Semiconductor Die Separation

Whilst the wafer is circular, the individual circuits or transistors are almost always square or rectangular, in order to be able to separate the slice mechanically into discrete units. The integrated circuit pattern is ‘stepped and repeated’ over the whole surface, but with a narrow gap separating adjacent images. This gap is generally referred to as the ‘street’ and is as narrow as possible consistent with the needs of the die separation process. The patterns cover the whole surface of the slice, which means that a number of the peripheral dice are incomplete by design. Formerly, parts of the wafer were designated for special test devices, but this is no longer common practice as yields have improved.

A number of different methods have been employed for separating a wafer into dice. These include:

- diamond scribing and breaking (the original technique)
- laser scribing and breaking
- back etching
- slurry sawing
- diamond sawing (the preferred current process)

Details of other processes are given in Alternatives for die separation.

Rationale for sawing

As with glass, silicon can be scribed with a diamond point, and then broken along the scribe line by careful bending. This is still occasionally the process used for small dice, such as those for transistors. The scribes are made in both directions, and the wafer, held between carrier films, is carefully ‘rolled’ in both directions to separate individual components. The advantage of scribing is that no material is lost, but there are several disadvantages:

- Even with a well maintained scribe diamond, consistent 100% break out yield is not achievable. At least some components will either be chipped or attached to sections of unwanted silicon.

- Because of the crystal structure of the silicon, the cleavage angle follows the crystal plane which, depending on crystal orientation, could be a 30–60° angle. Few breakouts are vertical, which makes it difficult for automatic die bonding, especially if the pick up tool used has been designed to make edge contact with the chip (a ‘collet’). Scribed dice also tend to ‘ride over each other’ during handling, whereas with vertical edges, the chips stick firmly against each other.
- Scribing will not work effectively on dielectrically isolated ICs or anything made of amorphous or polycrystalline silicon, which doesn’t fracture neatly.

Over the years, substantial improvements have been made in the techniques of separating dice by sawing. This is carried out on a precision machine with a saw blade rotating at high speed (Figure 1) operating with a localised stream of coolant (Figure 2).

Figure 1: Wafer-dicing spindle

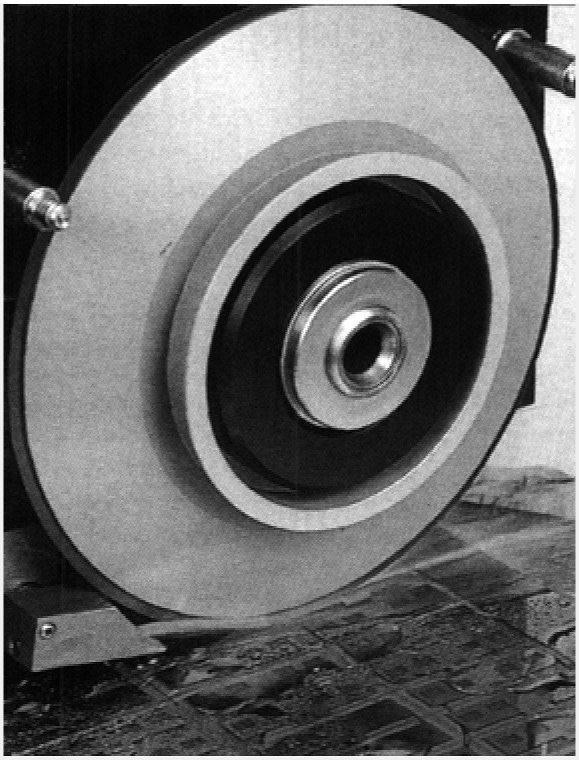
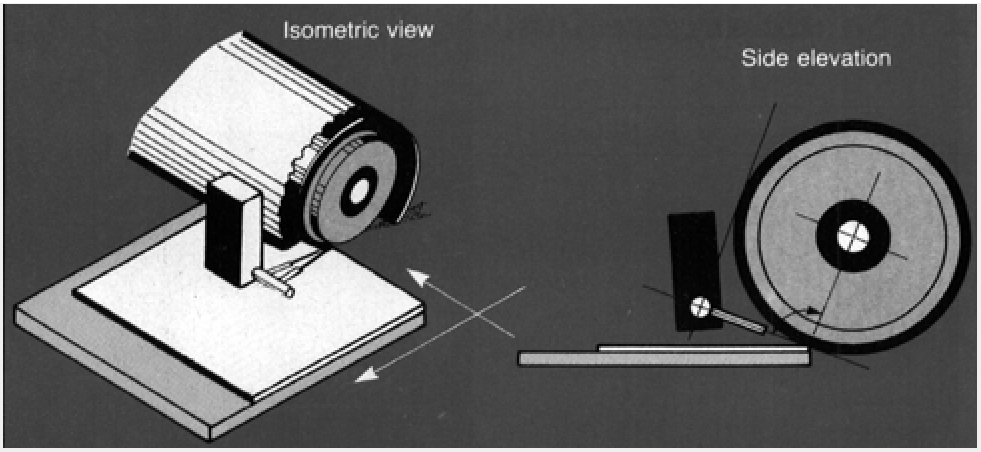
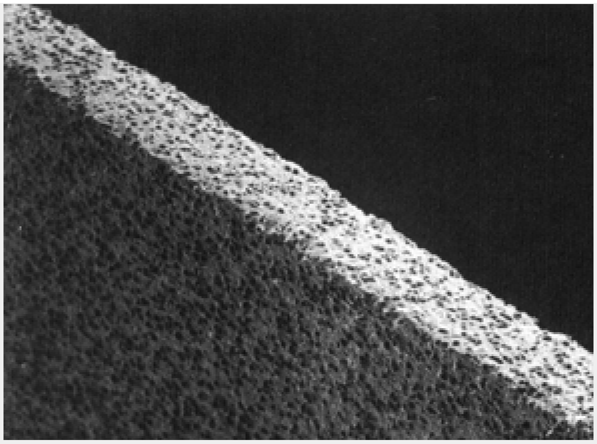


