CHIP-TO-PACKAGE INTERCONNECTIONS

NICHOLAS G. KOOPMAN
TIMOTHY C. REILEY
PAUL A. TOTTA

6.1 INTRODUCTION

Integration of circuits to semiconductor devices, driving the need for improvements in packaging, has been discussed in Chapter 1, "Microelectronics Packaging — An Overview." This is further illustrated in Figure 6-1, wherein the cost of interconnecting on silicon is compared with interconnecting on ceramic substrates and on organic boards, clearly showing the low cost of interconnecting on silicon. There is, however, a limit to the number of circuits and interconnections that can be made on a single piece of silicon, which is currently at about 50,000 for CMOS and 5-10,000 for bipolar. Since most current information systems require a greater number of circuits and interconnections, a number of chips need to be interconnected on to plastic, ce-
Figure 6-1. Cost of Interconnections. The relative cost of interconnection for DSR (double-sided rigid printed wiring board), MLB (multilayer printed wiring board), ceramic hybrids, gate array, and custom silicon. The curves nearly show the advantage of silicon integration in decreasing interconnection costs. After Goddard, Ref. [1], 1979.

Ceramic or coated-metal first-level packages. The electrical connections between the chip and the package, referred to here as chip-level interconnections, are the subject matter of this chapter. Since, for the systems considered here, no first-level package can accommodate all the required chips, a second-level package interconnecting the first levels is required. These interconnections, referred to as package-to-board interconnections, are reviewed in the chapter with that title.

There are three primary chip-level interconnection technologies currently in practice:

- Wirebonding (WB)
- Tape Automated Bonding (TAB)
- Controlled Collapse Chip Connection (C4).

This chapter reviews each of these, indicating the advances made over the last two decades in materials, design factors, fabrication processes, tools, assembly, rework, and reliability. Future trends in this very important area are highlighted.

Two other types of interconnect are discussed briefly: pressure contacts and optoelectronics. The most basic function of the chip-level interconnections is to provide electrical paths to and from the substrate for power and signal distribution. Electrical parameters such as resistance, inductance, and capacitance need to be quantified for each interconnection design, as each of these will affect the total system's performance. In addition, the electrical functions, the C4 and TAB designs provide mechanical support for the chip, which is augmented by the encapsulants used to seal the chip metallization. Since almost any electrical conductor is also a good thermal conductor, some interconnections together with package materials are designed for removal of heat from the chip. Thus each interconnection provides electrical, mechanical, and thermal functions which are discussed in the chapter.

6.2 CHIP INTERCONNECTION EVOLUTION

The silicon integrated circuit chip was created in the mid-1960s. Prior to its existence, chips or dice of silicon or germanium were made as discrete transistors or diodes for a number of years. These devices were typically packaged in hermetically sealed metal cans and interconnected with passive components such as resistors and capacitors on printed-circuit boards. The first transistorized generation of computers was built in this way.

At Bell Laboratories and at IBM, the chip was taken out of its hermetic enclosure and made into a smaller, less vulnerable component. The approach at Bell Laboratories was to use a combination of silicon nitride on the chip to protect junctions, and external gold wiring and beam leads to keep the interconnections free of corrosion. At IBM a thin layer of glass passivation sealed the chip surface and its aluminum-based wiring. Solder-bump interconnections joined the chip to its package. Both the Bell and the IBM chips were mounted face down in the “flip-chip” configuration on ceramic substrates that had thin-film or thick-film wiring and passive components. The chips were encapsulated in silicone gel, which prevented the formation of continuous water films. Thus began the nonhermetic hybrid module era, which was to accompany the second-generation transistorized computer era.

The first integrated circuits became available in manufacturing quantities in the late 1960s. As with their hybrid module predecessor, there were few circuits on each chip of silicon. Aluminum or gold-based thin-film wires were used to “integrate” the active and passive devices embedded in the silicon. The exponential growth of the circuit count per die during the following decade has been phenomenal and unpredictable. At the start there were one to five bipolar logic circuits on a chip, and the first bipolar memory chips had a modest 16 bits in a scratch pad memory application. In the early 1970s, the bipolar logic grew to about 100 circuits and monolithic memory to 128 bits to form the first commercial, bipolar main memory replacing ferrite core, as for example in the IBM 370 System. Today, the number of logic circuits has grown to about 10,000 per chip (bipolar) and one megabit memory arrays, with FET transistors replacing bipolar. Very Large Scale Integration (VLSI) is turning into Ultra Large Scale Integration (ULSI).

The integration and densification process in integrated circuits has caused the continuous migration of intercircuit wiring and connections from
boards, cards, and modules to the chip itself. The surface of the chip, with its multilayer wiring, has become a microcosm of the conductor and insulator configurations that were common on previous multilayer printed-circuit boards and on multilayer ceramic packages (Fig. 6-2). As many as four levels of wiring have been created on the chip. A typical logic chip with 700 circuits and three layers of wiring has approximately 5 m of aluminum wiring on a chip less than 5 mm square. There are over 17,000 via connections from level to level through a micron-thick insulator film of SiO\(_2\) (Fig. 6-3). Yet the wiring capability in the chip greatly lags behind the densification of the silicon devices. To this day, most of the area of the chip (approximately two thirds) serves as a platform for the wiring. Therefore, the vision of 100-megabit memory chips or 50,000-circuit bipolar logic chips requires an enlarged chip size, from the current 5 mm to 10 mm square or more.

The technology of chip surface wiring is truly a part of the packaging technology, but it is too complex and different from typical board and substrate technology to be treated adequately in this book. It is a fact, however, that all the physics and engineering of high-speed transmission line theory applies to the migrated chip wiring as well.

The progress in integrated circuits has led not only to enormous densification of circuits on a chip, but also to the total integration of a computer on a chip. Modern microprocessor chips in handheld calculators have the computing power of second-generation large-scale computers of the mid-sixties. The potential for wafer scale integration is re-examined often with BI-FET chips — Bipolar and FET transistors integrated on the same piece of silicon for optimum performance and cost — emerging as a candidate technology.

The advanced VLSI era has put great demands on the functionality and reliability of ever-increasing numbers of input/output (I/O) connections. An empirical relationship between I/O and the number of circuits to be wired (Rent’s rule) appears to be holding well for mid-to-large systems. This means that the I/O demand will jump from about 100 to 1,000 in the decade ahead.

Heretofore, serial wirebonding of one or two rows of I/Os around the perimeter of the chip has satisfied the needs of ceramic or plastic dual in-line packages. Automated wirebonding today is very fast, efficient, and reliable compared to the manual bonding of the sixties. But wirebonding appears to be yielding to TAB bonding, in which the perimeter density of connections can be doubled or tripled and all bonds made simultaneously. The solder-bump counterpart has evolved into an area array C4 configuration in which the entire surface of the chip is covered with C4s for the highest possible I/O counts. Unlike wirebonding, C4 and, usually, TAB demand bump formation on the surface of the chip when the chip is in wafer form. Bumping the chip has in the past been a deterrent to the widespread use of C4 or TAB in commercial devices. It is an added expense but, in the VLSI era, a necessary one. Typi-
6.3 CONTROLLED COLLAPSE CHIP CONNECTION (C4)

The solder-bump interconnection was initiated in the early 1960s to eliminate the expense, unreliability, and low productivity of manual wirebonding. Where the initial, low-complexity chips had typically peripheral contacts, this technology has allowed considerable extendibility in I/O density as it progressed to full-population area arrays. The so-called Controlled Collapse Chip Connection (C4 or C4) utilizes solder bumps deposited on wettable metal terminals on the chip and a matching footprint of solder wettable terminals on the substrate. The upside-down chip (flip chip) is aligned to the substrate, and all joints are made simultaneously by reflowing the solder (Fig. 6-4).

The C4 joining process has been described in the literature with numerous acronyms. C4, as originally used at IBM, denoted Controlled Collapse Chip Connection. In the industry, it has also been called CCB — Controlled Collapse Bonding — and Flip-Chip Joining, referring to the fact that it is opposite to the traditional back-side-down method of bonding, in which the active side of the chip, facing up, is wirebonded. The term C4 or flip chip is used in this book.

Figure 6-4. Controlled Collapse Chip Connection (C4). The upside-down chip (flip chip) is aligned to the substrate and all joints are made simultaneously by reflowing the solder. After Ohshima, Ref. [3], 1982 and after Fried, Ref. [2], 1982.

Figure 6-5. Terminal Metallurgy Design. The original SLT flip chip (27 mm square) with glass passivation, BLM sealing of via holes, and Cu ball bumps. After Totta, Ref. [4], 1969.

Two other acronyms are used in this section: BLM and TSM. These refer to the terminal connecting metallurgies at the chip and substrate, respectively. BLM stands for Ball Limiting Metallurgy and refers to the region of terminal metallurgy on the top surface of the chip that is wettable by the solder. TSM stands for Top Surface Metallurgy and refers to the terminal metallurgy on the substrate to which the chip and its associated solder ball are joined.

6.3.1 C4 History

The solder-bump interconnection of flip chips, the face-down soldering of silicon devices to alumina substrates, has been practiced for approximately twenty years [5]. First introduced in 1964 with the Solid Logic Technology (SLT) hybrid modules of IBM’s System/360, it was part of a design to eliminate the expense, unreliability, and low productivity of manual wirebonding [6]. The solder bump was also an integral part of a chip-level hermetic sealing system created by the glass passivation film on the wafer [7]. Most semiconductor devices of that era were, in contrast, protected by expensive hermetically sealed cans. The terminal metallurgy design was intended to reseed the access of “via” holes through the glass as well as to provide a means for testing and joining the chip (Fig. 6-5).

Initially, for the discrete transistors or diodes of the hybrid SLT, copper ball standoffs, embedded in the solder bumps, were used to keep the unpassivated silicon edges of the chips from electrically shorting to solder-coated
thick-film lands [4]. Later, in the integrated circuit era, the controlled collapse chip connection was devised. In this technique a pure solder bump was restrained from collapsing or flowing out on the electrode land by using thick-film glass dams, or stop-offs [8], which limited solder flow to the tip of the substrate metallization (Fig. 6-6).

Similarly, the flow on the chip is limited by a ball limiting metallurgy (BLM) pad, which is a circular pad of evaporated, thin-film chromium, copper, and gold that provides the sealing of the via as well as a solderable, conductive base for the solder bump. A very thick deposit (100–125 μm) of evaporated 95 Pb/5 Sn solder acts as the primary conduction and joining material between chip and substrate [9].

The early integrated circuit chips typically had peripheral C4 I/O pads like their wirebonded counterparts. The pads were 125 or 150 μm in diameter, on 300 to 375 μm centers. The pitch of connections was typically compatible with the resolution capability of thick-film (Ag/Pd) electrode screening on the ceramic substrates.

Occasionally it was convenient to have an “in-board” power pad or two in the thick-film technology, but larger numbers of inside I/O pads could not be used until metallized ceramic (MC) technology became available in the mid-1970s, as discussed in Chapter 7, “Ceramic Packaging,” and Chapter 9, “Thin-Film Packaging.” The narrower lines and spaces made possible with etched thin-film Cr–Cu–Cr on ceramic, associated with that technology, allowed wiring escape for double rows of I/O pads and many internal connections. Later, a “depopulated” grid of bumps allowed the interconnection of 120 pads on 700-circuit logic chips. The fully populated area grid array, in which every grid point is occupied by a solder bump, required the complexity of multilayered, cofired ceramic packages. In these packages the distribution of I/O wiring could be accommodated by via “microsockets” and buried layers of wiring as opposed to single-level wiring, where the “escape” of wires is geometrically restricted by the maximum number of lines per channel between I/O connections [10]. The progression of C4 geometry is shown in Figure 6-7.

An example of an area array C4 configuration is shown in Figure 6-8. The I/O count is 120 in an efficient square grid array, which is 11 C4 pads long by 11 pads wide on 250 μm (10 mil) centers. A 125 μm (5mil) solder bump is located at every intersection in the grid except one, which is displaced for orientation purposes [2]. Some packages, such as the cofired alumina ceramic MLC, (Fig. 6-9) use 9 to 133 area array chip sites per package to attain high bipolar circuit densities in IBM’s 4300 and 3081 series computers. Logic and memory chips are mixed as required. As many as 25,000 logic circuits or 300,000 memory bits have been packaged on a single TCM substrate in this technology [10, 11, 12, 13, 14]. The most populous logic flip chip to date is a circuit “computer-on-a-chip” which has 762 C4 solder bumps in a 29 x 29 area array (Fig. 6-10) [15]. Four layers of metal wiring in this chip are used, compared to the two or three levels normally used, previously.

Over time, the C4 technology has been extended to other applications. C4s are used on thin-film resistor and capacitor chips in hybrid module applications [17]. Solder pads for this application are very large — 750 μm in diameter. At the other extreme, Schmid [18] used C4s for precision registration and alignment in the joining of a GaAs waveguide. The C4s in this case were only 25 μm high. The most dense area array reported has been a 128 x 128 array of 25 μm bumps on 60 μm centers, resulting in 16,000 pads [19]. The photolith process for forming these is discussed later. C4s or solder structures similar to C4s are used for attaching chip carriers to boards and have become part of the surface mount revolution discussed in the first chapter. New applications of this technology are being explored continually.

6.3.2 Materials

Melting point has been a prime consideration in the choice of solder alloys for C4s. High-lead solders, especially 95 Pb/5 Sn, have been most widely used with alumina ceramic substrates because of their high melting point, approximately 315°C. Their use for the chip connection allows other, lower-melting-point solders to be used at the module-to-card or card-to-board packaging level without remelting the chip C4s.

![Figure 6-6. Controlled Collapse Chip Connection.](image-url)
A reverse order of assembly (e.g., modules-to-board, then chips-to-module) would require a reverse order of melting point. Josephson superconducting devices have been joined in such a fashion, using an alloy of 51 In/32.5 Bi/16.5 Sn (having a melting point of 60°C) for the chip C4s, while a higher-melting-point eutectic alloy, 52 In/48 Sn (having a melting point of 117°C), was used for pins and for orthogonal connections to the chip carrier [20, 21, 22].

Joining to organic carriers such as polyimide-Kevlar® [23] or printed-circuit boards [24, 25, 26] also requires lower processing temperatures. Here, intermediate melting point solders, such as eutectic 63 Sn/37 Pb (melting point 183°C), and PbIn alloys, such as 50 Pb/50 In (melting point of approximately 220°C), have been used. A listing of solder alloy compositions and melting points is shown in Table 6-1. Some phase diagrams relevant to C4 solder joints are shown in Figure 6-11.

