K. C. Norris
A. H. Landzberg

Reliability of Controlled Collapse Interconnections

Abstract: The use of solder pads to join multi-pad integrated circuit chips to modules provides a highly reliable, rugged interconnection technology. This paper reports some important aspects of the reliability evaluation that was carried out on the "controlled chip collapse" interconnection system developed by IBM. Included are an analysis of the mechanics of the system, a model to establish the relationship among different thermal fatigue testing conditions, and experimental verification of the model. In the course of this work, the chip failure rate of the interconnection as used in present designs was predicted to be better than 10^{-7}%/1000 hours for the mechanism studied.

Introduction

Integrated circuit devices require multiple connections between the integrated circuit chip and the module. Experience with chip mounting in SLT (Solid Logic Technology)[1] has been extremely favorable with regard to both process compatibility and reliability. An extension of this technology using "controlled collapse" solder joining[2] provides additional ductility and a self-aligning feature which improves process capability.

Any device joining system is likely to have an expansion mismatch between the chip and the module, which produces mechanical stress in the joints because of the difference in coefficients of thermal expansion between the silicon of the circuit chip and the ceramic used for the module substrate. This paper describes a physical model used to determine the effect of this stress on tin-lead solder fatigue.

The available literature on solder systems[4,5,7] reveals that tin-lead solder is quite susceptible to metal fatigue. In the controlled chip collapse interconnection this can take the form of thermal fatigue, which differs from mechanical fatigue in that fixed strain levels rather than fixed stress levels are exhibited. It therefore had to be determined whether thermal fatigue of the solder joint will occur within the usage time-period of the integrated circuit component, when temperature is changed (as during machine turn-on and shutdown) and large strains are initiated.

This question is addressed by first calculating the strain levels in the interconnection for an applied temperature change, then establishing a model to determine the effects of thermal fatigue, and finally verifying this model experimentally. Interconnection failure rate is then predicted for the component in an equipment environment.

The authors are located at the IBM Components Division Laboratory, Poughkeepsie, New York.

Description of interconnection and its application

In one technique of the controlled chip collapse technology, discs of 5-95 Sn-Pb (5% tin, 95% lead) solder are evaporated on a silicon wafer at metallized contact sites, which consist of a triple-layer deposit of chromium, copper, and gold. The solder is first reflowed in a reducing atmosphere; then the chip is joined, by reflowing the solder a second time, to a ceramic substrate on which silver-palladium paste, tinned with 10-90 Sn-Pb solder, is deposited. Simplified drawings which show the appearance of the interconnection at different stages in the process are shown in Fig. 1. A metallographic cross-section of a joint is shown in Fig. 2.

Temperature excursions occurring when equipment is turned on and off create displacements of the chip relative to the substrate; viewed with respect to the geometry of the joint, these displacements lead to local and average shear strains of significant magnitude. The thermal operating conditions of systems vary considerably, depending on such things as system size, application, and even altitude above sea-level (determines air density).

The thermal excursion which encompasses most applications is $\Delta T = 75°C$, at a rate of 1.5 cycles/day. For a chip 100 mils square, these conditions lead to average strains of about 0.5% (higher local strains would be anticipated due to the shape of the interconnection). Such strains are known to be significant in the fatigue of lead-bearing alloys.[4,5]

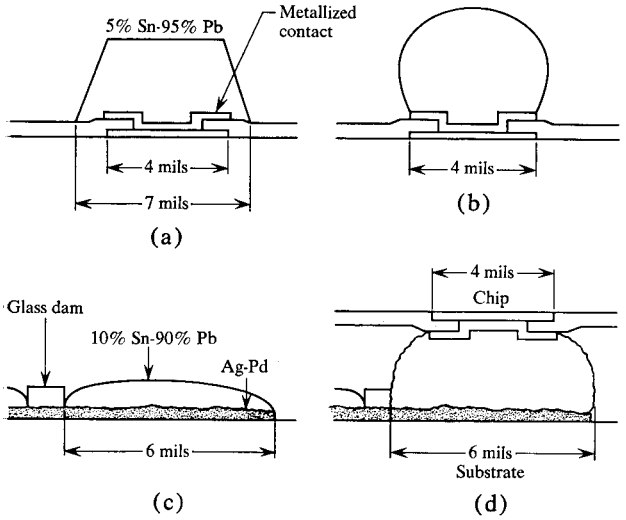


Figure 1 Construction of solder interconnection: (a) chip after solder evaporation; (b) chip after first reflow; (c) tinned substrate; (d) complete contact after joining.

