Assuring High Reliability of Lasers and Photodetectors for Submarine Lightwave Cable Systems

ASSURING HIGH RELIABILITY OF LASERS AND PHOTODETECTORS FOR SUBMARINE LIGHTWAVE CABLE SYSTEMS

Introduction

By R. L. EASTON, R. L. HARTMAN, and F. R. NASH*

(Manuscript received October 5, 1984)

The transatlantic optical fiber submarine cable system—employing semiconductor lasers and photodetectors and scheduled for service in mid-1988—is the first application of lightwave technology to transoceanic communications. This eighth transatlantic undersea system, known as TAT-8, represents a radical departure from its predecessors. The previous systems used coaxial cables as a transmission medium, with periodically spaced repeaters to amplify the analog signal back up to its original level, thus compensating for the cable attenuation. The technology and components required were well understood from previous use in similar ocean and land-based systems. In contrast, TAT-8 will use pulses of light traveling in glass fibers, with periodic regeneration of the pulse train. This will require the use of state-of-the-art semiconductor lasers and photodiodes for which there is no prior reliability experience for a land-based system.

There are several advantages which motivate this changing over from an analog-coaxial cable to a digital-optical fiber waveguide system. The coaxial cable technology has reached a level of maturity where only modest further cost reductions can be anticipated from future designs, whereas the lightwave technology is still evolving rapidly. Significant increases in repeater spacing and higher bit rates— making possible large cost reductions and increases in traffic capacity and improving the reliability of deep-water repeaters—can be anticipated. Furthermore, a digital submarine cable can be part of an

---

* Authors are employees of AT&T Bell Laboratories.

all-digital world network, which permits applications not possible if analog links are present. One very important example of such an application is the highly efficient circuit multiplication made possible by digital processing of voice signals. The price of obtaining the advantages offered by this new technology, however, is the use of state-of-the-art components.

The question that we consider in this issue is the following. If a lightwave submarine cable is built, is it possible in advance of deployment to assure that it will perform with no more than approximately three repairs in the desired 25-year system lifetime—a goal dictated by the costs of repairing a deep-water cable, including the provision of alternate transmission service during a period of up to several weeks that it may take to effect a repair. The answer to that question depends critically upon the behavior of the two newest state-of-the-art semiconductor components, lasers and photodetectors. Given that there already exists evidence suggesting feasibility (i.e., at least some individual devices appear to possess the potential for long life), the determination of whether adequately large numbers of lasers and photodetectors have sufficient reliability is conveniently broken up into two stages. Using the available knowledge concerning modes of failure about these and other similar semiconductor lasers and detectors, the first stage involves the design of the most suitable strategies for the timely detection of potential premature device failures. With a tentative best-approach in hand—one subject to alteration as subsequent data analysis might compel—stage two is concerned with determining whether any particular group of devices is adequate in fact.

This two-stage approach is mirrored by the division of papers in this collection. The emphasis in the first group is upon analyses of various reliability assurance strategies, including those proposed by us, as well as the more traditional ones. These analyses serve to critique, organize, and develop methodologies for reliability assessment. The second group of papers focuses upon the implementation of these strategies *and* the results of preliminary studies of unpackaged devices. These results provide compelling evidence for our conclusion that the device designs are consistent with the high-reliability goal and that the strategies which we have employed can provide for the selection of those devices that will meet that goal.

There were several motivations for producing the first group of papers. One was the absence in the literature of a comprehensive and critical discussion of methodologies for reliability assurance. Although the literature on reliability assurance is vast, the focus tends to be on specific failure modes of specific devices or upon mathematical modeling. No prior work provides an adequate organized way of

thinking about how to assure the long-term performance of new semiconductor components. While the first group of papers often employs the laser as a particular exemplary device, that group is intended to have potential applications for a broad range of semiconductor devices. It represents the papers that we wished had existed when we embarked upon this project.

Another motivation was prompted by the difference between the approach mandated for the 1988 deployment of TAT-8 and the approach that had governed all previous systems. Historically, the assurance of the high reliability required for submarine cable transmission was predicated upon the use of well-proven device types for which failure modes had been identified in previous systems and for which performance/reliability trade-offs were well understood. The desire to harvest the benefits offered by new technology requires the incorporation of rapidly evolving state-of-the-art components for which there has been no prior use and about which little is known in detail. Substantial emphasis must then be placed upon strategies for the early detection of system-threatening premature failures, which *may* occur. These strategies are based upon a priori considerations, rather than extensive manufacturing and field experience. There has to be as much focus upon the possible as upon the probable failure modes.