---

19 Trademark of DuPont Corp.
The choice of terminal metals, which is described in detail later, will likely depend upon the choice of solder. For example, silver and gold are poor terminal metals to be used with the SnPb alloy. In only a few seconds gold completely dissolves into the liquid solder. In these cases, another solder alloy could be used, such as Indium [18], which has a much lower solubility for gold; or one of the other lower-solubility metals could be used for the terminal. Thus, Cu, Pd, Pt, and Ni are very commonly used for both BLM and TSM thin films. All of these metals form intermetallics with Sn, which limits the reaction rates with PbSn solders. On the chip side, this terminal metal is normally sandwiched between an adhesion metal layer of Cr or Ti, and a passivation metal layer, usually of thin gold. The copper, palladium, or nickel thin films on the substrate are either similarly protected with gold [11, 12, 13, 14] or are dip-soldered. In the latter case, some of the solder for the C4 joint is supplied by the substrate [28, 29]. MLC substrates usually use a flash of gold on nickel [3, 11, 12, 13, 14, 17, 30, 31]. Thick-film substrates have the palladium or platinum alloyed with gold or silver and are dip-soldered prior to the joining operation. AuPt, AgPd, AgPdAu, and AgPt have been reported [8, 9, 32, 33, 34, 35, 36] as thick-film TSM pads.

6.3.3 Design Factors
Some of the factors affecting the material choices for the terminal and solder have already been discussed, but other variables must also be considered in C4 design. The joints must be high enough to compensate for substrate non-planarity, especially for the older version of thick-film substrates. Because solder surface tension "holds up the chip," a sufficient number of pads must be provided to support the weight of the chip. Typically this becomes a cause for concern only with very low I/O devices, such as memory chips or...
Table 6-1. Selection of Low Melting Solder Alloys. After Wassink, Ref. [27], 1984, reprinted with permission of Electrochemical Publications Ltd., Ayr, Scotland.

<table>
<thead>
<tr>
<th>°C</th>
<th>Sn</th>
<th>Pb</th>
<th>Bi</th>
<th>In</th>
<th>Cd</th>
<th>Other Elements</th>
</tr>
</thead>
<tbody>
<tr>
<td>16</td>
<td></td>
<td>24</td>
<td></td>
<td></td>
<td>76 Ga</td>
<td></td>
</tr>
<tr>
<td>20</td>
<td>8</td>
<td></td>
<td></td>
<td></td>
<td>92 Ga</td>
<td></td>
</tr>
<tr>
<td>25</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>95 Ga; 5 Zn</td>
<td>100Ga</td>
</tr>
<tr>
<td>29.8</td>
<td>10.8</td>
<td>22.4</td>
<td>40.6</td>
<td>18</td>
<td>8.2</td>
<td></td>
</tr>
<tr>
<td>47.2</td>
<td>8.3</td>
<td>22.6</td>
<td>44.7</td>
<td>19.1</td>
<td>3.3</td>
<td></td>
</tr>
<tr>
<td>58</td>
<td>12</td>
<td>18</td>
<td>49</td>
<td>24</td>
<td>21</td>
<td></td>
</tr>
<tr>
<td>61</td>
<td>16</td>
<td>33</td>
<td>51</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>70</td>
<td>13.1</td>
<td>27.3</td>
<td>49.5</td>
<td></td>
<td>10.1</td>
<td></td>
</tr>
<tr>
<td>70-74</td>
<td>12.5</td>
<td>25</td>
<td>50</td>
<td></td>
<td>12.5</td>
<td></td>
</tr>
<tr>
<td>72.4</td>
<td></td>
<td></td>
<td>34</td>
<td>66</td>
<td></td>
<td></td>
</tr>
<tr>
<td>79</td>
<td>17</td>
<td>57</td>
<td>26</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>93.3</td>
<td>42</td>
<td>51.7</td>
<td></td>
<td>44</td>
<td>8.1</td>
<td></td>
</tr>
<tr>
<td>95</td>
<td>31.3</td>
<td>50</td>
<td></td>
<td></td>
<td>14</td>
<td></td>
</tr>
<tr>
<td>98</td>
<td>16</td>
<td>52</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>103.0</td>
<td>26</td>
<td>53.5</td>
<td>20.3</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>96-110</td>
<td>22</td>
<td>28</td>
<td>50</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>112</td>
<td>48</td>
<td>43.5</td>
<td>56.5</td>
<td>25</td>
<td></td>
<td></td>
</tr>
<tr>
<td>127.7</td>
<td>57</td>
<td>59</td>
<td></td>
<td>75</td>
<td>25</td>
<td></td>
</tr>
<tr>
<td>139</td>
<td>43</td>
<td>62</td>
<td>38</td>
<td></td>
<td>10.2</td>
<td></td>
</tr>
<tr>
<td>144</td>
<td>49.8</td>
<td>32</td>
<td>100</td>
<td>43.1</td>
<td>33</td>
<td>1.5 Ag</td>
</tr>
<tr>
<td>156.4</td>
<td>57</td>
<td>67</td>
<td></td>
<td>9 Zn</td>
<td>35 Ag</td>
<td></td>
</tr>
<tr>
<td>170</td>
<td>67</td>
<td></td>
<td>36</td>
<td></td>
<td>15 Au</td>
<td></td>
</tr>
<tr>
<td>178</td>
<td>62.5</td>
<td>34</td>
<td>3</td>
<td>82.6</td>
<td>17.4 Zn</td>
<td></td>
</tr>
<tr>
<td>178</td>
<td>62.5</td>
<td>34</td>
<td>3</td>
<td></td>
<td>17.4 Zn</td>
<td></td>
</tr>
<tr>
<td>183</td>
<td>64.9</td>
<td>38.1</td>
<td>91</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>183</td>
<td>62</td>
<td>38</td>
<td></td>
<td>91</td>
<td></td>
<td></td>
</tr>
<tr>
<td>215</td>
<td>85</td>
<td></td>
<td></td>
<td>82.6</td>
<td>11.5 Ag</td>
<td></td>
</tr>
<tr>
<td>221</td>
<td>96.5</td>
<td>100</td>
<td>89</td>
<td>100</td>
<td></td>
<td></td>
</tr>
<tr>
<td>232</td>
<td>200</td>
<td>92.2</td>
<td>2.8</td>
<td>89</td>
<td></td>
<td></td>
</tr>
<tr>
<td>288</td>
<td>91.5</td>
<td>2.5</td>
<td>95</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>304-312</td>
<td>5</td>
<td>95</td>
<td>0.5 Zn</td>
<td>100</td>
<td></td>
<td></td>
</tr>
<tr>
<td>318</td>
<td>99.5</td>
<td>100</td>
<td>88 Au; 12 Ge</td>
<td>91 Au; 3 Si</td>
<td>100 Zn</td>
<td></td>
</tr>
<tr>
<td>321</td>
<td>100</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>322</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>356</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>370</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>420</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

Figure 6-11. Some Phase Diagrams Relevant to Soldering. Temperature, along the vertical axes, is given in °C; Concentration, along the horizontal axes, is given in mass percent (Hansen and Smithells give concentrations in atomic percentages). After Wassink, Ref. [27], 1984, reprinted with permission of Electrochemical Publications Ltd., Ayr, Scotland.

carriers, which are very bulky. Numerous studies have been published showing the interrelationship between BLM and TSM size, solder volume, chip weight, and C4 height. Figure 6-12 shows these relationships over a very wide range [17].

Extra "dummy" pads, added to supplement those needed for simple electrical connection, have often been used to enhance the mechanical behavior, reliability, or thermal performance of the assembly [38].
is important from the standpoints of both electrical design and reliability. The effect of distance to neutral point (DNP) is discussed later in this chapter as it relates to thermal cycle fatigue.

As VLSI chips become more and more dense, higher I/O counts will drive full area arrays of terminals. In this case, the pad size and location are fixed by the chip size and available real estate per pad allocated by a fully populated area array.

The number of C4 pads as a function of chip size and pad geometries is shown in Table 6.2 wherein the possibility of 155,000 pads on a 20 mm chip is indicated. Figure 6.13 shows the pronounced density advantage of area array versus a single-perimeter row, as pad sizes and spacings decrease.

### 6.3.4 Fabrication Processes and Tools

Metal mask technology [4, 5, 9, 17, 21, 33, 34, 39, 40] is most widely used for terminal fabrication. BLM and solder are evaporated through holes in a metal mask (Fig. 6.14) and deposited as an array of pads onto the wafer surface. This is a low-cost batch process, simultaneously involving many chips per wafer and many wafers per evaporation. DC sputter cleaning of the via hole is commonly used to remove undesirable oxides and photoresist just prior to evaporation [4]. This step assures low contact resistance to the aluminum metallurgy of the device and good adhesion to the SiO₂ or polyimide insul...
The multilayered structure of the BLM can be described by using $Cr-Cu-Au$ as an example. A typical evaporator would have numerous metal charges with thermal energy supplied by resistance, induction, or electron beams (e-Guns). $Cr$ is evaporated first to provide adhesion to the passivation layer, as well as to form a solder reaction barrier to the aluminum. A phased layer of $Cr$ and $Cu$ are co-evaporated next, to provide resistance to multiple refows. This is followed by a pure $Cu$ layer, to form the solderable metalurgy. A flash of gold is then provided as an oxidation protection layer. This is necessary because the wafers are normally exposed to air before going on to the next step of solder evaporation. This is performed through the same mask as the BLM but in a separate evaporator. This requires a "thick" (of the order of 100µ thick) mask for the solder. While lead and tin are usually in the same charge (single molten pool), the higher vapor pressure component, $Pb$, deposits first, followed by tin on top of the lead. Reflow in an $H_2$ ambient furnace at about 350°C melts and homogenizes the pad and brings it to a spherical shape. In addition $H_2$ assures reduction of oxides of $Pb$ and $Sn$.

Photolithographic processes and combinations of photolith and metal masks are becoming more and more popular for fabricating terminals [19, 30, 41, 42, 43, 44, 45, 46, 47, 48]. Most common is a sequence of blanket deposition of the BLM, application of photoresist, development of a pattern in the resist followed by electrodeposition of the solder; then removal of the resist and subetching of the BLM, using the plated solder bumps as a "mask" [19, 41, 43, 45, 47]. An alternative sequence is to blanket-deposit $P$...

Figure 6-13. Number of Pads vs. Pad Separation for Different Chip Sizes. Input/output terminal trends. After Nyc, Ref. [37], 1986.

Figure 6-14. Metal Mask Technology. (a) Tooling for alignment of mask to wafer. (b) Masking and evaporation of chromium/copper/gold. (c) Masking and evaporation of lead/tin. (d) Reflowed solder bump. After Brownell, Ref. [34], 1974, reprinted with permission of ISTHM.
sist, subetch BLM through the resist, then deposit solder by a variety of techniques including solder dip [42], solder ball placement [48], or metal mask evaporation [44]. The single masking processes (unmask) are significantly simpler and cheaper than multiple mask techniques, although they are not as flexible in providing varying amounts of solder. In some applications higher volumes of solder are able to decrease strain, as discussed in section 6.3.6, "C4 Reliability," on page 382.

The final operation before dicing of the wafer into individual chips is the electrical testing of each chip. Mechanical probes are used to contact the soft solder bumps fabricated earlier.

The formation of wettable-surface contacts on the substrate (providing a mirror image to the chip contacts) is achieved by thick- or thin-film technologies. Thin-film contact technology is similar to the BLM described previously, but thick-film technology involves development of wettable surfaces by plating nickel and gold over generally nonwettable surfaces such as Mo or W (conductors usually used within the ceramic substrate). Solder flow is restricted by the use of glass or chromium dams where necessary.

Various thin and thick-film processes that are typically used are shown in Figure 6-15.

6.3.5 Assembly/Rework

Once the BLM, TSM, and solder are in place, as described previously, the joining of chips to the substrate using C4 technology is straightforward. Flux, either water-white rosin [45, 49, 50, 51] for high-lead solders or water-soluble flux [20, 21, 52] for low-lead and other low-melting solders, is normally placed on the substrate as a temporary adhesive to hold the chips in place. Such an assembly is then subjected to a reflow thermal cycle involving either individual chip joining using a local heat source to bond one chip at a time [23] or, joining an assembly of several chips to a substrate simultaneously using an oven [50].

One of the greatest features of the C4 process is its self-alignment capability arising from the high-surface-tension forces of solders [40]. Chip pads and their counterparts on the substrate may be separated by as much as three times the average bump radius, but if the mating surfaces touch and are reasonably wettable, self-alignment will occur. As many as a million C4 bonds can be made in one hour using automated tools.

Once the chip-joining operation is complete, cleaning of flux residues is accomplished with such solvents as chlorinated solvents or xylene for rosin flux and water for water-soluble flux. The assembly is then electrically tested.

Rework: Reworking may be necessary to replace defective chips on multichip modules or for engineering changes. In this case, one or more chips may be removed from the module and replaced with new ones. This process is described by Puttitz [50] for large multilayer ceramic modules.

Device debonding can be achieved by mechanical means (torque or pull) or by melting the solder and directly lifting the chip from the surface with a vacuum pencil. Residual solder is left on the substrate, which must be removed prior to joining the new chips (see Fig. 6-16). The excess solder is removed from the substrate pads by means of a hot gas tool (Fig. 6-17), which ensures that adjacent sites are not contaminated with the excess solder.
The TSM microsockets are then ready for chip joining (Fig. 6-18). Multiple chip replacements per chip site have been performed using this method.

6.3.6 C4 Reliability

This subject is covered in detail in Chapter 5, "Package Reliability," but since one of the major limitations of this technology is how large a chip area can be bonded and still remain reliable, a brief discussion is included here.