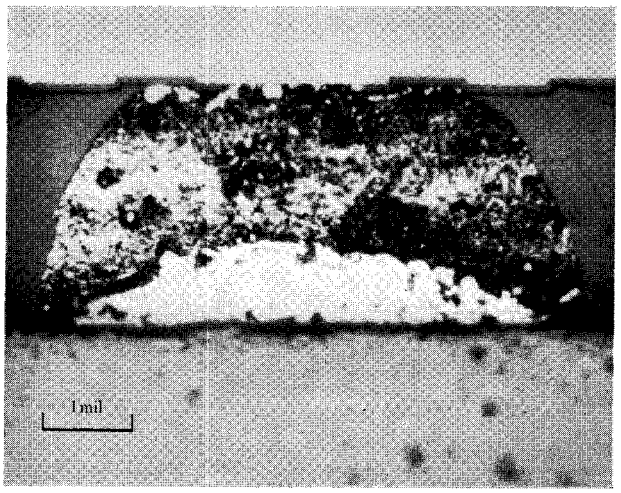


Figure 2 Metallographic cross section of controlled collapse joint.

Thermal fatigue has not been observed in SLT modules, which have three closely spaced interconnections 10 mils apart; the strains in SLT interconnections are too small to create a fatigue condition—less than one tenth of that experienced by a 100 mil chip.

Thermal fatigue model

In order to obtain early information, a model to evaluate the fatigue behavior of solder interconnections under accelerated conditions and to predict fatigue life under machine operation has been developed. It takes into account the time and temperature factors that have been found to be significant, as well as mechanical strain.

• *Effect of strain on thermal fatigue*

The Coffin-Manson equation[3] relates plastic strain, $\Delta\epsilon_p$, and the number of cycles, N, to failure:

$$N(\Delta\epsilon_p)^n = C,$$

where n is an empirical constant observed to be 2 for nearly all metals and C is a proportionality factor. In the case of a tensile test, $\Delta\epsilon_p = \epsilon_f$, the fracture ductility, and $N = \frac{1}{4}$, so that $C = \epsilon_f/2$, which has been verified for a large number of metals provided the test temperature is sufficiently low that slip occurs and no creep or relaxation effects are present.[1] However, since creep and grain boundary sliding predominate in the deformation of lead alloys above room temperature, the use of $C = \epsilon_f/2$ is invalid for tin-lead solder.

The thermal displacement caused by a change in temperature of 75°C leads to an average total strain (elastic strain + plastic strain) of 0.5% for a 100 mil chip, more than an order of magnitude greater than the elastic limit strain (0.02% at room temperature). Hence, the total strain may be considered equal to the plastic strain.

Following the analysis of Goldmann,[10] who assumes that the interconnection is a spherical segment, the plastic strain amplitude at any cross section is:

$$\Delta\epsilon_p = \delta\left(\frac{V}{\pi r^2 h^{1+\beta}}\right)^{1/\beta} = \alpha_{rel} l \Delta T\left(\frac{V}{\pi r^2 h^{1+\beta}}\right)^{1/\beta},$$

where

δ = displacement of chip relative to substrate;
α_{rel} = relative thermal expansion coefficient of chip to substrate, approximately 3.5×10^{-6} in/in °C, ($\alpha_{silicon} = 2.5 \times 10^{-6}$ in/in°C, $\alpha_{Al_2O_3} = 6 \times 10^{-6}$ in/in°C);
l = distance from chip neutral point to interconnection;
ΔT = temperature excursion for the cycle;
h = height of the solder;
V = volume of solder;
r = radius of cross section under consideration;
β = exponent from plastic shear stress-shear strain relationship $\tau = K\epsilon^\beta$.