Figure 2: Dual nozzle positioning for efficient cooling



A typical blade is a thin stainless steel disc, to the edge of which fine diamond grit has been resin-bonded (Figure 3). The resultant saw (more properly, a grinder) is able to cut slits in the silicon which are only 70µm wide. The once traditional method of scribe and break is now used only rarely for silicon and is generally limited to brittle material such as gallium arsenide.

Figure 3: Metal-bonded saw blade showing 3–6µm grit size (magnified x80)



Saw blades

Three basic manufacturing processes are used in making saw blades:

- Resin bonded blades, where the resin that bonds the diamonds allows dull diamonds to fracture or pull out of the bond and expose new sharp diamonds. Resin blades are thus self-sharpening and cut well on hard materials.

- Electroplated blades, where nickel bonds the diamonds. Although harder than the resin, individual diamonds are more exposed, allowing the blade to cut freely.
- Metal bonded blades, where diamonds are sintered with metal powder to produce a durable long lasting blade. Although they may require occasional ‘dressing’, metal bonded blades are the most commonly used.

Electroplated blades are used for cutting silicon, gallium arsenide and hard materials such as alumina. Resin blades are used for cutting very soft material such as ferrites and glasses and very hard or brittle materials (alumina or sapphire).

Saws generally vary in width between 20µm and 50µm, but should be matched to the street size. By using the widest possible blade that can be fitted within the street, the less blade wobble there will be, and consequently the longer will be its life.

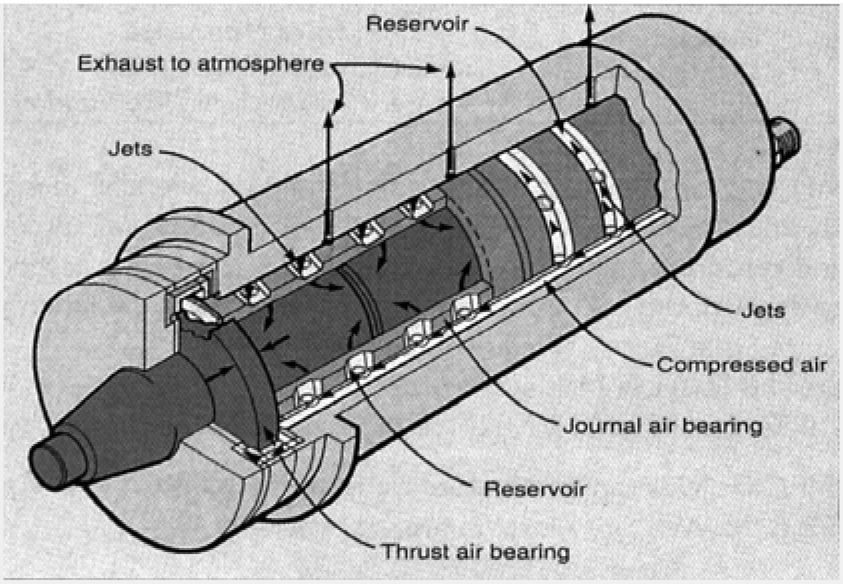
A range of diamond grit sizes is available, but for silicon the most efficient grit size is 3–7µm. The balance is between cut quality (which favours fine grit) and cutting speed and blade life (which are highest with large diamonds).

Gallium arsenide and similar materials are difficult to saw because they are brittle and chip easily, a problem compounded by the crystal orientation of the material. Gallium arsenide wafers require a smaller grit side of 2–4µm to keep chipping at a minimum. Another key factor is that gallium arsenide wafers have narrow streets, in order to conserve material.