A question often raised regarding flip-chip bonding is the ability of the joint to maintain structural integrity and electrical continuity over a lifetime of module thermal cycling. A thermal expansivity mismatch between chip and substrate will cause a shear displacement to be applied on each terminal. Over the lifetime of a module, this may lead to an accumulated plastic deformation exceeding 1,000% [53]. A quasi-empirical model was developed by Norris and Landsberg [32] that relates the cyclic lifetime to cyclic deformation parameters. It is based on the Coffin-Manson relationship [54] between fatigue life and plastic strain amplitude but with two terms added to account

See the in covering the "Coffin-Manson Equation," on page 295, in Chapter 5, "Packability."
for time-dependent behavior: a frequency term, where lifetime increases with frequency to a low power, and a maximum-temperature term, where lifetime decreases with maximum temperature. With the assumption of a lognormal failure distribution, a product sample may be tested in an accelerated thermal cycle; then, based on the statistical lifetime to electrical failure, the projected field lifetime may be extrapolated. Using this technique, an interconnection failure rate projection was made for logic chips in System/370 [32] of no more than $10^{-9}/1,000$ hours per bond at the end of life. By the end of 1975, 540 billion MST interconnection hours had been accumulated with no wearout failures reported [55], yielding a 50% confidence-level-estimated failure rate of $1.3 \times 10^{-10}/1,000$ hours, in good agreement with the earlier projection.

These results demonstrate the inherent capability of solder interconnections to withstand high-strain accumulations; the results also demonstrate the approximate validity of the projected failure rate. However, the early MST chip is only about $1 \times 1 \text{ mm}$ in size and has only 12 peripheral bumps. As discussed in Chapter 5, "Package Reliability," the mismatch shear deformation is proportional to the distance between a given pad and the neutral point referred to as DNP (the point on the chip that remains stationary relative to the substrate during a thermal excursion). Because the neutral point is near the chip center, the maximum value of DNP is roughly proportional to chip size. Moreover, the Norris-Landberg model and most subsequent experiments show that lifetime is inversely proportional to shear deformation, and thus DNP, raised to a power that approaches two. Thus, with the evolution from MST to much larger and denser C4 footprints, thermal wearout was considered with renewed interest.

The wearout model subsequently has come under close scrutiny. It has been suggested, for instance, that a dwell time factor be added [56]. Also recommended is a complete reformulation starting with the constitutive equation for solder that incorporates crack growth and creep deformation [57]. It has been further suggested that several competing mechanisms come into play in thermal wearout, including cavitation. Recent experimentation [58] also suggests that wearout is much more complex than implied by the simple equations discussed above. Scanning electron microscopy photos of joints at various stages of thermal cycling show that mismatch between the solder and chip plays a role in the damage. Also, low and high volumes of solder joints fail by different mechanisms, thus emphasizing the effect of C4 shape.

A further complicating factor in modeling wearout is that thermally induced strains are not uniform within the joint. An initial attempt [59] was made to incorporate joint shape into failure-rate projections. Despite the simplicity of the model, reasonable agreement has been obtained in experiments where joining geometry has been intentionally varied and the joints mechanically tested by single-cycle or cyclic torquing of the chip [17, 59].

6.3 CONTROLLED COLLAPSE CHIP CONNECTION (C4)

More sophisticated techniques, taking into account time- and temperature-dependent solder properties, will be required to understand geometric effects fully. Chip bending also must be included in a complete analysis [46, 60].

In summary, simple models that accompanied the introduction of C4 joining have, until now, proven adequate to estimate field behavior and as an aid in product design. This can be attributed not only to their simplicity and their qualitative rationality, but also primarily to the fact that existing products have experienced wearout failure rates too low to be of concern. As chips grow larger and pad counts in the hundreds become common, new or revised models will be required that reflect a greater understanding of the wearout mechanism and better precision in failure projection.

6.3.7 Thermal Mismatch Reliability: Extensions

Together with a greater understanding must come an extension of the C4 in its ability to accommodate larger and denser chips without affecting system reliability. Existing modeling and testing techniques, however imperfect, have been used to evaluate various extension schemes, several of which show promise. They may be subdivided into strain reduction, geometry or shape improvement, and alternate solders.

The thermal mismatch displacement across the pad can be kept to a minimum by arranging the footprint to minimize the DNP; for example, by deleting corner pads or, in an extreme case, by a quasi-circular array [5].

Joint geometry is dictated by the wetting areas on chip and substrate, solder volume, and the weight of the chip. Unless the chip is very heavy, the joint has the shape of a spherical segment [17, 59, 61], and its height is uniquely determined by interface radii and volume. The previously mentioned geometric model [59] claims that wearout depends upon shape and that a joint can be geometrically optimized to extend lifetime. Although verifications have been primarily by mechanical testing, with its inherent limitations, the following optimization philosophy and sequence are probably valid:

1. Consistent with other design and process constraints, the interface areas should generally be as large as possible.
2. There is an optimum ratio of substrate-wetting area to chip-wetting area, which must be determined experimentally for each particular material set. For thin-film copper lands on ceramic (discussed in Chapter 7, "Ceramic Packaging"), the ratio is about 1.2. The optimum ratio is characterized by a roughly even distribution of thermal cycle fails between chip and substrate (solder crack near BLM or TSM intermetallics).
3. For fixed interface radii, there is an optimum solder volume. A model exists, but because verification has been by mechanical testing only [17], thermal cycle testing for the optimum is recommended.

The preceding discussion pertains to joints whose shapes are dominated by solder surface tension. Under these conditions the joint takes the shape of a doubly truncated sphere, truncated at each end by the contact metallurgies. Normally, optimization of the spherical segment joint is of limited extendibility value, improving lifetime by less than 50%. Enforced changes in shape away from a spherical segment, on the other hand, can produce very large effects. Heavy chips that depress the joint [17] severely reduce lifetime, while stretched or elongated joints substantially extend lifetime. Mechanical testing has shown an order of magnitude difference in fatigue life between an hourglass and a barrel-shaped joint, with the fracture location of the hourglass joint shifted to the center of the joint [62, 66]. Stretched pads have been fabricated by a number of techniques including using two different solders on the same chip [62]. Called SST (Self-Stretching Soldering Technology), this technique makes use of the surface-tension forces of larger bumps of one solder to stretch the lower volume functional bumps (Fig. 6-19). Two additional concepts being pursued in Japan are illustrated in Figure 6-20 wherein solder columns are being stacked to achieve improved fatigue life.

In addition, solder has been cast into helical copper coils to form very high-aspect-ratio solder columns. It has been applied to joining leadless ceramic chip carriers (LCCC) to glass-epoxy printed-circuit boards (Fig. 6-21). This structure has not been scaled down to sufficiently fine dimensions to be applicable to integrated circuit chip interconnections. Freestanding cast-solder pillars (without the copper helix) have been developed for package-to-board interconnections [26].

Among solders, which have been evaluated as alternatives to 95 Pb/5 Sn, the Pb/In system has shown the most fatigue enhancement. Thermal cycle lifetime has been shown to be quite sensitive to composition with a minimum at 15 to 20% In [49, 67]. There is a two times improvement over 95 Pb/5 Sn at 5% In, three times at 10%, and twenty times at 100% In [67]. Pure In is being used for optoelectronic device joining [18, 19]. Early work on integrated circuits emphasized 50 Pb/50 In as a compromise between ultrahigh thermal cycle reliability and processing constraints. Implementation of this alloy was limited by two factors: increased corrosion susceptibility in nonhermetic packages [49, 67, 69], and a substantially accelerated thermomigration rate over Pb/Sn alloys [70]. By the latter phenomenon, the thermal gradient between chip and substrate causes a condensation of vacancies at the chip BLM region, leading to premature high resistance or mechanical wearout. Later work emphasized low (3–5%) In alloys [49], which provided less fatigue enhancement but were not susceptible to the corrosion and thermomigration difficulties of 50 Pb/50 In. Testing 95 Pb/5 In

![Figure 6-19. Self-Stretching Soldering Technology (SST). After Satoh, Ref. [62], 1983.](image)

Figure 6-19. Self-Stretching Soldering Technology (SST). After Satoh, Ref. [62], 1983.

at several thermal cycle frequencies showed a similar relationship to that found by Norris and Landzberg for 95 Pb/5 Sn.

By far, the largest effort to reduce the strain drastically is by matching the substrate thermal expansions to that of silicon, as demonstrated in Figure 6-22. This initial study using a polyimide-Kevlar® organic substrate

![Figure 6-20. Some Methods Being Pursued to Extend C4 Life. (a) stacked solder bumps using polyimide. After Matsui, Ref. [63], 1987 ©IEEE. (b) stacked solder bumps using multiple solders. After Fujitsu, Ref. [64].](image)

Figure 6-20. Some Methods Being Pursued to Extend C4 Life. (a) stacked solder bumps using polyimide. After Matsui, Ref. [63], 1987 ©IEEE. (b) stacked solder bumps using multiple solders. After Fujitsu, Ref. [64].
[23] has been followed by a number of packaging efforts that take advantage of the improvement in lifetime of the C4 connection as a result of thermal expansion match. The use of AlN, SiC, Si, and glass-ceramics as first-level packages (as discussed in Chapter 7, "Ceramic Packaging") and polyimide-Kevlar®, copper-Invar-copper boards with polyimide and copper thin-film wiring (as discussed in Chapter 13, "Coated-Metal Packaging") are being actively pursued for direct chip-attach applications. Gallium arsenide has also been matched using sapphire [18]. Power cycling is replacing thermal cycling to evaluate these material combinations [60, 71, 72]. Process constraints, wireability, dielectric constant, and heat dissipation must, of course, be among other factors considered in selecting an alternative substrate material. Epoxy resins with coefficients of thermal expansion similar to that of solder (260 x 10⁻⁵°C) are also being pursued to improve thermal fatigue life as shown in Figure 6-23.

Among other reliability concerns that have been reported in the literature, thermomigration has been shown to be a major concern for applications where high-temperature gradients are coupled with high-diffusivity, low-melting-point solders [70]. Corrosion has been encountered in high-humidity testing of PbSn [69]. Solder void defects are addressed in [55]. Palladium depletion of AgPd thick films by PbSn prompted a switch to AgPdAu ternary alloy for the substrate electrode [73, 74].

A more fundamental problem receiving increased attention is that of soft errors in devices caused by alpha-particle emission of trace quantities of radioactive materials in the packaged assembly [29, 75, 76]. In the case of C4, high-lead alloys are mostly used for computer applications, and lead almost invariably brings with it trace amounts of uranium and thorium. This

---

**Figure 6-21.** Solder Columns. Close-up view of the solder columns connecting a LCCC to a glass-epoxy printed-circuit board. After Cherian, Ref. [65], 1984.

**Figure 6-22.** Effect of Thermal Expansion Coefficient on Fatigue Life. After Green, Ref. [23], 1978, ©IEEE.

**Figure 6-23.** C4 Life Extension by the use of Thermal Expansion Matched (to Solder) Resins. After Nakano, Ref. [68], 1987.
problem will demand more attention as VLSI devices become more dense and as the critical charge levels in the device become smaller.

6.3.8 Heat Dissipation with C4

Heat dissipation is covered in great detail in Chapter 4, "Heat Transfer in Electronic Packages," and only a brief discussion is included here as it pertains to C4 bonding. A traditional constraint of all flip-chip bonding has been the poor heat dissipation capability compared to back-side bonding. Early chips with 10 to 20 peripheral C4 bumps bonded to an alumina substrate could dissipate approximately 0.5 watt [77]. While the interconnection portion of the heat transfer path was at least an order of magnitude more resistive than a comparable backbond [9, 78] a comparison of overall thermal resistance showed only about a 50% increase for flip chips for a typical module [78, 79]. Thermal patterns for flip chips depend upon device location, size, metallization, and number of the terminals, and upon the thermal resistance of the substrate. Thus, numerical analysis techniques are required for reasonable performance projections [80, 77].

Area array bumps, in addition to meeting the density needs of VLSI, also provide enhanced dissipation because of the increased number of joints (effectively approaching a continuous heat transfer layer) and their greater proximity to the devices. Flip chips are becoming competitive thermally with backbonding. For instance, an air-cooled module with six 4.5 mm chips, each having an 11 x 11, 0.25 mm grid array of solder joints, can dissipate approximately 1.5 watts per chip [10]. Higher power levels have been achieved with new, high-thermal-conductivity ceramics such as AlN [81] and SiC [82].

Flip chips may have another thermal advantage compared to backbonding. Because the back of the chip is free of mechanically or electrically delicate surface features, it is thus amenable to direct contact by a wide variety of heat sinks whose thermal conductivity is often better than the plastic or ceramic package they are backbonded to. An example is the IBM multi-chip module [80, 83], where spring-loaded pistons transfer heat from the back of each chip to a water-cooled plate, augmenting the traditional solder joint thermal path. Four watts per chip or 300 to 400 watts for a 100-chip module can be dissipated. This is further discussed in Chapter 7, "Ceramic Packaging."

Finally, liquid immersion cooling [84] and cryogenic applications [20, 21] have also been demonstrated for C4 interconnected structures. The latter application, for Josephson devices that operate at 4.2 K, was especially noteworthy, because all of the materials were subjected to a very large ΔT between room temperature and operating temperature, and because the material's intrinsic properties, such as resistance to cracking, are quite degraded at low temperature. Orthogonal solder connections are also used to join silicon slices at right angles to each other. For such a joint, matched expansion materials are required [20, 21, 22].

6.3.9 C4 — Future Trends

As VLSI proceeds to denser chips, C4 densification will follow. Bumps of 25 microns, resulting in over 10,000 pads per chip in experimental devices, have already been shown [19]. Photolithographic processing will be a must for such devices, and with that, area arrays will become more and more common. Significant activity in matched-expansion substrates, or alternatives that alleviate the fatigue limitations, is being actively pursued. High-thermal-conductivity packaging materials and innovative configurations to improve heat dissipation of the devices are also being evaluated. More attention will be given to defects, for both yield and reliability needs. Efforts to purify all packaging materials for low alpha emission will continue.

The range of applications for C4-like structures is becoming much wider, given the more diverse materials used in new electronic devices and substrates (for example, In C4's on GaAs chips on sapphire substrates for optoelectronics).

As the competitive position and capability of solder-bump interconnections become markedly improved with these recent developments, it is likely that C4s will be more generally used in the future.