A final motivation springs from the observation that certain traditional strategies fail to take full advantage of the character of failure which exists in semiconductor components. For example, it is observed that detectors fail suddenly, without warning. In such an event, traditional population-inference approaches are the best available. It is also observed, however, that mechanisms controlling laser lifetime lead to gradual degradation. In this event, information about the early behavior of an individual laser can be very important in assessing the prospects for longevity of that device. What is needed for lasers is an individual-inference scheme to supplement the traditional population-inference approach; substantially greater reliability assurance is the goal. Such a scheme is discussed at length in the first group of papers.

All of the papers in this collection have a common approach. It is that reliability assurance is to be determined by extrapolation in temperature rather than time. Either one or the other is inevitable when the desired system lifetimes are tens of years; clearly it is unacceptable to wait for 25 years to determine whether a device operating in field conditions will actually live 25 years. For a slowly degrading device such as a laser, both temperature and time extrapolations are possible in principle. A straightforward application of the purely temporal option would involve aging each laser at the

use temperature for as long as it took to determine the aging law (e.g., the time dependence of the operating current, $I_{op}(t)$, to maintain a constant optical power output), and then using this law to compute the behavior at the end of the designed system lifetime. Experience with aging long-wavelength (1.3-μm) lasers suggests that the data follow a sublinear aging law that is not universal, e.g., $\beta t^{1/2}$, but which appears to have a universal form,* e.g., βt^m, $m < 1$. Thus, the parameters β and m will be different for each laser. The problem, however, is that even if an empirical aging law were adequately parametrized for the aging-test period (approximately 1 year), the temporal approach offers no assurance that the same law or the same parameters are valid at 25 years. The very nonuniversality of the aging law—i.e., the fact that the law is not, for example, $t^{1/2}$ for every laser— raises the suspicion that each laser is being controlled by different processes. The unsatisfactory nature of the temporal extrapolation is that an empirical law established, however well, in some test period is then assumed to hold for a period of time that is more than an order of magnitude longer than the test period. This is not the approach which we have adopted, because there is *no* basis for any confidence in such a temporal extrapolation.

The approach which underlies virtually all of the work in this collection is that of thermal extrapolation. Semiconductor reliability experiments show that failure can be accelerated by increasing the temperature of the device. It can be said that increasing the temperature makes the internal device clock run faster. The degradation rate (or reciprocal of the lifetime) is often found, and hence usually assumed, to be proportional to $\exp[-E_a/kT]$ where the empirically determined activation energy, E_a, has a reasonably well-defined value for a particular type of device. Proof of the validity of this acceleration factor between test and use temperatures is, of course, an important part of a reliability assurance program. Once E_a is known, it is possible to age devices at elevated temperatures for relatively short periods of time (approximately 0.1 year), which are equivalent to examining the actual behavior of the device for a system lifetime (25 years) at a lower use-temperature. The behavior of a device at the end of 25 years can be determined without waiting 25 years. It is this ability to "crystal ball" the future which makes the temperature extrapolation approach so desirable; temporal extrapolation has no comparable virtue. This does not preclude using observations over a relatively long test period (though still short compared to system life) under use conditions to

---

* Aging should proceed sufficiently long so that an empirical sublinear law of the form $\beta \ln(bt + 1)$ could be distinguished from βt^m, $m < 1$. If insufficient aging time were available to make such a distinction, the more conservative form, βt^m, could be chosen.

enhance confidence that the accelerated aging scheme accurately models the behavior of the laser in the system.