If the design dimensions and materials are held constant, the expression assumes the form

$$\Delta\epsilon_p/\Delta T = \text{constant},$$

and the Coffin-Manson equation becomes

$$N(\Delta T)^2 = \text{constant},$$

with the strain-related exponent n taken to be 2. In making the substitution above it has been assumed that thermal

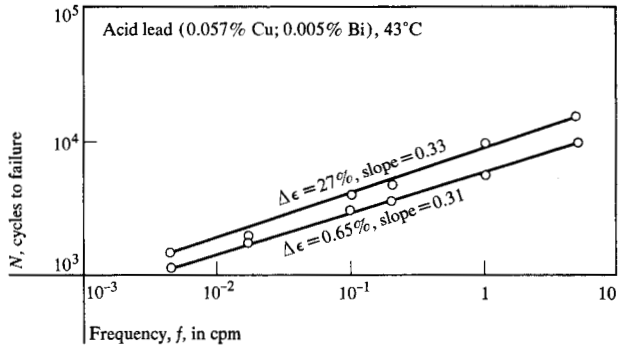
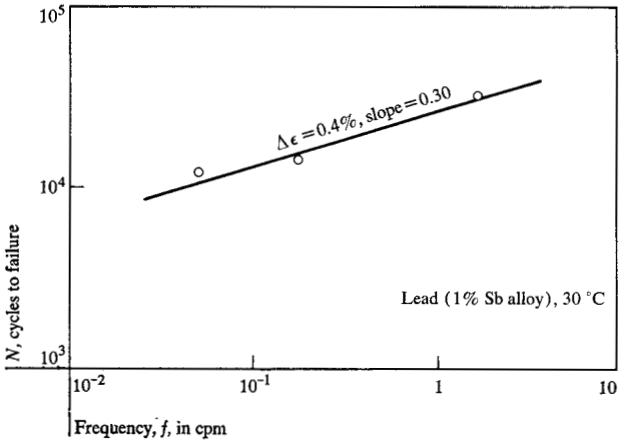


Figure 3 Fatigue vs. cyclic frequency, acid lead. (After J. F. Eckel.[4])

Figure 4 Fatigue vs. cyclic frequency, Pb-1% Sb alloy. (After Gohn and Ellis.[5])



fatigue failures occur at a cross section of common $\Delta\epsilon_p$. Thermal cycling tests have shown this to be a valid assumption; most fractures occur at or near the chip-to-solder interface, which is the area of maximum shear strain.

The Coffin-Manson equation was found to be inadequate for projecting the thermal fatigue of solder interconnections; in laboratory experiments it was found to yield very pessimistic estimates of fatigue lifetimes. Furthermore, for tin-lead solders, room temperature must be considered as elevated temperature (greater than half the absolute melting temperature), and a reasonable fatigue model must attempt to account for the effects of relaxation phenomena. Accordingly, modifications to account for time- and temperature-dependent properties of solder were introduced.

• *Effect of frequency of cycling on thermal fatigue*

Eckel[4] has shown that cycling frequency has a pronounced effect on the fatigue life of acid (0.057% copper, 0.005% bismuth) lead. His data, plotted in Fig. 3, relate cycles-to-failure and frequency f, approximately by

$$N/f^{\frac{1}{3}} = \text{constant}.$$

Limited data from Gohn and Ellis[5] for a lead-antimony alloy obey the same relationship; those data are plotted in Fig. 4. Coles and Skinner[6] have extended and generalized Eckel's results to fatigue in the creep regime:

(1) $N \propto f^k$, the extremes being $k = 0$ for pure fatigue, where frequency is not a factor, and $k = 1$ for creep rupture, where failure depends only on time.
(2) A hold period or dwell time in creep fatigue may be regarded as a frequency adjustment.

The consistency of the existing literature with respect to the exponent of $\frac{1}{3}$ has led to its addition as an empirical correction term in the Coffin-Manson equation to account for time-dependent properties of solder. No attempt is made to deal individually with specific time-dependent properties (strain-rate effects, relaxation, etc.); instead, all such effects upon fatigue life are contained in the empirical frequency factor.