6.4 WIREBONDING

6.4.1 Definitions and Types

As pointed out previously, wirebonding is the most common chip-connection technology in the microelectronics industry and the most common wirebonded assembly is in a "plastic package." Chapter 8 is devoted to this topic; therefore, this section will emphasize wirebond interconnections rather than wirebond packages. Wirebonding starts with mounting the chip, back side down, with die-bonding epoxies and metals onto a substrate (die bonding). The wires are then bonded, one at a time, to the chip and substrate by one of three processes: ultrasonic (U/S) bonding, thermocompression (T/C) bonding, and thermosonic (T/S) bonding. The shape of the bond wires, or "flying leads," is a direct result of the bond cycle; these are shown in Figure 6-24. Thermocompression and thermosonic techniques can produce either a ball bond or a wedge bond [87]. The ultrasonic technique is associated with a wedge bond.
The three most common wirebonding methods are as follows:

**Ultrasonic (U/S) Bonding** — In U/S bonding, the wire is guided to the bonding site, then pressed onto the surface by a stylus ("bonding wedge"). While the wire is firmly clamped between the bonding tool and the bond pad, a burst of ultrasonic vibrations is applied to the wedge. The combination of the pressure and the vibration accomplishes a metallurgical cold weld between the wire and the pad metallization (Fig. 6.25).

 Aluminum-alloy wire is most commonly employed in ultrasonic bonding, but gold, copper, and other metals are also used in special applications.

**Thermocompression (T/C) Bonding** — Thermocompression bonding is accomplished by pressing the wire against the bond site metallization at an elevated temperature. Practically all T/C bonding is performed using gold wire, and practically all wire loops are formed using a "ball bond" at the first bond site and a "wedge bond" at the second (Fig. 6.26). The bonding tool is a capillary of alumina, tungsten carbide, or other refrac-

**1. Wedge-bond**

**2. Wedge-bond**

Figure 6.25. Ultrasonic (U/S) Bonding. After Singer, Ref. [87], 1984, reprinted with permission from Semiconductor Magazine.

tory material, and the bonding surface is heated to 300° to 400°C. Gold wire is particularly suitable for T/C bonding because it deforms readily under the bonding capillary at elevated temperatures, exposing clean metal, without formation of any oxide that inhibits joining.

**Thermosonic (T/S) Bonding** — The principal features of U/S and T/C bonding are married in T/S bonding (as in Fig. 6.26). Gold wire ball-and-wedge bonds are made, as in the T/C technique, but the capillary is driven by a burst of ultrasonic power at each bond to augment metal joining. As a consequence, the substrate temperature may be lower than in T/C bonding. Thermosonic bonding has been particularly successful in attaching wires to hard-to-bond thick-film hybrid substrate metallizations.

### 6.4.2 Wirebond History

Wirebonding was the earliest technique to be applied to device assembly [89]. Bell Laboratories published the first results in 1957, on a technique called "Thermocompression Bonding" [90]. Several years later, bonding machines became commercially available and the technique began to be generally applied. Later, ultrasonic bonding and thermosonic bonding techniques were introduced that enabled bonding temperatures to be reduced for ther-
Figure 6-26. Thermocompression (T/C) Bonding. Thermosonic ball-wedge bonding of a gold wire. (a) Gold wire in a capillary. (b) Ball formation accomplished by passing an H₂ torch over the end of the gold wire or by capacitance discharge. (c) Bonding accomplished by simultaneously applying a vertical load on the ball while ultrasonically exciting the wire (the chip and substrate would be heated separately to about 150°C). (d) A wire looped and a wedge bonded under load and ultrasonic excitation. (e) A wire broken at the wedge bond ready to go to step (a). (f) The geometry of the ball-wedge bond that allows high-speed bonding. Because the wedge can be on an arc from the bond, the bond head or package table does not have to rotate to form the wedge bond [4]. After M. Szcz, Ref. [88], 1982.

mally sensitive devices and for surfaces that were difficult to bond ultrasonically at room temperature.

The initial bonders were manually operated and rather labor intensive. Automatic wirebonders with pattern recognition and computer control now are available to make the complete bond in less than 0.2 second per wire[87].

Die attachment, which precedes wirebonding, was originally accomplished by forming a gold-silicon eutectic between the die and the substrate. Now polymeric materials are also in common use in plastic packages as well as for nonconductive packages [91]. Hydrolyzable chlorides in these materials have been reduced from approximately 1,000 ppm to only a few ppm, thereby greatly reducing corrosion concerns [92]. Solders and silver-filled glasses are also being used for large chips [91]. Dramatically improved bond yield and reliability have come about from the development and use of new evaluation techniques, such as the ball bond shear test, and new surface analysis methods, such as ESCA. Originally, only gold and aluminum wires were available. Now, feasibility has been demonstrated for numerous other wire materials, including palladium [93] and copper [94, 95]. Even aluminum, long the standard for ultrasonic bonding, has been joined using the ball bond technique [96].

Whereas early chips had very few wires, the transition from discrete devices to LSI to VLSI has brought larger chips with high pad counts, requiring wirebonding to be extended to as many as 224 perimeter wirebonds on a chip. Even though wirebonding is basically a perimeter lead technique as opposed to an area array, it remains the dominant chip bonding technology in the industry.

6.4.3 Die Bond Materials

Shukla[91] has reviewed die-bonding materials. The die-bonding process includes a large choice of materials, which may be broadly classified into the three categories described below.

**Solders:** These include single, binary, and ternary metallic compositions and may be further classified into two subgroups, hard and soft solders. Hard solders (Au-Si, Au-Sn, Au-Ge) have rather high flow stresses (onset of plastic flow), thereby offering excellent fatigue and creep resistance. The disadvantage of using hard solders stems primarily from their lack of plastic flow, which leads to high stresses in the silicon chip because of the thermal expansion mismatch between the die and the substrate. The soft solders, on the other hand, are low-melting binaries and ternaries (e.g., Pb-Ag-In and Pb-Sn compositions), which have a high degree of plastic strain capability.

**Organic Adhesives:** Epoxies and polyimides filled with precious metals have found widespread acceptance as die-bond materials in low-cost packaging (plastic packages). They offer the advantage of lowest processing temperatures and lower cost compared to the gold-based hard solders. Numerous papers have claimed the lowering of thermal stresses in silicon by the use of organic adhesives. These adhesives are typically filled with a metal to provide the thermal and electrical conductivity required in most applications (silver being the most common filler). However, the use of organic adhesives in high-reliability packages (hermetically sealed) has been slow in acceptance because of the poor thermal stability of these materials, especially when filled with silver. The poor thermal stability coupled with the out
trapped solvents and other gaseous species remains a challenge in the application of these materials to high-reliability VLSI applications.

**Glass Adhesives:** Silver-filled specialty glass materials have emerged in the past few years as an alternative in VLSI die-bond applications. These materials offer the possibility of a void-free die-bond interface with excellent thermal stability. However, the currently available materials require high processing temperatures (400°C) and oxidizing ambients for optimum adhesion, thereby creating some special processing concerns. In addition, the glass die-bond materials also need solvents and binders for processing purposes, leading to the problem of complete solvent removal similar to the case of organic adhesives.

### 6.4.4 Bonding Wires

Wirebonding materials have been reviewed by Gehman [86]. It is customary to specify the mechanical properties of bonding wire by setting acceptable ranges for the break strength (BS) and elongation (EL). Tensile properties are determined from a standard stress strain curve (Fig. 6-27).

**Aluminum with 1% Silicon Wire:** Ultrasonic bonding using aluminum-alloy wire was developed in the late 1950s as a means of avoiding the “purple plaque” detected when Au wire was bonded to Al chip metallization. “Purple plaque” refers to a purple-colored Au-Al intermetallic compound sometimes associated with early bond failures, as discussed in Chapter 5, “Package Reliability.”

Very pure Al is too soft to be easily drawn to a fine-diameter wire, so from the beginning, alloying to toughen the metal was required. Combining Al with Si was a “safe” choice so far as the danger of semiconductor poisoning was concerned, and Al/Si alloying in commercial practice was well established. The nominal composition of Al + 1% Si emerged as the standard alloy for U/S bonding wire, and it still is.

In retrospect, from a metallurgical standpoint, 1% Si as a solute in Al for U/S bonding wire was an unfortunate choice. The equilibrium solid state solubility of Si in Al at 20°C is of the order of 0.02% by weight. Only at temperatures above about 500°C is Si at 1% in equilibrium solid solution. Thus, at ordinary bonding temperatures, there is always a tendency for Si to precipitate, forming a silicon second phase. When uncontrolled, excessive Si segregation may degrade wire bondability or bond integrity.

**Aluminum–Magnesium Wire:** Al alloyed with 0.5% to 1% Mg can be drawn into fine wire. The finished product exhibits breaking strength and elongation similar to that obtained in Al + 1% Si. This alloy wirebonds satisfactorily and is superior to Al + 1% Si in resistance to fatigue failure at first-bond heels and to break-strength degradation after exposure to elevated temperatures. An early attribution of device field failures to the presence of Mg in the wire alloy has been conclusively disproven. In comparable production volumes Al/Mg wire should be no more expensive than Al + 1% Si.

The advantage of Mg over Si as a solute in Al wire alloys is that the equilibrium solid solubility of Mg in Al is about 2% by weight. This compares, as noted previously with about 0.02% Si equilibrium solid solubility in Al at 20°C. At 0.5 to 1.0% Mg concentration there is no tendency toward second-phase segregation as is the case with Al + 1% Si.

**Gold Bonding Wire:** In producing Au bonding wire, the control of surface finish and surface cleanliness is the highest priority. These precautions are required to prevent clogging of bonding capillaries. The mechanical parameters break strength, and elongation are not unimportant, but adequate control over these properties is more easily achieved with 99.99% pure Au (the usual purity specification) than with the more complex alloy Al + 1% Si.

Like Al, very pure Au is very soft. It is difficult to draw it into fine wire. Relatively small concentrations of impurity atoms, however, render the metal workable. Any number of foreign atomic species may be added to pure gold to toughen the metal, but Be and Cu are the two alloying elements currently
in use. Cu-doped Au wire generally contains 30 to 100 ppm Cu by weight; Be-doped material has typically 5 to 10 ppm Be by weight.

Be-doped wire, sometimes referred to as "Be-stabilized," is about 10% stronger as-drawn than Cu-doped ("Cu-stabilized") wire. Cold-worked wire, however, is too stiff and strong for T/C bonding; therefore the finished temper is generally adjusted by an annealing procedure. Be-doped material retains a 10% to 20% higher break strength compared to Cu-doped wire after the customary annealing operation. The increased strength for Be-doped wire is advantageous for automated T/C bonding; since this process involves high-speed capillary movements that generate greater wire stresses than in slower, manual bonders. For manual bonding Cu-doped wire is satisfactory.

**Gold-Wire Substitutes:** The remarkable fluctuation in the price of Au during the last decade periodically causes increased attention to the development of gold-replacement materials for T/C and T/S bonding. Another reason for substituting for gold wire is the need to find a stronger wire material that would allow the use of finer wire diameter, thereby allowing potentially finer bond spacing. Adequate substitutes for Au have been developed for a few particular applications, but no generally suitable material is yet available.

Ag wire, for example, may be drawn to finer diameter, balls nicely on T/C and T/S bonders, has excellent thermal and electrical conductivity, and is thermocompression-bondable to conventional semiconductor metallizations. Long-term reliability of Ag-Al junctions, however, is suspect. Intermetallic compound formation is reputed to be much more serious than with Au/Al [86]. Susceptibility to corrosion may be an additional problem. It should be possible to develop chip metallizations compatible with Ag wire, but there seems to have been little exploration of this approach.

The possibility of ball bonding of Al wire is of great interest. Considerable success has been shown with ball bonding of Al using special tooling to form the ball [98]. Bond-breaking strengths are compared in Figure 6-28.

Other metals used include copper [94, 95] and palladium [99]. Both these metals are stronger than gold, and care must be taken to avoid cracking of the silicon under the bond pad. Figure 6-29 shows a nicely formed palladium ball [99].

### 6.4.5 Design Factors

Two important factors in wirebond design are wire diameter and pad spacing at the chip and substrate. Practical constraints of wirebond tooling and motion and properties of the wire generally limit wirebonding to a perimeter footprint.

![Figure 6-28. Comparison of Bond-Breaking Strengths. Relation between strain and the bond-breaking force. After Onuki, Ref. [98], 1984 ©IEEE.](attachment:image)

The shape and length of the bonded wire are quite important to its reliability. For example, too taut a loop can lead to thermal cycle failures; and, at the other extreme, too large a loop can cause the wires to lean over and short to each other or the device. [101]. Bond pads can be staggered on different levels to reduce the propensity of wire shorting, as shown in

![Figure 6-29. Formed Palladium Ball. SEM micrograph of Pd wire. After Bischoff, Ref. [99], 1984 ©IEEE.](attachment:image)
Table 6-3. Spacing Guidelines. Minimum pitch in microns for 110 x 110 micron wirebonding pad arrangement. After Dehaine, Ref. [100], 1984 © ISHM.

<table>
<thead>
<tr>
<th>Method</th>
<th>Wire Diameter</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>μm</td>
</tr>
<tr>
<td>Gold ball</td>
<td>25 30 38</td>
</tr>
<tr>
<td>Aluminum wedge</td>
<td>150 162 175</td>
</tr>
<tr>
<td></td>
<td>100 112 125</td>
</tr>
</tbody>
</table>

Figure 6-30. Tool proximity to the adjacent wire limits density, as illustrated in Figure 6-31. Spacing guidelines for various wire sizes are shown in Table 6-3.

For most applications 25- to 38-micron-diameter wire is adequate for signal and power connections. However, high power or current requirements of power devices on some hybrid circuits need the current-carrying capability of 125- to 150-micron wires [103]. Wirebonding is usually applied directly to the aluminum, the most common metallization used in the semiconductor industry. Given that some of the aluminum remains exposed after wirebond-

Figure 6-32. Planarity Limitation of Thick Film. After Noble, Ref. [102], 1983 © ISHM.
ing, terminal pads to "seal the chip" may be added to avoid corrosion concerns. Substrate pads include thick-film and thin-film types similar to those used for C4 (see section 6.3.1, "C4 History," on page 367). Thick-film applications, however, require special attention to planarity effects, as shown in Figure 6-32.

6.4.6 Fabrication Processes and Tools

Tooling advances in wirebonding have been very dramatic. It is a marvel of engineering that all of the following operations can be performed with precision in less than 0.2 second: 1) forming the ball; 2) locating and positioning the bonding tip; 3) moving the ball to the chip or surface while automatically compensating for different bond-pad heights; 4) applying precise power and force and dwell; 5) feeding the wire while the tip forms the wire loop and moves into position for the bond at the opposite end of the wire; 6) forming that bond; 7) then excising the wire. Bonding rates are thus in the 18,000 per hour range. Some of the various wirebond tool manufacturers are shown in Table 6-4.