For the purpose of guiding the reader through this volume, the remainder of the introduction will be devoted to a brief description of the contents. In the first paper, "Selection of a Laser Reliability Assurance Strategy for a Long-Life Application," critiques of traditional approaches to reliability assurance provide relevant nomenclature and concepts. To overcome the perceived deficiencies of conventional approaches, to deal with anticipated types of particularly deleterious or misleading degradation behavior in semiconductor devices, and to take full advantage of the character of long-term laser degradation, an alternative strategy for providing reliability assurance is proposed. Its regimes are called "stabilization," "purge," and "truncation." As previously noted, lifetime estimates will rely upon temperature extrapolation. Confidence in these estimates would be substantially undermined if the long-term degradation rates measured in elevated-temperature accelerated-aging tests reflected only an initial transient behavior *or* if there existed weakly temperature-activated, suddenly occurring, early (infant) failure mechanisms that might be expected to remain undetected if temperature were the only degradation accelerant used. By overstressing all available driving mechanisms for degradation, purging (a timely identification of infant failures) and stabilization (a rapid driving to completion of transient behavior) are accomplished. The result should be a population of devices whose longevity is controlled exclusively by a gracefully occurring degradation (wear-out). Truncation is the process by which devices are sorted according to their initially measured degradation rates and only those degrading sufficiently slowly are kept for use. The success of the rate-monitoring process to implement truncation depends upon a prior elimination of devices prone to sudden failure and the stabilization of transient behavior.

The starting point in the second paper, "Methodology of Accelerated Aging," is at the end of the stabilization and purge regimes. It is assumed that a long-term thermally accelerable wear-out mode of degradation controls the behavior of all devices; all infant mortalities have been eliminated and all transient behavior has been stabilized. In approximately one-half of this paper, a conceptual framework is presented for understanding the information provided by accelerated aging, using temperature as an exemplary accelerant. Considerable analysis is given on the important distinction which must be made between, for example, a thermal activation energy (all measurements of failure or degradation caused by operation at higher temperatures are performed at a lower reference temperature) and a thermal extrapolation energy (measurements of failure or degradation are performed

at the same temperatures at which the failure or degradation occurs). The related discussion makes clear why there can be considerable variations in the experimentally determined values of these often undifferentiated energies. In a given experiment, the amount of degradation is determined not just by the initial and final aging points in time-temperature space, but also by the particular path between these points. As a consequence, it is next shown that the simplest possible model of accelerable degradation is, in mathematical terms, a differential form. Differential-form modeling is developed and used for extrapolation outside the scope of the usual integrated-form models. In the event that the projected device lifetime exceeds that of the system, it is next shown how to achieve a considerable saving in aging time and cost by not aging the device to failure, but only for an *effective* system lifetime, i.e., the equivalent, but shorter in real time, lifetime of a device aged at some elevated temperature. Finally, a procedure is given for assessing the effectiveness of truncation, i.e., a procedure for quantifying the uncertainty in projections based on the initially observed degradation rates.

This truncation-assessment procedure is given a more detailed quantitative formulation in the third paper, "A Statistical Approach to Laser Certification." A statistical "correction factor"—obtained by comparing the *actual* degradation that occurs for a population of devices aged for an effective (i.e., temperature-contracted) system lifetime with the *predicted* degradation based upon observations of initially occurring degradation rates—is important in quantifying the risk associated with a reliability assessment dependent upon initially observed rates.

The group of experimental studies commences with "1.3-μm Laser Reliability Determination for Submarine Cable Systems," whose main purpose, consistent with the temperature extrapolation approach for lifetime estimates, was the determination of the activation energy associated with the long-term (wear-out) degradation and characterization of the aging behavior at elevated temperatures. The lasers used were neither stabilized nor purged. To the extent that the initial degradation behavior displayed undesirable characteristics, the elimination of those characteristics would be the goal of a stabilization procedure. Indeed, a substantial transient behavior (large initial degradation rates) was observed which required approximately 1 kilohour to stabilize even at elevated ambient temperatures (60°C). The more graceful wear-out degradation occurred only after this initial effect had been stabilized. Because of a considerable variation observed in post-stabilization rates of degradation among nominally identical lasers, it was decided to abandon the traditional *isothermal* approach to determining activation energies, in which the median degradation rates

of two groups of lasers aged at different temperatures are compared. Instead, a *step-temperature* technique, in which each laser is consecutively aged at two elevated temperatures, was employed to determine an activation energy for each individual laser. All such values were found to lie in the range 0.8 to 1.3 eV. Assuming that failure is defined by a 50-percent increase in operating current, I_{op}, above its initial value, and assuming that the increase in I_{op} is linearly dependent on time, the estimated activation energies were used to predict more than adequate lifetimes at ocean bottom temperatures for the lasers examined. The above assumptions are conservative; other studies have shown that transmitter packages continue to perform adequately,[1] beyond the 50-percent increase in I_{op}, *and* that long-term degradation is actually temporally sublinear[2] over several effective (temperature-contracted) system lifetimes. This conclusion of adequate lifetimes for an unpurged, untruncated population leads to high confidence that a properly screened population will exhibit a comfortable additional margin.