• *Effect of maximum temperature of the cycle on thermal fatigue*

When strain is applied at a continuously changing temperature, it is anticipated that fatigue life will be decreased in the upper temperature regions of the cycle due to temperature-related effects such as increased grain boundary sliding. Data from the Southern Research Institute[7] for 5-95 Sn-Pb solder cycled at constant frequency, constant strain, and various temperatures are plotted in Fig. 5 to show this effect. From the ratio of lifetimes under two temperature conditions an empirical temperature factor which places emphasis on the peak temperature attained during the cycle may be defined:

$$\Phi(T_{\max}) = \frac{\text{Life } (T_{\max 1})}{\text{Life } (T_{\max 2})}.$$

From Fig. 5 it can be seen that the ratio of fatigue life at 85°C (maximum machine condition) to that at 150°C (typical test condition) is approximately a factor of 2.

• *Thermal fatigue model*

By adding the empirical correction factors for time dependent and temperature dependent effects, the complete model for thermal fatigue of solder interconnections is obtained:

$$\frac{N_{\text{Lab}}}{N_{\text{Machine}}} = \left(\frac{f_L}{f_M}\right)^{\frac{1}{3}} \left(\frac{\Delta T_M}{\Delta T_L}\right)^2 \Phi(T_{\max}).$$

This model is limited to use in describing solder fatigue, since the empirical frequency and temperature modifications were devised to apply to tin-lead solder.

Experimental results and discussion

All of the experiments to verify the solder fatigue model were performed using a 112-mil chip with 23 controlled

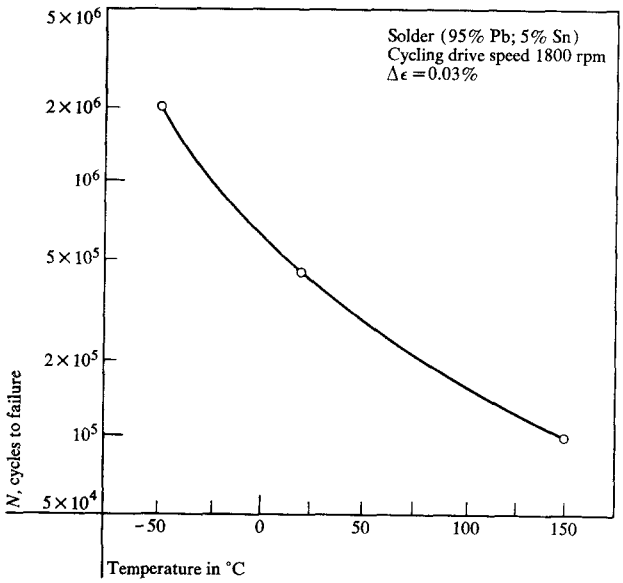


Figure 5 Fatigue vs. temperature, 5-95 Sn-Pb solder.[7]

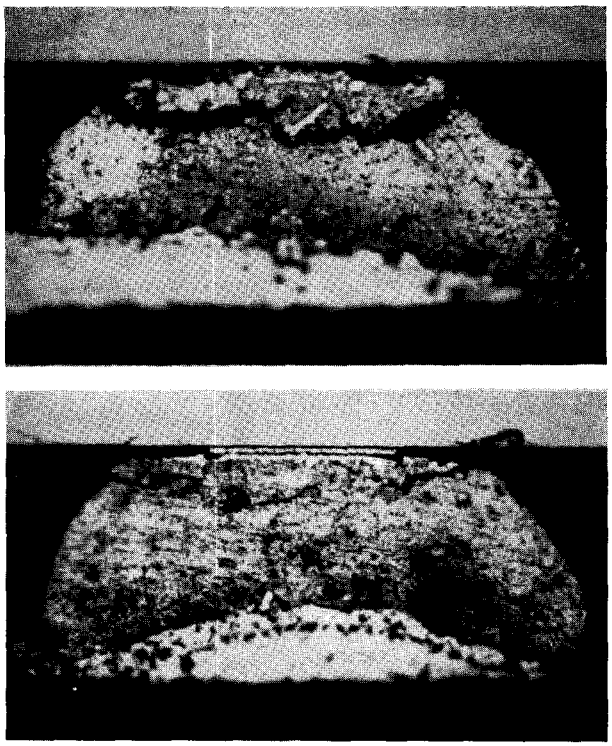


Figure 6 Two typical fatigue failures.

collapse interconnections. On all of the samples tested, the metallized contacts at the chip were 4.0 mils in diameter, and the tinned area on the substrate was 6 mils square. Because plastic strain amplitude, and consequently fatigue life, are closely linked to dimensional considerations, all of the samples in each experiment were designed with identical geometry. In some experiments, the evaporated solder volume on the chip was 60 mil^3, leading to local plastic strain at the chip interface of 0.051%/°C; in the remaining tests, the evaporated solder volume was 90 mil^3, which leads to local plastic strains at the chip interface[10] of 0.039%/°C.