Optical pattern recognition is one of the main factors that has made totally automatic wirebonding systems possible [87]. There are still problems with reflectivity, varying chip heights (affecting focus), and heat-wave distortion. To overcome these problems, the toolmaker (K & S) incorporated a twin-camera system that automatically recognizes dies from 0.25 to 12.5 mm units or larger. A new lens system is currently being developed to provide three fixed magnification levels (fields of view), with the selection of each under program control. This is intended to solve the problem of focusing on devices of different heights. Multiplexing of pattern recognition systems to a number of bonders is also under development.

Hybrid devices pose some interesting wirebonding problems beyond those normally associated with monolithic wirebonding. Some hybrids have over 100 components and thousands of wires, thereby requiring larger memories than some automated bonders have available. One may find a wide range of materials (epoxy, gold, aluminum, solder, preforms, eutectics, and glass) being used in one assembly. ICs and chip components are of varying heights and sizes, requiring different wire lengths. Special features like programmable reverse motion and calculated loop heights to accommodate short and long wire lengths are being incorporated into new hybrid bonders.

Fully automatic chip mounters for die bonding are also available [104]. Toshiba has a silver paste mouter that operates at 0.58 second per chip; Mitsubishi, a solder chip mouter operating at 0.7 second per chip; and at Hitachi, a fully automated in-line assembler is programmed to use either Au – Si eutectic or Ag paste by mode switching [104].

### Table 6-4. Wirebond Tool Manufacturers

<table>
<thead>
<tr>
<th>Company</th>
<th>Model</th>
<th>Ball</th>
<th>Wedge</th>
<th>Large wire automatic</th>
<th>Semi-automatic</th>
<th>Manual</th>
</tr>
</thead>
<tbody>
<tr>
<td>ASM America</td>
<td>AB300</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>ASM America</td>
<td>AB500</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Foton</td>
<td>8050</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Hughes</td>
<td>HMC-2460</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Hybond</td>
<td>523</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Hybond</td>
<td>552</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>K &amp; S</td>
<td>1419</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>K &amp; S</td>
<td>1470</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>K &amp; S</td>
<td>2406</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>K &amp; S</td>
<td>4123</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>K &amp; S</td>
<td>4124</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Mullen Equip.</td>
<td>8-150-05</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Mullen Equip.</td>
<td>8-166</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Mullen Equip.</td>
<td>TPB-460</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Orthodyne</td>
<td>20</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Orthodyne</td>
<td>40</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Sem. Equip.</td>
<td>1300</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Sem. Equip.</td>
<td>1350</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>West Bond</td>
<td>44XY</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>West Bond</td>
<td>47EZ</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

6.4.7 Assembly and Reworking

Assembly of single chip modules is straightforward, involving die bonding, sequential wirebonding, encapsulation, and testing (see Chapter 10, "Package Sealing and Encapsulation"). Parts that do not pass the tests are normally discarded.

In multichip assemblies, more factors need to be considered. Reworking a defective wirebonded chip is not nearly as simple as a C4-joined chip. Each of the wires must be removed, the die debonded, a new device placed, and new wires sequentially bonded on the substrate pads that had previously been bonded. Yield improvement activities are paramount for multichip assemblies.

Individual chip yield as a function of yield per wire can be calculated from the simple relation [100]:

\[ Y_D = Y_W^N \]

where

\[ Y_D = \text{yield per device}. \]
### Table 6-5. Chip Yield vs. Wire Yield for a 224-Wire Device

<table>
<thead>
<tr>
<th>Yield Per Wire</th>
<th>Yield Per Device</th>
</tr>
</thead>
<tbody>
<tr>
<td>0.9</td>
<td>(5 \times 10^{-11})</td>
</tr>
<tr>
<td>0.99</td>
<td>0.105</td>
</tr>
<tr>
<td>0.999</td>
<td>0.799</td>
</tr>
<tr>
<td>0.9995</td>
<td>0.894</td>
</tr>
<tr>
<td>0.9999</td>
<td>0.978</td>
</tr>
</tbody>
</table>

\(Y_{fr}\) = yield per wire.

\(N = \text{number of wires per device.}\)

Table 6-5 shows the wirebond yield that can be expected from a 224-wire device ([100]).

In a similar way the yield per substrate assembly as a function of chip yield and chips per substrate can be determined (Fig. 6-33).

Clearly, extremely high wirebond yields are necessary for an even moderate yield of multichip modules with high chip and wire counts.

---

**Figure 6-34. Ball Shear Test Setup.** (a) Detailed sketch of manual ball-shear probe. (b) Sketch of manual shear probe in use. After Harman, Ref. [106], 1984 ©ISHM.

Some causes of defects [100] are:
- wire misalignment at package level
- aluminum bondability at chip-level
- wire sagging and "essing"
- environmental dust

Improvements in these areas have come about by improved pattern recognition systems, pretesting chip lot bondability, using sufficiently heavy wire, and performing the bonding in clean room environments. The rule of thumb is do it right the first time rather than try to fix it afterward.

An important diagnostic tool leading to improved yield and reliability is the ball bond shear test (Fig. 6-34). Normally, simple pull testing is performed to test wirebond integrity. However, the bonded area of a well-bonded ball bond has been estimated to be five to ten times larger than the cross-sectional area of the wire. Thus, variations in actual bond strength may be masked by a low breaking strength of the wire. Ball bond shear testing has potentially greater sensitivity in assessing the integrity of the bond. This is generally supported by the results shown in Figure 6-35. Note from this figure that conventional wire pull testing showed no degradation with time, whereas the ball shear test showed roughly a 60% strength reduction under the same condition.

One key question in the use of ball shear testing is how to set numerical values for acceptance criteria. In a recent work a correlation was presented between ball shear test results and wire pull test results [107]. This correlation can be quite useful in establishing acceptance criteria.

Use of ball shear testing has resulted in significant improvement in yield for gold T/C bonds (Fig. 6-36) and improved understanding of the role of trace layers of organic contamination from simple exposure of substrates to laboratory air (Fig. 6-37).

Another potentially very important use of the ball shear test is its extension to nondestructive ball shear testing. This could be very useful for
products of high intrinsic or strategic value, for example, in military applications. Detailed experimental results are provided in Panousis and Fischer [107].

Figure 6-35. Shear vs. Pull. Gold ball bond shear force and pull force vs. time at 200°C. The ball bonds were made with 1-mil (25-μm)-diameter gold wire, bonded to integrated circuit aluminum bonding pads. Note the change of scale from shear force (left) to pull force (right). Error bars represent ±1 standard deviation from the mean. After Harman, Ref. [106], 1984 © ISHM.

Figure 6-36. Yield Improvement. A comparison of the ball shear strength distribution transistor wirebonds before and after instituting the ball shear test for early production. T/C bonds were made using 1.5 mil- (38-μm)-diameter gold wire. After Harman, Ref. [106], 1984 © ISHM.

Figure 6-37. Effect of Trace Organic Contamination. Effect of surface contamination on the thermocompression bonding of gold pulse bonded. Data are for T/C ballbonds made with 25-μm-diameter gold wirebonded to gold-chromium metallization with a bonding force of 98 gf. The contamination in this figure resulted from exposing substrates to laboratory air for various lengths of time. After Harman, Ref. [106], 1984 © ISHM.

6.4.8 Wirebonding Reliability

As wirebond yields have been improved, so also has wirebond reliability. Most reliability failures are due to manufacturing defects, and as manufacturing practices are refined, reliability failures have decreased.

The die bond, whether it be solder, epoxy, or glass, can fail by thermal cycle cracking. Many of the comments made for C4 reliability regarding expansion matching, strain reduction, and materials selection apply here. See also comments regarding solders versus organic and glass adhesives for the die bond. The plastic package discussed in Chapter 8 has its unique problems, especially regarding encapsulate interactions, corrosion, and wire breakage.

The most widely known reliability problem in wirebonding technology involves the Au-Al interface. Gehman [86] notes that the intermetallic compounds AuAl₂ (purple plague) and Au₄Al₃ (white) form -n--- suitable
conditions when Au wire is bonded to Al metallization, or Al wire is bonded to Au. "It has been demonstrated that the intermetallic compounds themselves are not harmful, but their presence indicates that bond integrity may have been otherwise degraded. Because these compounds are brittle, wire flexing, coming from vibration or thermal cycling, may more easily induce metal fatigue and stress cracks, leading to bond failure. Of more concern is the behavior at elevated temperatures, where Al diffuses rapidly into the Al-rich AuAl2 phase, leaving behind voids (Kirkendall voiding) at the Al–AuAl2 boundary. Likewise, Kirkendall voiding is also observed at the junction of Au–metallic and Au–rich AuAl2 phases. If excessive intermetallic compound formation occurs, these voids may coalesce, resulting in an open circuit in bond pad metallization or in bond lifting. C.W. Horsting has pointed out an additional Kirkendall voiding failure mode associated with the presence of impurities in plated gold metallization. Ni, Co, Fe, and B have all proved deleterious. Some intermetallic compound formation is unavoidable if a circuit containing Au–Al junctions is heated into the 400° to 450°C temperature range, as in sealing a ceramic dual in-line package, for instance. Catastrophic failures can be prevented, however, by minimizing the time spent at elevated temperatures, by assuring the purity of gold metallization, and by following accepted design rules for metallization thickness" [86].

Gold-plating impurities, especially thallium, have been shown to form low-melting eutectics that weaken grain boundaries in thermocompressive bonding bull [108, 109, 110]. Stress-induced creep in the wire may lead to wire failure.

6.4.9 Future Trends

Dramatic advances have been made with wirebond rates by automatic bonders, and resultant cost reductions are significant. It is expected that wirebonding will remain a widely used technique through the 1990s, particularly for low I/O chips up to about 224. It is one of the simplest and fastest techniques to implement. However, densification much below 100 μm pitch is not anticipated, as shown in Table 6-3. With the restriction of perimeter pads, it will be dominant in the lower end of the chip I/O spectrum.

Efforts are continuing to improve material purity and to evolve new combinations of materials that accompany changes in chips and substrate metallurgies. The use of aluminum ball bonding for improved bonding rates and of copper wire for joining to copper thin and thick films is increasing. Further emphasis on clean-room processing, typical in the silicon facility, will become more common in the package assembly as well.

6.5 TAPE AUTOMATED BONDING

6.5.1 Introduction

Another approach for chip packaging has emerged, that uses an alternative to wirebond or C4 for chip interconnection. This approach is Tape-Automated Bonding (TAB), a concept initiated in the 1960s by GE and developed in recent years by many major semiconductor manufacturers. It was originally developed as a highly automatable technique for packaging large volume, low I/O devices but has been applied, as well, to high I/O devices, up to and exceeding 300 chip interconnections. An early precursor to the TAB structure currently in use is the beam lead structure, in which electroplated leads are prepared on silicon chips for subsequent connection to second-level packaging [111]. Over the years TAB has evolved into areas of application, as illustrated in Figure 6-38, requiring increased electrical performance for high-end computer products, such as very high-speed integrated circuit (VHSIC) [112] and supercomputers by NEC Corporation [68] and ETA Systems [113]. The TAB process involves bonding silicon chips to patterned metal on polymer tape, e.g., copper on polyimide, using thermocompression (T/C) bonding. Subsequent processing can be carried out in strip form through operations such as testing, encapsulation, and burn-in, followed by excising of the individual packages from the tape and attachment to the substrate or board by outer lead bonding. TAB technology, in principle, can be considered both a chip connection as well as first-level package, the latter role in view of groundplane TAB and as a result of bonding to second level directly.

A key factor, central to the topic of this section, that affects the materials and geometry of the TAB package is the T/C bonding process. T/C bonding is the joining, under pressure at elevated temperature, of the TAB tape leads to the metal structures on the silicon device. These joints are made in a single operation, thereby differing from the serial approach of wirebonding. Another distinction with respect to wirebonding is the absence of ultrasonic energy, which is frequently used during wirebonding. This imposes a restriction on the metal surfaces to be joined, since ultrasonic wiping is not always available to disrupt the oxide layer on the interconnection pads of the silicon chip. Many T/C-bondable structures have been demonstrated that utilize different metallurgies and bonding reactions. These bonding reactions include diffusion welding of similar or dissimilar metals, and transient liquid metal reactions. The nominal bonding parameters (temperature and force) must normally be programmed to several values throughout the bonding cycle to optimize joint reliability. Machine characteristics such as bonding planarity, thermal and mechanical properties of the bonding tool (thermocouple), and temperature uniformity may strongly affect the quality of T/C bonds.
The sections that follow present examples of current structures and materials for TAB packaging, along with factors that influence the bonding processes.

TAB essentially consists of several steps: (1) Inner Lead Bonding (ILB), in which the metallized tape is bonded to specially prepared electronic chips; (2) encapsulation, during which the chips are encapsulated on the active, bonded surface or on both surfaces; (3) singulation, an optional step in which the packaged chips are separated and placed in an intermediate carrier to facilitate testing and/or burn-in; and (4) Outer Lead Bonding (OLB), during which the packages are excised from the continuous tape or carrier, placed on a printed wiring substrate, and bonded, usually with solder. Steps (1), (2) and (4) are shown schematically in Figures 6-39, 6-40 and 6-41. Numerous variations on these procedures have been used commercially; the choices and some of the underlying factors are discussed in this section.

6.5 TAPE AUTOMATED BONDING

Figure 6-39. TAB Inner Lead Bonding (ILB).

6.5.2 TAB Tape
The tape used for TAB is prepared in many forms, involving different materials, widths (ranging from a few mm to 70 mm or more), plated surfaces, and geometries. The three basic variations are one-, two- and three-layer tape, shown in Figure 6-42.

One-Layer Tape: One-layer tape is made of etched metal, normally copper, having a conventional thickness of about 70 μm. Because the leads are shorted, bonded devices are not testable on tape (after wafer testing) until excised from the tape. Furthermore, because the tape metallization must be relatively thin to accommodate the fine-dimensional (approximately 50 to 250 μm) lithography, the length of unsupported lines is limited by their fragility. This places a practical upper limit on the lead length between ILB and OLB points. The actual number of I/Os that one-layer tape may allow is usually determined by the OLB pitch, since this determines the OLB footprint size and thus the lead length between OLB and ILB.