Saw practicalities

Cutting wheels, usually driven by brush-less DC motors, rotate at 40–60,000 rpm. Because of their high speed, they need precision bearings. Standard practice is to use air bearing spindles (Figure 4) to give high accuracy and reliable performance, combined with low spindle vibration to reduce chipping. Coolants are normally used.

Figure 4: Air bearing construction and operation



Air bearing spindles

These are critical components in manufacturing equipment such as wafer sawing and grinding. Air bearings allow higher speeds, higher levels of stiffness, with less vibration and less tool movement. The principle of an air lubricated bearing is shown below.

Compressed air flows into a reservoir surrounding the bearing, from which it passes through small holes into the space between the inside of the bearing and the shaft and escapes continuously from the ends of the bearing. When a load is imposed on the nose of the shaft, the pressure distribution automatically changes to resist the force. Spindles have been designed to achieve rotational accuracy of 0.125µm TIR in both axial and radial directions.

With any grinding process, there is some edge damage, and the risk of crack propagation. Depth of cut, feed rate, spindle speed, cooling nozzle design, diamond particle size, and blade flange design are among the many variables that affect the results and, as with back-grinding, it has been reported that faster feed rates both reduce blade life and induce more device chipping.

One way to reduce this problem is to use a two-pass method. A wide saw with tapered sides (the term ‘wide’ is only relative!) is followed by a narrow saw made with finer diamond grit to complete the cut. Such machines have two spindles, one of which leads the other.

The machine needs to be heavy and rigid enough to prevent vibration, and the blade must be mounted so that it runs exactly parallel to the chuck surface, with no end-play or vibration. Sensors are fitted to control depth of cut, and a blade wear monitoring system may be fitted to ensure that the blade height remains within operating parameters. The chuck must also be very flat to guarantee that the saw will cut to the programmed depth within ± 5µm. With tapes typically 75µm thick, greater deviations in depth of cut may result in tape tearing, or alternatively a less than complete cut.

The normal mode of operation is that the slice is aligned and stepped in X and Y, using a vision system to centre the cutter in the street, and the head carrying the rotating saw is moved across the slice. Overall cut placement accuracy of ± 4µm can be achieved, using a Moiré linear encoder, high magnification optics and sophisticated pattern recognition.

Handling and cleaning

During any sawing operation, the wafer has to be kept rigid. Originally, wafers were waxed to support plates, and wax mounting has been surprisingly common although demounting and cleaning the dice after cutting is a major disadvantage, and handling is relatively messy, requiring the support plate to be heated.

Wax has the advantage that it accommodates a non flat substrate, and is sometimes the only method which can be used, but current practice is to stick the wafer to a sheet of adhesive film stretched on a frame, either of metal or (lighter-weight) plastic.

The adhesive has to be carefully formulated to give sufficient adhesion, yet allow release during the die bonding process, and without clogging up the diamond saw, which cuts through the silicon into the top of the adhesive layer. Users are also looking for a tack characteristic which doesn’t age, and a material which doesn’t leave a residue on the chip.

A typical tape is flexible PVC with a synthetic acrylic adhesive bonded to one side. It is tough, has high tear strength and elongation. There are variants in tack level, depending on the needs of the application, and different thicknesses to allow for variations in required dicing saw penetration, but 75µm is a typical thickness.

Sawing debris needs to be removed:

- If colloidal silicon produced during sawing dries on the wafer surface, or bonding pads become contaminated, this will create lead bonding problems.

- Residue may collect in the kerf, producing die attach pick up problems similar to those which result from chipping.

The most common process is to wash in water, with high pressure jets during sawing to dislodge silicon, followed by wafer scrubbing after sawing *without* allowing any time for the wafer to dry.

The frame and film now become the transport and handling medium for the individual dice into which the wafer has been cut. For small volume users, dice may be visually inspected and transferred into a ‘waffle tray’ for subsequent placement, but this introduces unnecessary handling and direct use of the sawing mounts is the usual solution.

At die attach, the film is put on a die matrix expander. In its simplest form, this is an ‘embroidery hoop’ which holds the film and brings it down over a heated plate, when the plastic is evenly stretched in all directions, pulling the dice far enough apart to be picked up by the vacuum tool or collet.

Die marking

One is faced with a requirement only to bond good chips, and on most wafers a proportion at least will have failed electrical probe test. How are these to be distinguished from the others? The usual practice is for the prober to identify faulty chips by depositing a spot of ink.

While early transistor practice used ink containing ferric oxide, which allowed faulty parts to be removed magnetically, current probers use a black ink which can be visually detected by the vision system within the die placement machine. This is helpful in detecting fault spots which may have been added other than at the probe test stage, for example, during visual inspection.

[

Source : http://www.ami.ac.uk/courses/topics/0266_sds/index.html