Electrical measurements of the samples were made at regular intervals to determine whether degradation or failure of the interconnections had occurred. Failure of the chip occurs when a single interconnection becomes an electrical open. Because plastic strain amplitude depends on the distance from the chip neutral point,[10] the corner interconnections, or those close to the corners, are preferred sites for failure; over ninety per cent of the failures were observed at these locations.

Laboratory testing was carried out in Tenney and Delta thermal cycling chambers. In the former, a test fixture cycles between a hot chamber and a cold chamber in a preset cycle; in the latter, the samples are both heated with a heater and cooled with liquid nitrogen in a single chamber. In either type of chamber, the temperature excursion and frequency of the cycle can be preselected.

In a typical laboratory cycle of 0–150°C, 2 cycles per hour (cph), the heating cycle takes approximately eight minutes with a seven minute dwell at the upper temperature, and the cooling cycle takes approximately seven minutes with an eight minute dwell at the lower temperature.

When the thermal fatigue tests were carried out to failure, cracks were observed to originate at the chip interface and propagate through the bulk solder of the interconnection. Typical solder fatigue failures are shown in Fig. 6. Parameters of the log-normal distribution,[8] which was found appropriate to describe the statistics of interconnection fatigue failures, are used to correlate the test results. Fatigue life of a sample cell is defined by the distribution parameter N_{50}, mean-cycles-to-failure; comparisons among different stress cells are made by using the N_{50} of the cells.

To determine how accurately the thermal fatigue model predicts failures, a test was designed to compare fatigue under machine temperature conditions with that under accelerated test conditions. Two groups of 20 modules with evaporated solder volume of 60 mil^3 were cycled at 25–85°C, 3 cph, and 0–150°C, 2 cph, respectively. From the thermal fatigue model, the predicted acceleration of the 0–150°C test over the 25–85°C test was estimated to be thirteen times; actual test results showed the acceleration to be 12.35 times (Fig. 7). The figure also shows that good agreement is obtained between the predicted and observed fatigue life distributions.

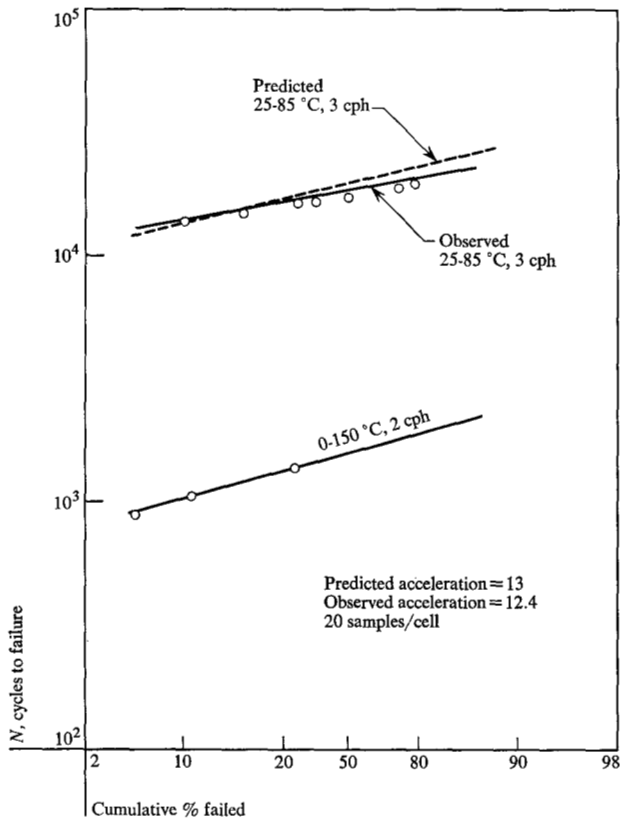
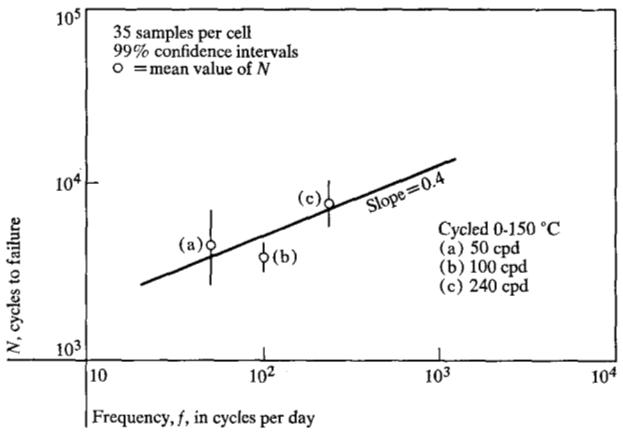