Two-Layer Tape: Two-layer tape consists of a polymer film onto which is plated a patterned metal layer, usually copper, having a thickness of about 20 to 40 μm. This structure and the three-layer structure, described subsequently, can support isolated leads, allowing device testing on tape. TAB tape may be fabricated from a variety of materials, including polyimide, polyester, and polyimide acid [114]. (Epoxy/ fiberglass and other materials are also being used for three-layer tape, which does not have to be patterned by etching the polymer.) An important aspect of two-layer tape preparation is the adhesion between polymer and metallization. A common approach for promoting adhesion of metal to polymer is the sequential sputtering of thin layers of chromium and copper (on the order of 1 μm) onto the
Figure 6-40. TAB Encapsulation. A single-orifice nozzle dispenses encapsulant on the bonded chip.

Figure 6-41. TAB Outer Lead Bonding (OLB). After the TAB package is excised from the tape and placed on the substrate or board, it is normally solder bonded.

polymer film. This adherent metallization provides a plating base for thick copper deposition. Once the metal pattern is defined, the polymer material is photolithographically patterned and etched to provide sprocket holes and bonding windows for both ILB and OLB leads. Typical metal features may be defined to widths and spacings of 50 μm (or slightly less for thin metal layers); typical polymer features are normally much coarser, being on the order of several hundred microns. An alternative approach for two-layer TAB tape, which has been commercialized by 3M, is the spray deposition of a polymer (polyimide) onto unpatterned copper. Both the copper and polymer are subsequently patterned by etching. A limitation associated with this approach is the cure-shrinkage of the polymer, which produces stress-induced warpage of the structure; this places a limit on the practical size of the tape patterns.
Three-Layer Tape: Three-layer tape consists of blanket metallization, usually in the form of sheets of electro-deposited or rolled copper, which are laminated using an adhesive to prepunched polymer film. This foil may be etched or surface-plated under conditions that roughen the surface, thereby enhancing the adhesion. Following lamination the adhesive is normally cured. The lead pattern is then photolithographically defined and etched. One key difference between three-layer and two-layer tape is the punching rather than etching of the polymer tape. The tape punching is carried out with a series of metal dies, which for a new tape design requires slightly greater setup time and expense than is associated with two-layer tape, which requires only a new mask set. One other distinction between two-layer and three-layer tape is that a "tie-bar" must normally be used to connect each polymer area on the three-layer tape. This tie-bar must be cut or removed at some point in the TAB assembly process. Many kinds of adhesives have been used in three-layer tape, including polyimides, epoxies, acrylics, and phenolic-butyrals. Because TAB packages must normally withstand some elevated temperature processing during ILB, encapsulation-curing, burn-in, and OLB, the choice of adhesive for three-layer tape or the adhesion layer for two-layer tape may be primarily based on thermal stability. Table 6-6 shows a process flowchart of the three basic procedures for tape fabrication. Note the metal etching required for three-layer tape; this places a slight limitation on the refinement of the metal pattern, compared to two-layer tape, which is pattern plated. Tape is commonly fabricated in widths having multiple patterns, thereby requiring a slitting operation to produce finished tape. At this point, no accepted standard tape widths (apart from 35 mm) have been established. However, recent proposals have been made to standardize on 35, 48, and 70 mm, using "super-35 mm" rather than the standard 35 mm sprocket holes [115]. Figures 6-43 and 6-44 show examples of TAB tape having a low and high number of I/Os.

All three types of tape are normally plated to optimize the ILB or OLB operations (or to yield a compromised optimization of the two), as well as to provide required shelf life and corrosion resistance. Relevant issues regarding leadframe solderability and reactions during soldering are described in Chapter 11, "Package-to-Board Interconnections." ILB materials issues are discussed in the next section. The conventional choices of plated layers on TAB tape are immersion-plated or electroplated gold or tin. The former process, involving an exchange reaction in which gold or tin is substituted for copper at the copper surface, is a self-limiting reaction resulting in a relatively thin (on the order of 1000 Å) layer which may be porous. The suitability of immersion-plated tin (or for that matter, electroplated tin) may be limited by the potential for tin-whisker formation after plating from improperly controlled plating baths [116, 117, 118]. An approach used by some tape suppliers to reduce the potential lead-shorting associated with tin-whisker growth is the application of a screen-printed coating (usually epoxy) over most of the leads prior to Sn plating. The plated leads near the ILB or OLB points are refloved during bonding, thereby removing whiskers or the propensity for whisker formation in these areas. Another whisker elimination procedure is an annealing treatment of the plated tape; this allows relaxation of the internal stresses that cause whiskers. An alloy of Pb-Sn may also be electroplated onto the tape metallization without concern for whisker formation. One extra plating step designed to improve bonding reliability is an electroplated nickel layer between the copper leads and gold surface layer to prevent interdiffusion. The use of nickel as a diffusion barrier has been studied by several groups [119, 120, 121, 122] with the conclusion that the nickel layer is unnecessary for reliable inner and outer lead bonding [120, 123]. It may also be used, however, to maintain the purity of a final plating bath that is prone toward copper dissolution. Electroplating of tape metallization requires that all leads be shorted to contact points. This affects the usable area on the tape and may require, where it is desirable to test ILB-bonded devices prior to OLB, a special cutting or punching operation to delete common con-

---

Table 6-6. TAB Tape Fabrication Steps.

<table>
<thead>
<tr>
<th>Layer</th>
<th>All Metal</th>
<th>Additive Process</th>
<th>Subtractive Process</th>
</tr>
</thead>
<tbody>
<tr>
<td>1.</td>
<td>Slit metal foil</td>
<td>1. Deposit metal adhesion layer/common electrode</td>
<td>1. Punch adhesive-coated polymer</td>
</tr>
<tr>
<td>2.</td>
<td>Apply photoresist</td>
<td>2. Apply photoresist (2 sides)</td>
<td>2. Laminate metal foil</td>
</tr>
<tr>
<td>5.</td>
<td>Etch metal pattern</td>
<td>5. Pattern plate</td>
<td>5. Expose</td>
</tr>
<tr>
<td>7.</td>
<td>Clean</td>
<td>7. Strip photoresist</td>
<td>7. Coat back side for lead protection</td>
</tr>
<tr>
<td>10.</td>
<td>Surface plate</td>
<td>10. Surface plate</td>
<td>11. Surface r-</td>
</tr>
</tbody>
</table>
Figure 6-43. Low I/O TAB. (a) A 28-lead three-layer tape. Photograph courtesy of Shindo Denshi Ltd. (b) A 16-lead one-layer tape. Photograph courtesy of 3M Electronic Products Division. (c) A 40-lead three-layer tape. Photograph courtesy of Mesa Technology.

Figure 6-44. High I/O TAB. (a) A 328-lead two-layer tape. Photograph courtesy of 3M Electronic Products Division. (b) A 204-lead three-layer tape. Photograph courtesy of Mesa Technology. (c) A 308-lead three-layer tape. Photograph courtesy of Shindo Denshi, Ltd.
nections (see reference [124]). Examples of several tape formats are shown in Figure 6-45.

Two primary options are available for the structure that allows chip and tape to be bonded together. These two structures are referred to as Bumped-Chip and Bumped-Tape and are shown schematically in Figure 6-46. The bonding pedestals are needed in either case to isolate the thermocompression bonding forces to the point where interconnection is to occur. Furthermore, since most electronic devices have a passivation layer of $SiO_2$, $Si_3N_4$ or polyimide on top of the active silicon, an unprocessed interconnection pad on the device would be recessed and would have limited accessibility to a planar bonding lead. Considerable attention has been paid to the intervening metallurgical layers between the device interconnection pad (normally aluminum, Al-Si or Al-Cu) and bump; this is discussed in detail in the next section. One well-recognized advantage of bumped tape is the potential elimination of any chip processing beyond the opening of contact holes through the passivation layer. That is, the bumped tape may be directly bonded to the aluminum pad; this is also discussed subsequently. The formation of bumps on the tape is relatively straightforward for single-layer tape and involves double-sided etching of a thicker copper structure (about 70 $\mu$m); an example is shown in Figure 6-47. Formation of bumped two- or three-layer tape is considerably more difficult but has been demonstrated using conventional tape processing techniques [126, 127]; however, this approach is not, at present, commercially viable. Conventional tape processes for making bumped tape, which involve double-lithographic processing, suffer from...
Figure 6-47. Bumped One-Layer TAB Tape. The flat-topped bumps at the end of the leads are bonded to the chip I/O pads. Photograph courtesy of Mesa Technology.

sional stability limits associated with the carrier tape material; re-registration may become the limiting process. A recently commercialized approach is a bump-transfer process, developed by Matsushita [125], in which standard TAB tape leads are bonded to prepatterned gold bumps. These bumps are lifted from the substrate on which they are prepared and used as bumps for subsequent bonding to a silicon chip, (see Fig. 6-48). Other processes have been cited in the patent literature [128, 129].

Most TAB packages are designed for peripheral interconnections to the active device. To increase the number of potential chip interconnections and to improve silicon design flexibility, an area array concept for TAB (ATAB) is being pursued [130, 131]. This allows internal I/Os rather than the peripheral lead attachment commonly associated with TAB. The first examples of this approach employ solder interconnections, similar to C4 joints rather than T/C bonded joints. The tape for ATAB requires two metal layers, giving the geometric limitation that exists for single-metal-layer tape (see Fig. 6-49); this is the limit in fanning out the signal lines from a dense area array to the peripheral leads extending to the board. The introduction of two-metal-layer tape has been delayed by the added cost and complexity of preparing interlevel vias, as well as of the additional metal layer.

Several arrangements are used in mounting the TAB packages to the printed-circuit substrate (Fig. 6-50). These include chip-up or chip-down orientations, back bonding to promote heat transfer to the substrate, or heat-sinking for natural- or forced-convection cooling. One reason for using both chip-up and chip-down configurations, as demonstrated [132], is to avoid the requirement of a mirror-image silicon device or an extra wiring level on the second-level package when attempting to interconnect several TAB-packaged devices to second-level substrates that have very regular wiring patterns. This includes display panels or memory cards. Component and computer manufacturers have used TAB tape as an intermediate packaging element to connect the device to a leadframe for plastic packages [133]. Other applications include interconnections for hybrid modules [134], display panels [132], so-called smart cards (credit card structures containing silicon elec-
tronics), and multilayered ceramic modules for high-performance computers (see Fig. 6-51) [135, 136]. Gang-bonding associated with TAB is faster than wirebonding, especially for moderate-to-high I/O devices. It has also been argued to be a more reliable method of making a large number of interconnections to the device [100]. A TAB package, like an MLC substrate, is not limited to a single chip but may be used to package several chips.

Once the ILB process has been completed, the exposed surface of the device is normally encapsulated (Fig. 6-52) with one of many commercially available materials such as filled epoxies (containing low-expansion fillers), silicones, or flexibilized epoxies [137, 138, 139]; for current encapsulation materials and methodologies, see Chapter 10, “Package Sealing and Encapsulation,” and Chapter 12, “Printed-Circuit Board Packaging.” Along with the active surface, the entire device may be encapsulated. This practice re-
duces chip bending stresses, arising from encapsulant cure shrinkage plus differential thermal expansion between chip and encapsulant upon cooling, and may help to limit corrosion arising from chip edge exposure. However, this full encapsulation requires slightly more complicated processing. Once the encapsulant is applied, the tape is cycled through a curing oven to drive off encapsulant solvents or to effect cross-linking reactions.

After encapsulation, parts are often tested and subjected to a burn-in procedure to eliminate those with incipient defects. This may present a considerable advantage in hybrid yield with respect to interconnection approaches that do not allow pretesting or burn-in prior to bonding [140]. Under some circumstances a simple dc voltage bias, applied to the part during a many-hour exposure to elevated temperature, is a sufficient burn-in test. However, where device testing using active signals is warranted during burn-in, it may be most expedient to separate TAB frames into individual frame carriers, similar to photographic slides, to allow active testing [141].

6.5.3 Inner Lead Bonding (ILB)

TAB Inner Lead Bonding (ILB) normally consists of a thermocompression bonding step in which a hot (300° to 600°C) thermode applies pressure between the TAB leads and the chip interconnection structure in a single operation. The objective is a low-resistance contact to the active circuits, with the contact joints having high mechanical strength, usually at or exceeding the pull strength of the TAB leads. The choice of bonding conditions is a compromise between the high temperature, pressure, and bonding time required for improved joint reliability and the need to limit the exposure of the chip to these conditions to avoid chip degradation. Manufacturing throughout is another factor that influences the cycle time of the bonding operation and may support the application of higher temperatures and shorter bonding and thermal ramp times, thereby exposing the chip to greater levels of thermal shock.

**Interface Metallurgy (IFM):** For the situation in which planar tape is bonded to a bumped chip, the aluminum or Au-alloy chip interconnection pad is normally coated with several layers of metal to provide low contact resistance and adhesion to the pad. The top layer of the interface metallurgy (IFM) is normally a noble metal, to provide an inert surface for bonding or plating. For bumped tape this triplate metallization may also be applied to the active chip, although in this case the IFM would be the bonding surface rather than a plating base. Table 6-7 gives a listing of many of the IFM structures used in TAB processing. These structures generally include an adhesion layer (titanium or chromium), a diffusion barrier layer (copper, nickel, tungsten, palladium, or platinum), and a bonding layer (usually gold). The adhesion layer is chosen to make good electrical contact to the aluminum pad as well as to adhere to the passivation layer surrounding the pad. This provides improved corrosion resistance compared to alternative bonding structures in which aluminum is exposed [142] and, along with the chip passivation, provides a nearly hermetic chip structure [143] (Fig. 6-53). The barrier layer prevents unwanted interdiffusion between the several metals of the bonding structure.

The interface metallurgy may be evaporated or sputtered; and when IFM is used as a plating base for bump plating, this is done as a blanket deposition. When bumped tape is bonded to an IFM structure, the IFM may be patterned during deposition using metal masks or lift-off techniques, or it may be patterned by photolithography and etching of a blanket deposition. Extensive studies on thin film metal deposition have been performed to evaluate aspects of metal adhesion [145], film stress [146], and other parameters that may affect the IFM [147, 148]. Industrial experience has shown not only that a variety of IFM structures provide for successful T/C interconnection, but also that the choice of cleaning, deposition conditions, or metallurgies for the IFM may lead to unacceptable interconnections, as referenced below.

