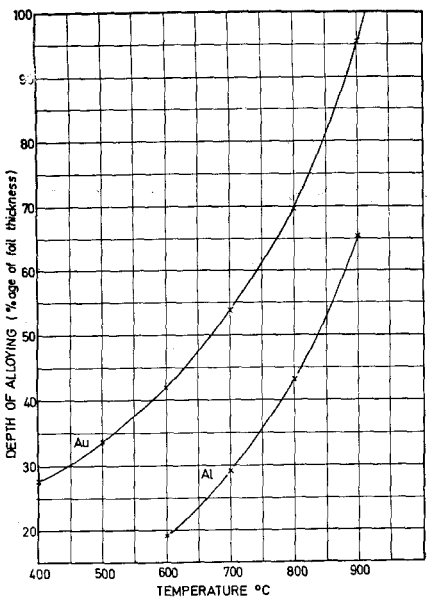


Figure 14 Depth of penetration of Au and Al into Si.

The gold/silicon phase diagram commonly consulted dates from 1920 [13]. The silicon then available was not of semiconductor levels of purity, and the eutectic composition (~31% silicon) was determined by extrapolation of the liquidus line. In 1961 [14] a eutectic composition of 18.6% silicon was determined. By disregarding the three points nearest the eutectic in the early diagram, interpolation between the remaining points and the corrected eutectic composition can be essayed (fig. 15). Depths of penetration measured metallographically agree reasonably well with those calculated from the interpolation (fig. 16). Since the present work was completed, and while writing this paper, a phase diagram [15], determined from both cooling curves and depths of alloying, has been published; the interpolation is virtually identical with this.

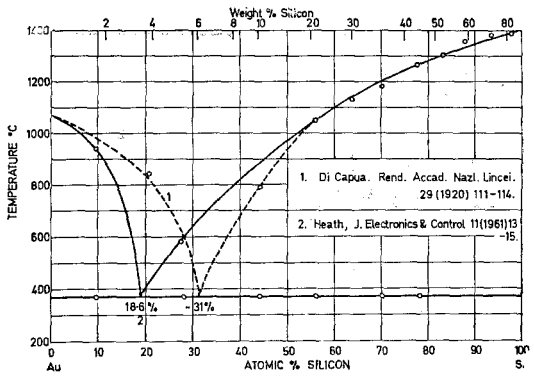


Figure 15 Gold/silicon phase diagram.

If the support plate material is actively involved in the alloying process, as for example in Si/Al/Mo joints, where silicon is removed from solution to form molybdenum silicide at the Al/Mo interface, and further silicon is dissolved to replace it, penetration into the silicon is increased. Actual depths of penetration are approximately twice those calculated from the aluminium silicon phase diagram without taking the formation of molybdenum silicide into consideration. The great variations in mean penetration depths of aluminium shown in fig. 16 may be explained by the unstable nature of molybdenum alloys [16] formed on silicon (111) surfaces.

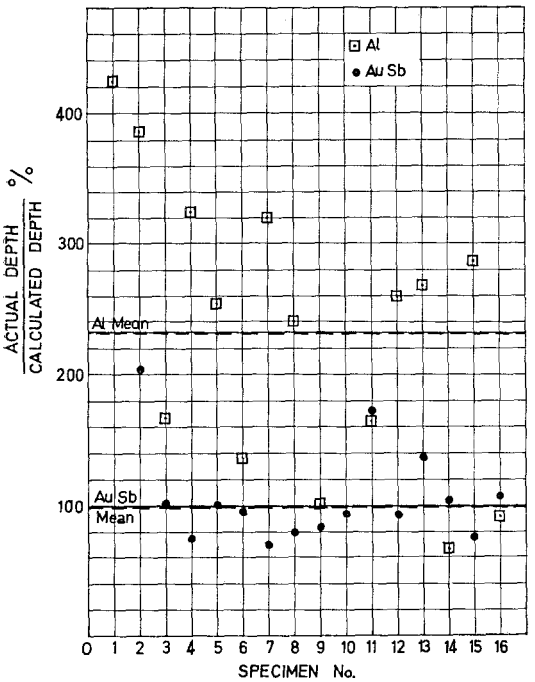


Figure 16 Calculated and actual depths of penetration of Al and Au into Si.

4. Conclusions

The technology of large area alloying in semiconductors is required to produce uniform wetting, dissolution and re-growth of the crystal, and controlled depth of penetration. These requirements can be met by: choice of appropriate crystallographic orientation; fresh surface preparation before thermal cycling; uniform contact pressure between alloying metal and semiconductor; fast heating; highest maximum temperature compatible with the extent of solubility of the semiconductor in the alloying metal; radial temperature uniformity; maintenance of a favourable temperature gradient, where the melt is hotter than the solid, throughout the thermal cycle. Thermal flow considerations are paramount in the design of a jig to achieve this.

Acknowledgements

The authors thank Dr E. Eastwood, Director of Research, The English Electric Company Ltd, for permission to publish this article.

References

1. TCHANG-IL CHUNG, *J. Electrochem. Soc.* **109** (1962) 229.
2. H. D. BARBER and E. L. HEASELL, *J. Appl. Phys.* **36** (1965) 176.
3. E. P. EERNISSE and H. W. THOMPSON JR, *ibid* p. 2652.
4. Y. F. CHANG and H. W. THOMPSON JR, *ibid* **34** (1963) 3136.
5. A. N. KNOPP and R. STICKLER, *Solid State Electron.* **9** (1966) 1119.
6. J. H. WERNICK, *J. Chem. Phys.* **25** (1956) 47.
7. *Idem, Trans. AIME* **209** (1957) 1169.
8. W. A. TILLER, *J. Appl. Phys.* **34** (1963) 2757.
9. P. F. SCHMIDT *et al, Electrochem. Tech.* **3** (1965) 49.
10. A. BONDI, *Chem. Rev.* **52** (1953) 417.
11. L. BERNSTEIN, *Semi. Prod.* **6** (1963) 16.
12. A. J. WALL and D. R. MILLNER, *J. Inst. Met.* **90** (1962) 394.
13. C. DI CAPUA, *Rend. Accad. Nazl. Lincei.* **29** (1920) 111.
14. E. G. HEATH, *J. Electron. Control* **11** (1961) 13.
15. W. GERLACH and B. GOEL, *Solid State Electron.* **10** (1967) 589.
16. J. J. CASEY *et al, J. Electrochem. Soc.* **114** (1967) 201.