Figure 7 Fatigue tests for (a) 25-85°C; (b) 0-150°C.

Figure 8 Fatigue vs. cyclic frequency.



To obtain additional information on the accuracy of the fatigue model, a matrix of tests with 35 samples apiece was conducted on samples with an evaporated solder volume of 90 mil^3 to correlate factors for maximum temperature and frequency. Three frequencies, 50, 100, and 240 cycles per day (cpd), and two maximum temperatures, 110°C and 150°C, with a strain level corresponding to ΔT of 150°C were used. Results for the 0—150°C tests at 50, 100, and 240 cpd are presented in Fig. 8; regression analysis yields a slope of 0.4 for these data, a fair correlation with the slope of $\frac{1}{3}$ in view of the scatter observed in the data. The tests that were conducted with f = 240 cpd and maximum temperatures of 110°C and 150°C gave a ratio of mean fatigue lifetimes $N(110)/N(150) = 1.3$, which is in good agreement with the 1.45 predicted by the fatigue model.

For the strain amplitude effect, data from three different test conditions were correlated by assuming that the frequency and maximum temperature relationships of the model apply and normalizing with respect to these factors to obtain the strain-related slope $1/n$. Twenty modules with evaporated solder volume of 60 mil^3 were tested at each of the following conditions:

(1) −60 to 150°C, 3 cph, to 15% failure;
(2) 0 to 150°C, 2 cph, to 33% failure;
(3) 30 to 110°C, 30 cph, to 50% failure.

These data, which are plotted in Fig. 9 for 15% failure and 33% failure levels, yield a value of n = 1.9, in excellent agreement with the typically observed strain exponent, n = 2.

Failure rate determination

Development of a model to describe thermal fatigue of tin-lead solder provides a method for obtaining accurate assessment of controlled chip collapse interconnection reliability in a machine environment. Four groups of thirty-five samples each with 90 mil^3 evaporated solder volume were thermally cycled under the following accelerated conditions:

(1) 0°C to 150°C, 2 cph;
(2) 0°C to 150°C, 4 cph;
(3) 0°C to 150°C, 10 cph; and
(4) −40°C to 110°C, 10 cph.

Each group was tested to at least fifty per cent failure.

Using the modified Coffin-Manson model, the number of cycles to failure for all solder fatigue failures was converted to equivalent machine cycles to failure and ultimately to equivalent machine hours; this was based on a 25°C to 85°C temperature excursion (normal room ambient to maximum junction temperature) at a rate of 1.5 cycles per day (typical machine frequency). The log-normal lifetime model[8] was then used to predict failure rate.