Once the IFM is deposited on the devices, a thick photoresist (either a wet film or, as is most common, a dry film) is applied to the wafer and patterned. At this point a bump is formed by plating a thick (approximately 10 to 30 µm) layer of gold or, to minimize the amount of gold used, a duplex copper/gold structure [149, 150]. The photoresist is then stripped away and the underlying plating base may be etched or sputtered away (Fig. 6-54). For T/C bonding in which the bumps may be deformed at low bonding stresses, it is desirable that the electroplated bumps be made of soft gold [150]. However, electrodeposited gold may be quite hard, because of grain refinement and possible other factors [151]; e.g., nominally pure gold bumps may have a Vickers Hardness exceeding 120 [152]. Because the as-plated bump hardness can be so high and can lead to “Si-cratering” during T/C bonding, the bumped wafers may be annealed prior to T/C bonding [153]. The sputter deposition or evaporation of the bump structure [147] has now become commercialized by Siemens, using a blanket copper evaporation plus etch process. These relatively hard bumps are bonded using Sn reflow rather than thermocompression bonding.

**ILB Process:** The ILB process utilizes a heated thermode that may be a solid block or, for bonding of peripheral leads only, a bladed tool (Fig. 6-55). The primary difference between the two is the thermal mass of the thermode; this affects the programmability of the bonding temperature cycle and, as well, the thermal shock imposed on unheated chips. For example, to minimize the thermal shock to devices being bonded with a solid thermode, which is normally kept at a fixed bonding temperature, the chips are often preheated before bonding [154]. The temperature ramp experienced by the chip under such bonding conditions may well exceed 10° K/s. A bladed
thermode may be heated and cooled during a bond cycle, thereby allowing more process flexibility. A deficiency in a bladed thermode is the mechanical fragility compared to the solid tool. In both cases the loading cycle may be programmed through the bond cycle. Several studies have been performed to optimize the temperature/pressure/time cycle during T/C bonding for a given interconnection structure [116, 149, 156]. Figures 6-56 and 6-57 show the high degree of sensitivity of bond strength to temperature and dwell time during bonding. The temperature affects the kinetics of the metallurgical reactions, contaminant desorption or breakdown, and diffusion-dependent

<table>
<thead>
<tr>
<th>Company</th>
<th>Adhesion Metal</th>
<th>Barrier Metal</th>
<th>Bump Metal</th>
<th>Plating Metal</th>
<th>Lead</th>
</tr>
</thead>
<tbody>
<tr>
<td>GE</td>
<td>Cr</td>
<td>Cu</td>
<td>Au</td>
<td>Sn</td>
<td>Cu</td>
</tr>
<tr>
<td>CH Honeywell</td>
<td>—</td>
<td>—</td>
<td>Au</td>
<td>Sn</td>
<td>Cu</td>
</tr>
<tr>
<td>Bull</td>
<td></td>
<td></td>
<td>Au</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Sharp</td>
<td>Cr</td>
<td>Cu</td>
<td>Au</td>
<td>Sn</td>
<td>Cu</td>
</tr>
<tr>
<td>Honeywell</td>
<td>Ti</td>
<td>Pd</td>
<td>Au</td>
<td>Sa</td>
<td>Cu</td>
</tr>
<tr>
<td>Fairchild</td>
<td>—</td>
<td>—</td>
<td>Au</td>
<td>Sa</td>
<td>Cu</td>
</tr>
<tr>
<td>Signetics</td>
<td>Ti</td>
<td>Pt</td>
<td>Au</td>
<td>Sn</td>
<td>Cu</td>
</tr>
<tr>
<td>Nippon Electric</td>
<td>Ti</td>
<td>Pt</td>
<td>Au</td>
<td>Au</td>
<td>Cu</td>
</tr>
<tr>
<td>National</td>
<td>Cr</td>
<td>Cu</td>
<td>Au/Cu</td>
<td>Au (Thin)</td>
<td>Cu</td>
</tr>
<tr>
<td>Semiconductor</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>BTI</td>
<td>Ti</td>
<td>Pt</td>
<td>Au</td>
<td>Au</td>
<td>Cu</td>
</tr>
<tr>
<td>Philips</td>
<td>Al</td>
<td>Ni</td>
<td>Solder/Cu</td>
<td>Au/Al</td>
<td>Cu</td>
</tr>
<tr>
<td>Siemens</td>
<td>—</td>
<td>—</td>
<td>Au</td>
<td>Sn</td>
<td>Cu</td>
</tr>
<tr>
<td>RCA</td>
<td>Ti</td>
<td>Pt</td>
<td>Au</td>
<td>Sn</td>
<td>Cu</td>
</tr>
<tr>
<td>Motorola</td>
<td>—</td>
<td>—</td>
<td>Au</td>
<td>Sn</td>
<td>Cu</td>
</tr>
<tr>
<td>TI</td>
<td>—</td>
<td>—</td>
<td>Au</td>
<td>Sn</td>
<td>Cu</td>
</tr>
</tbody>
</table>


Figure 6-53. Improved Chip Hermeticity with TAB Bump Structure. A potential corrosion-resistance advantage may accompany the full coverage of the aluminum bonding pad.

mass transport mechanisms, as well as the plastic flow strength of the deforming members. Local pressure is also an important parameter in T/C bonding; higher pressure tends to shift the bond pull-strength curves, shown in Figure 6-56, to lower temperatures [157, 158]. One guideline that is generally obeyed for optimum T/C bonding is the condition that both deforming elements being joined have similar flow strengths [95].

Figure 6-54. Gold Bump Processing. After Oswald and deMira 1977. Reprinted with permission of Solid State Technology.
Thermocompression bonding is a relatively complex phenomenon that requires an understanding of (1) the gross mechanics of the deforming members; (2) the operative deformation-welding mechanisms at the bonding surfaces at the given stress and temperature, and (3) the effect of surface roughness and surface contaminants. These must be balanced against tooling limitations and constraints imposed throughout by the manufacturing conditions. Regarding the mechanics of T/C bonding, it has been observed experimentally that the bonding stresses, deformation, and the degree of bonding are not homogeneous across the bonding surface [159, 160]. The general problem of localized loading has been considered analytically [159, 160, 161], and the results may be interpreted for the case of thermocompression bonding in which the triaxial constraints limit the amount of local plastic shear deformation at the center of the bonding pads, which limits the effective bonding area. This has also been observed in wirebonding [88]. The bonding area increases with total plastic strain [160], consistent with reported observation [119] that greater than about 30% deformation is required for effective T/C bonding. One notable finding to the contrary is the work [162] in which the plastic deformation for gold-plated copper bumped tape bonded to a bare aluminum pad was minimized for optimum bonding.

The effect of surface roughness on the dominant deformation mechanism (plasticity, dislocation creep, or diffusional creep) for Cu-to-Cu T/C bonding have been reported [163, 164, 165, 166]. The bonding effectiveness was given in terms of fractional area bonded, with the dominant mechanism deduced from interfacial void shape. Although the bonding times evaluated (10 to 10^5 s) were considerably longer than those associated with TAB-ILB (0.1 to 5 s), and bonding stresses studied (7 to 35 MPa) were lower than ILB stresses (35 to 200 MPa), these studies provide a framework for analyzing the operative bonding mechanisms for T/C bonding for TAB interconnections.

A study by Spencer [155] on T/C bonding of gold-plated copper shows the important effect of the total deformation on the time-at-temperature required for effective bonding (Fig. 6-57). In this work the kinetics of bonding to a given strength level match those of gold surface diffusion.

Surface contamination in Au-to-Au [157, 167] and Cu-to-Cu [95] bonding has been studied. Although the mechanisms by which the variety of
Figure 6.57. Interdependence of Dwell Time, Temperature, and Deformation on Bond Strength. Arrhenius relationship between the dwell time of the bond and the interface temperature at several deformation levels. Data from plots of bond strength vs. dwell time, at constant strain as shown in part (a), were used to study the thermally activated process responsible for the bonding process, shown in part (b). Specimens of gold-plated copper were tested in shear. After Spenser, Ref. [155], 1982.

controlled contaminants and cleaning conditions affect the T/C bonding are not clear, the presence of organic films on the bonding surface is significant (Fig. 6.58). When copper is present at the bonding surface, bond quality is adversely affected by the presence of an oxide [95]; oxygen per se [168] and water vapor [153] may also adversely affect bondability through oxidation.

Another common form of ILB bonding involves the presence of a liquid phase, unlike the solid-state processes described previously. One approach is to use a high lead solder bump on the chip as used in the C4 [113, 169]. Another technique is to form a low-melting alloy between gold and tin [147, 116]. Usually this is accomplished by bonding tin-plated leads to gold bumps, although bonding of a gold-plated lead to a tin-capped bump has also been reported [116]. This so-called eutectic bonding is an effective approach for low-stress, low-temperature bonding, with general applicability to TAB packaging (Fig. 6.56). It is observed that the wettability and thus bondability in this ILB system is improved when bonding is carried out in an inert atmosphere [116].

Several studies have been published regarding the performance of T/C bonds subjected to annealing, thermal cycling, or vibrations [119, 121, 122, 170, 171]. For particular structures some joint strength degradation was observed [119, 120, 170], which was believed to be associated with the evolution, at rather elevated temperatures, of Kirkendall voids at the Cu-Au interface. For Au–Au T/C solidus bonds, postannealing improved bond strength [167]. In other TAB ILB studies no significant effect of aging was seen [122, 126]. No degradation was observed in the study of thermally cycled Au–Sn TAB-ILB joints. Aging before bonding was extensively studied for gold-plated copper leads [119, 123, 171] and a shelf life/temperature relation was developed for acceptable bonding. Very limited data has been published for temperature/humidity/voltage bias corrosion testing of T C

...
some of these data serve to point out the need for proper fabrication procedures. Experiments have been performed to evaluate acoustic emission techniques for ILB monitoring or pull-testing failure analysis [173, 174]; these experiments have not demonstrated any particular advantage at this point.

**TAB Tooling:** Commercial tooling is available from several sources for TAB-ILB. These range from high-speed devices, requiring minimal chip and tape inspection, to slower tools capable of full, automatic inspection and alignment. The choice is determined by the silicon and tape cost differential (an expensive chip may be most cost-effectively bonded to an inexpensive TAB tape only when the tape frame is fully inspected) or by the accuracy of the tape sprocket-hole registration and chip size with respect to ILB alignment requirements (sprocket-hole tolerances or chip-dicing tolerances may require vision-based, chip-to-tape alignment for fine ILB dimensions). One critical element in the performance of an ILB tool is the planarity between the chip-bonding platform and the thermode. Given the limited vertical displacement of the thermode during bonding, a lack of coplanarity between the chip and the thermode may result in only partially bonded chips or excessive, localized bonding loads that result in chip fracture. Another focal point in the ILB process is the thermode. Studies have shown that the temperature measurement technique and temperature uniformity are particularly important to reliable bonding [149, 175, 176]. Heat transfer between thermode and silicon must be optimized to minimize thermal shock [162] for example, by delaying full-load application to reduce heat transfer and to reach the desired bonding temperature in a reasonable dwell time [176]. Preheating of the chip may be beneficial for ILB but may accelerate undesirable interdiffusion of the IFM or, for bonding to bare aluminum, may accelerate pad oxidation [162]. Some of these issues and other manufacturing factors are addressed [144, 177, 178]. One issue for which little information is available, but which has an important impact on ILB tooling, is the thermode dressing and cleaning that are periodically required.

### 6.5.4 Outer Lead Bonding (OLB)

Several aspects of interconnecting a TAB-packaged chip to the next level of packaging differ from practices used with more established package types, which are described in Chapter 11, “Package-to-Board Interconnections.” TAB packages may be joined to a variety of substrates ranging from leadframes for plastic packages (pin-through-hole or surface mount) to high-performance elements on a computer module. Intermediate TAB structures are used for isolated, high I/O devices on printed-circuit boards or densely spaced packages on a carrier structure, such as a memory card or display unit.

---

**Figure 6-59.** Down Setting. Prevent Inner Lead Shorting to Chip.

After inner lead bonding and encapsulation, each TAB frame constitutes a package ready for testing, burn-in, and final-stage assembly. Prior to OLB it may be necessary to remove the common connections (required for electroplating two- or three-layer tape) by punching or cutting; it may also be necessary to singulate each frame into a carrier to allow dynamic testing (see section 6.5.2, “TAB Tape,” on page 411). The first step in the OLB procedure is the excising or cutting of the TAB package from the tape carrier using a metal die. Then the leads are normally formed into “gullwing” shapes (Fig. 6-59) to provide mechanical compliance. This can be accomplished along with the excising operation in a single die. At this point the package is rather fragile, given that the leads are thin (approximately 20 to 60 μm vs. 100 to 250 μm for other surface-mounted packages). To achieve reasonable manufacturing yields, it is therefore necessary to place or bond the TAB package without further handling. Some applications allow the use of TAB without lead forming [136, 179], although this is not without some risk of fatigue failure.

Although the fragility of the excised package does place constraints on the board attachment, it also provides an advantage with respect to stiffer lead frames. This advantage is realized under conditions that cause stress on the board interconnection, such as those arising from differential thermal expansion or board bending (which occurs during handling or plug-in operations). As the stress approaches or exceeds the yield strength of the lead material, fatigue failure begins to become a consideration, but for the typically compliant TAB structures addressed here, this is normally not an issue. Furthermore, the stresses applied by the thermally stressed lead to the typical solder joint are also reduced with respect to the stresses transmitted by stiffer leads, e.g., from PLCCs or other surface-mounted packages. Thus, thermal fatigue of the solder interconnection is reduced for the lead thickness associated with TAB.

The outer lead bonding of TAB packages is normally accomplished using heated blade structures. The heated blade forces the TAB leads against
the bonding pads, which are usually coated with solder or solder plus flux. The joint is usually allowed to solidify before the bonding thermode is removed. Several other solder-bonding techniques may also be used for TAB OLB (see Chapter 11, "Package-to-Board Interconnections"), such as vapor phase reflow [149] or hot-air techniques. However, the fine lead pitch and lack of lead rigidity may require alteration of processes used for conventional SMT packages. Thermocompression bonding may also be executed for OLB [135, 136], mostly for joining to ceramic substrates. Some data on OLB pull strength and reliability is given in reference [171].