The thermal fatigue data from these experiments, plotted in Fig. 10, have a mean time to failure, t_{50}, of 355,000 hours and a deviation σ, $[\ln(t_{50}/t_{15})]$ of 0.379. The 1% cumulative failure point occurs at approximately 150,000 hours,

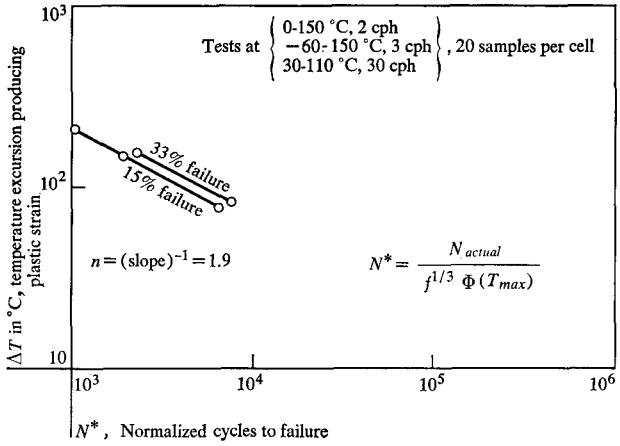


Figure 9 Relationship of strain to thermal fatigue.

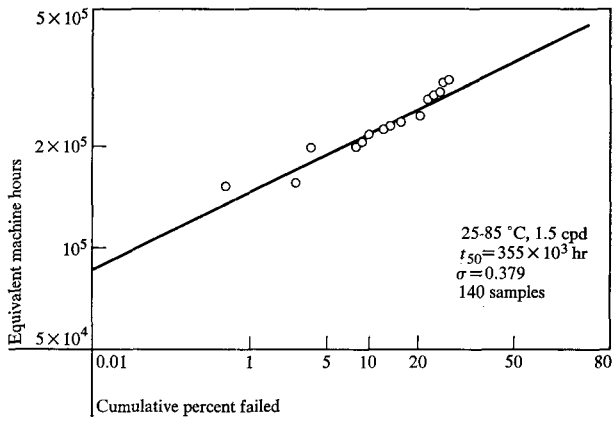


Figure 10 Solder fatigue failure distribution.

well beyond the design lifetime. The instantaneous failure rate at any time t is given by:

$$\text{F.R.} = \frac{1}{R(t)\sigma t(2\pi)^{\frac{1}{2}}} \exp\left\{-\frac{1}{2}\left[\frac{\ln(t) - \ln(t_{50})}{\sigma}\right]^2\right\},$$

where $R(t)$ is the reliability function of the distribution. Using this formula, the chip failure rate due to solder fatigue of controlled chip collapse interconnections at machine lifetime is predicted to be $10^{-7}\%/1000$ hours; in earlier stages of machine operation it is less than $10^{-7}\%/1000$ hours.

Acknowledgements

The authors express their appreciation to B. E. Blake for his support in establishing necessary testing programs and for his assistance and encouragement in preparing this manuscript. We thank D. S. Chhabra whose insight contributed to the development of the model proposed in this paper. We also thank H. C. Kammerer and S. S. Im who carried out extensive testing programs in evaluating the reliability of controlled chip collapse interconnections.

References

1. E. M. Davis et al, "Solid Logic Technology: Versatile, High Performance Microelectronics," *IBM J. Res. Develop.* **8,** 102 (1964).
2. L. F. Miller, "Joining Semiconductor Devices with Ductile Pads," *1968 Hybrid Microelectronics Symposium* (Society for Hybrid Microelectronics) Chicago, October 29, 1968.
3. S. S. Manson, *Thermal Stress and Low-Cycle Fatigue,* McGraw-Hill Book Company, New York 1966.
4. J. F. Eckel, *Proc. ASTM* **51,** 745 (1951).
5. G. R. Gohn and W. C. Ellis, *Proc. ASTM* **51,** 721 (1951).
6. A. Coles and D. Skinner, *J. Royal Aeronaut. Soc.* **69,** 53 (1965)
7. Southern Research Institute, "Fatigue Properties of Solder Materials," *I.D. No. 2281–59,* August 10, 1960.
8. Lynn R. Goldthwaite, "Failure Rate Study for the Lognormal Lifetime Model," *Bell Telephone Systems Technical Publications, Monograph 3814.*
9. L. F. Coffin, Jr., "Low-Cycle Fatigue," *Metals Engineering Quarterly,* November 1963, p. 15.
10. L. S. Goldmann, "Geometric Optimization of Controlled Collapse Interconnections," *IBM J. Res. Develop.* **13,** 251 (1969).

Received December 11, 1968