6.5.5 Design Factors

Many factors not previously addressed may affect the TAB package configuration. An increasingly important factor associated with VLSI packaging is heat-dissipation capability. A distinction with TAB, vis-à-vis other silicon interconnection approaches, is the relative freedom with regard to the approach taken for thermal dissipation (see for example, Fig. 6-50). Both faces of the chip are accessible to heat sinks, although an encapsulation layer may impose a lower limit on the thermal resistance when the heat sink is attached to the active chip face (on the order of 3 to 10 K/W for conventional encapsulant materials). Back bonding the chip to a substrate or heat sink provides a low thermal resistance path between chip and ambient. The leads themselves are either thermally conductive and, depending on the environment and the other thermal dissipation paths, can play an important role in conducting heat from the chip to the substrate. The thermal resistance between the chip and leads in TAB packages is lower than that in wirebonded packages. This arises because the ILB points used in TAB structures are constructed of high-conductivity metals, having cross sections larger than wires used for wirebonding. The typical value of the thermal resistance from chip to air of a TAB package soldered to a glass-epoxy board in natural convection is about 57 K/W [180]. That is, if the ambient air and the maximum chip temperatures are 25°C and 85°C, respectively, then the package will dissipate slightly more than 1 watt. Other enhancements, such as forced-air convection, back bonding plus higher-thermal-conductivity substrates, or heat sinks, can greatly increase this power dissipation level.

Electrical performance requirements will affect the TAB configuration. For example, an increase in the number of power and ground leads to the chip will decrease the effective package inductance, which, for VLSI chips, will allow better circuit utilization (see Chapter 3, "Package Electrical Design"). This may decrease the allowable ILB and OLB spacing and, for the latter, may require wider tape. The OLB spacing is influenced by the desire to limit the package lead length, especially for high-inductance paths, and the desire to reduce board area. This tends to promote refinement of the OLB spacing. On the other hand, the wireability of the board or substrate to which the package is attached must also be considered. It is often desirable to have a coarse OLB pattern, simply to be able to fan out the signal lines from a high I/O chip to minimize wire-flow congestion problems beneath the package. The above trade-off will be resolved differently for different systems. Standards for OLB width and spacing have not been fixed but should be established soon.

Mechanical factors also have an effect on the TAB process and configuration. An example is shown in Figure 6-59 of a common practice known as "down setting," in which the leads are bent away from the chip surface to avoid shorting to the chip edge. Careful handling is necessary at any point during TAB processing, sometimes even after board assembly; for example, at least one manufacturer chooses to protect the TAB package after the OLB step by applying a second, silicone encapsulation layer. Specialized tools are necessary for splicing, and specialized spacer tape is often used for bonded chips on tape.

6.5.6 Future Trends

It is expected that the breadth of application of TAB will increase, especially as its reliability becomes better established. Two examples of current application for very high performance involve the use of TAB for supercomputers of the NEC Corporation and the ETA Corporation. TAB packages are intrinsically lower in cost than certain alternatives, such as ceramic packages. TAB has an advantage with respect to many wirebonded packages in the areas of electrical and thermal performance. The manufacturing automatability and the compatibility of the package with finer second-level structures give further impetus for its development. It is expected that as the level of integration increases, along with the number of I/Os and level of heat dissipation, TAB packages will evolve to have finer ILB dimensions (less than 50 μm width and spacing), finer OLB dimensions (less than 100 μm width and spacing, as the substrate technology allows), improved electrical performance, especially as two-metal-layer tape is employed [181] (effective package inductance less than a few nanohenrys), and improved thermal performance (up to 5 to 10 W per chip heat dissipation).

6.6 PRESSURE CONNECTS

Pressure connects are not actually bonded connections but maintain contact only by an external, continuously applied force (such as a metal spring or elastomer retainer). Several types have been described. Citizen Watch [182], for example, joins Au bumped chips onto Au bumped substrates using conductive rubber contacts embedded in a polyimide carrier (Fig. 6-60).

Typical loading is 2 g per pad to provide adequate deformation of the rubber for needed contact on all pads; thus a 300-pad chip would require a total loading of 600 g...
6.7 OPTOELECTRONIC INTERCONNECTIONS

Continuous loading. The conductive rubber is not a "good" conductor: a 200 μm x 200 μm pad has 35 mΩ resistance. Thus, it has been applied in special applications (such as liquid-crystal display panels) that can tolerate the high resistivity.

Significantly lower resistances per contact, of the order of 10 mΩ, have been achieved by using an all-metal system. However, this requires a "flexible" substrate to which the chip is joined. Tektronix [183] has joined GaAs chips to a flexible circuit, as shown in Figure 6-61. The bumps are plated on the thin-film flexible circuit. Two hundred pads per chip have been fabricated with 50 μm pads on 100 μm centers.

Silicon Connection has a very similar approach called "SICONS" [182]. Applications include ROMS in ATARI computer games. A recent chip-connection approach using the cure shrinkage of a polymer to hold two finely bumped surfaces in contact has been recently described by Matsushita [184].

Although such applications are limited at present, there are several reasons for the increased research in these areas. This type of contact eliminates the chip-size restrictions of the C4 type of full-bonded connection; there is no fatigue mechanism. However, this is replaced by the concern of a sliding or point contact developing high resistance by debris or oxidation/corrosion. The second reason for research activity is the ease of assembly and reworking — no connections need to be "broken" or unbonded in order to take the package apart.

Figure 6-61. GaAs Chips Joined to a Flexible Circuit. (a) IC flex circuit elastomer pressure pad sandwich cross section. (b) I/O pressure bumps establish firm elastomer contact. After Smith, Ref. [183], 1985.

Optical photons have little or no interaction with other photons and have little "resistance" in optical media (glass, air, etc.). Optical signals can be sent with much lower power, higher speed, and higher density than can electrical signals. An example of relative data transmission rates was given by IBM [185]: a new amplifier chip, or data receiver, in a controller used a fiber optics transmission line to send data back and forth to the computer. It was able to operate at 400 million bits per second, compared to 25 million bits per second for comparable copper transmission lines: "The text of a 20 volume encyclopedia can be received in under three seconds!"

Optical devices such as Light Emitting Diodes (LEDs) and laser diodes are used as optical signal generators or receivers to convert optical to electrical signals and vice versa. These devices are linked to fiber optics, which are the most common medium for long-distance communication. Figure 6-62 shows three different schemes for getting the fiber optic signal into or out of an optical device. Figure 6-62a utilizes free space as the connection medium [186]; Figure 6-62b utilizes a butt joint directly to the top of the LED [187]; and Figure 6-62c utilizes a thin-film wave guide to funnel down the signal to the fine device dimensions [18]. Note the chip mounting could be face up, flip
chip or side mounted. In the flip-chip example, very fine (4 μm high) C4 joints made of pure In are used to make electrical connections to the same side of the chip as the optical connection. Wirebonds are illustrated in Figure 6-63, in an alternative configuration [188].

Alignment of the optical fiber to the device is critical. It may be noted that the integrated circuit technology developed initially for silicon is having significant application for optical devices. The self-alignment capability of the C4 jointed to great advantage in the Coplanar Waveguide GaAs device (Fig. 6-62c). The precision etch capability of silicon is being used to make grooves for aligning and holding the optical fibers shown in Figure 6-64. An integrated fiber optic transmitter, shown in Figure 6-65, utilizes a semiconductor laser array chip, a cylindrical lens, and an array of optical fiber light guides, all joined to a single silicon substrate, which is wirebonded.

Once on chip, there are numerous optoelectronic structures to couple, guide, modulate, detect, and emit the laser signals. Although a detailed discussion is not within the scope of this chapter, a schematic of these typical building blocks is shown in Figure 6-66.

It is currently a challenge to join optical devices, optoelectronic devices, and straight LSI or VLSI devices and to interconnect these into a compatible package assembly. Only relatively simple assemblies with relatively simple functions, have been developed to date, yet the interconnection requirements are becoming progressively more sophisticated. The different materials involved, tolerances required, assembly hierarchy, thermal expansion tolerances, and other material properties and interactions combine to make this field broad and to receive increasing attention.
6.8 ELECTRICAL PARAMETERS OF INTERCONNECTIONS

The electrical performance of the packaged chip may well be determined by the mode of chip interconnection. This arises not from the intrinsic properties of the interconnection but from the package geometry which is, in part, im-
posed by the chip-level interconnection approach. For example, a wirebonded silicon device is normally associated with a leadframe, which, for fabrication reasons or board-level interconnection geometry requirements, may require rather long leads, well separated from a ground or reference plane. (Alternative wirebonded structures, such as pin-grid arrays or direct-chip attach to a substrate, may provide a much different electrical signal environment.) High-end computer packaging designers are aware of the electrical performance issues associated with packaging, usually for their own customized applications (see Chapter 1 and Chapter 16). For this reason, as well as because low-performance systems are not so often package-limited, focus on mid-range structures are frequently gated by package configuration and performance. To be considered are chips having 84 and 180 I/Os packaged using each of the three primary chip interconnection approaches addressed in this chapter: C4 solder-joined, wirebonded and TAB-bonded packages.

The packages to be compared are: (1) a C4-joined, metallized ceramic single-chip carrier having swaged pins; (2) a C4-joined, multilayer ceramic, single-chip carrier having brazed pins; (3) a wire-bonded pin-grid array; (4) a wire-bonded plastic-lead chip carrier (PLCC) (84 I/O only); and (5) a TAB-packaged device. The pin grid for (1), (2), and (3) is fixed at 2.54 mm; the PLCC is fixed at 1.27 mm lead pitch; the TAB package is fixed at 0.5 mm lead pitch. (Note, package 2 is a somewhat artificial configuration, given that usually more than one C4-joined chip is joined to an MLC module.) The objective here is to give a rough basis for comparing package performance, using the lead geometry of the packages. The basis for comparison is the calculated range in the value of lead inductance. This provides a starting point for estimating the effective package inductance, \( L_{eff} \). This quantity is used to predict the inductance-based voltage swing (or \( \Delta I \) noise), \( \Delta V = L_{eff} \Delta I \frac{dI}{dt} \), where \( n \) is the number of simultaneously switching drivers, having a given current surge \( dI/dt \) (as discussed in more detail in Chapter 3). Shown in Table 6-8 is the lead inductance calculated for various packages using a three-dimensional inductance program [191]. Lead inductance may be used to estimate the effective package inductance, given a detailed knowledge of chip configuration and the number and location of power and ground leads (all of which should be optimized for the package and chip of interest [192]). The relation between lead inductance and \( L_{eff} \) is quite geometry-specific and will not be expanded upon further. Signal coupled noise, which is equally dependent on a detailed package description, will also not be considered here. Referring to the table of package inductances, it should be noted that for most of the package geometries chosen, a broad range of lead inductance is observed, indicating the need for personalization of signal and power/ground assignments.

<table>
<thead>
<tr>
<th>Substrate Size (mm)</th>
<th>24</th>
<th>24</th>
<th>24</th>
<th>30</th>
<th>18</th>
</tr>
</thead>
<tbody>
<tr>
<td>Lead Length (mm)</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>(min.)</td>
<td>2</td>
<td>1</td>
<td>2</td>
<td>12.5</td>
<td>6.5</td>
</tr>
<tr>
<td>(max.)</td>
<td>17</td>
<td>13.5</td>
<td>17</td>
<td>17.7</td>
<td>9.0</td>
</tr>
<tr>
<td>Lead Inductance (nH)</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>(min.)</td>
<td>3.5</td>
<td>2.5</td>
<td>7.0</td>
<td>16.0</td>
<td>5.0</td>
</tr>
<tr>
<td>(max.)</td>
<td>13.3</td>
<td>6.3</td>
<td>19.0</td>
<td>23.0</td>
<td>7.2</td>
</tr>
</tbody>
</table>

6.9 DENSITY OF CONNECTIONS

A dramatic difference in attainable connection density between area arrays and peripheral pads was shown previously in (Table 6-2) and Figure 6-13. With the advent of aren TAB and multiple wirebond layers, improvements in density are being made (Fig. 6-67). High performance and I/O technologies in both leading edge CMOS and bipolar chips will favor area array technologies over peripheral. The present area of applicability of wirebond, TAB, and C4 is shown in Figure 1-16 on page 30 in Chapter 1, "Microelectronics Packaging — An Overview."

6.10 SUMMARY

A typical process comparison of C4, wirebond, and TAB for a multichip module application is shown in Table 6-9. Although process costs and density are major influencing factors, reliability, component and tool...
Table 6.9. Chip Interconnection Assembly Comparison for Multichip Module.

<table>
<thead>
<tr>
<th>C4</th>
<th>Wirebond</th>
<th>TAB</th>
</tr>
</thead>
<tbody>
<tr>
<td>1. Flux site</td>
<td>1. Solder or epoxy preform place</td>
<td>1. Peel tape into bond position</td>
</tr>
<tr>
<td>2. Align and place chip(s)</td>
<td>2. Align and place chip</td>
<td>2. Align and place chip</td>
</tr>
<tr>
<td>3. Reflow to bond all pads (On all chips)</td>
<td>3. Die bond</td>
<td>3. Inner-lead-bond (One-chip-at-a-time)</td>
</tr>
<tr>
<td>5. Test</td>
<td>5. Test</td>
<td>5. Test/burn-in (optional)</td>
</tr>
<tr>
<td>7. Align and place assembly on S/S</td>
<td></td>
<td></td>
</tr>
<tr>
<td>8. Outer-lead-bond</td>
<td></td>
<td></td>
</tr>
<tr>
<td>9. Test</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

The choice of assembly technique also depends upon constraints not discussed in the text so far. A manufacturer building an assembly for in-house use may opt for a technology for which he or she already has manufacturing capability, even though that technology may not be optimized for performance. Likewise, a small company relying completely on vendors for chips and substrates must choose from what is available rather than what is most desirable. This leads to the conclusion that while technical trade-off discussion is desirable, the decision to use one technology over another is more than technical.

By far, the two main developments having the biggest impact on chip interconnections technologies are (1) materials evolution (expansion coefficient, thermal conductivity, purity), opening options in many areas, and (2) tool developments, including highly automated assemblers with pattern recognition, computer control, and low-contamination manufacturing environments.
6.11 REFERENCES

6.11 REFERENCES


REFERENCES


L.S. Mok. Private communication, 1986.


M.F. Bregman. Private communication.