The next paper, "Implementation of the Proposed Reliability Assurance Strategy for an InGaAsP/InP, Planar Mesa, Buried Heterostructure Laser Operating at 1.3 μm for Use in a Submarine Cable," is an attempt to put into practice the reliability assurance strategy described in the first paper of this volume. The most important aspect of this work was the discovery of a regimen in which high currents (250 mA dc) and high ambient temperatures (150°C) could be used to compel a rapid stabilization (10 hours) of a large initially occurring transient degradation phenomenon (referred to previously) that appears to affect many lasers. The unstabilized presence of this transient behavior can lead to erroneously large or small estimates of the activation energy and hence to misleadingly pessimistic or optimistic lifetime predictions. The survivors of the purge-stabilization process were subjected to the truncation process that sorted the aging rates in an elevated temperature (60°C) burn-in operation. The largest observed increase in operating current that occurred in *one* effective system lifetime was a negligible 3 percent, and observations conducted for at least *five* effective system lifetimes showed the degradation to be well behaved and the cumulative largest increase in operating current to be only about 20 percent. These results are in accord with those previously mentioned that predict more than adequate longevity for the subcable lasers.

The final experimental study, "Reliability of InGaAs Photodiodes for SL Applications," is concerned with developing and verifying a purge procedure for photodiodes. Unlike the lasers, which *appear* to have no suddenly occurring failures, such failures, which are not preceded by cognizable warning signs, can occur to a significant extent

in InGaAs photodiodes. A conventional elevated temperature (200°C) burn-in can promote these sudden failures but only at the expense of time (approximately 1 kilohour) and some consumption of the useful life of "good" diodes. A high-temperature (200°C), high-reverse-bias (20V) purge lasting 10 hours was developed to eliminate weak devices outright and to identify those devices likely to become early failures. The effectiveness of this purge was confirmed by a subsequent long-term elevated temperature burn-in.

It is to be understood that all of the experimental work presented on components is based upon data acquired in the period mid-1982 to mid-1983, and it represents an early snapshot in a continuing program that is now investigating the reliability and performance of packaged components and subassemblies such as transmitters. Additional work on lasers and detectors is also currently being performed.

Some redundancy will be noted in this collection of papers. This is inevitable and desirable as individual workers strive to refine the perspectives in which new results should be viewed. Minor differences exist in the more speculative conclusions drawn by some of the authors. These are important because they focus attention on work for future study. Particular conjectures may be applicable to the populations studied, but may not be valid for the subpopulations resulting from stabilization, purging, and truncation. Having said all of this, it is well to restate that important conclusion, unanimously supported by the authors, which is that our critical components, the lasers and detectors, have more than adequate reliability for the submarine cable application.

REFERENCES

1. C. E. Barnes et al., unpublished work.
2. J. A. Shimer and P. E. Fraley, unpublished work.

AUTHORS

Robert L. Easton, B.A. (Liberal Arts), 1950, University of Chicago; B.S.M.E. and M.S.M.E., 1953 and 1954, respectively, California Institute of Technology; AT&T Bell Laboratories, 1954—. At AT&T Bell Laboratories Mr. Easton has worked on undersea cable system development, test equipment development related to power meter design, TAT-7 negotiation, TASI-E development, and, most recently, lightwave submarine cable development. Member, Tau Beta Pi.

Robert L. Hartman, B.A. (Mathematics), 1958, Columbia College; B.S. (Engineering Physics), M.S. and Ph.D (Physics), Columbia University, in 1959, 1961, and 1968, respectively; AT&T Bell Laboratories, 1968—. At AT&T Bell Laboratories, Mr. Hartman is presently engaged in studying the physics of III-V compound semiconductor devices, particularly InGaAsP/InP and (Al, Ga)As-GaAs double-heterostructure injection lasers. He has contributed extensively to the understanding of the degradation mechanisms in these devices

and authored or coauthored several key papers in this field, especially those involving the discovery and analysis of dark-line defects, the discovery of strain effects upon laser reliability, the attainment of long operating lifetimes in LEDs and injection lasers, the first measurement of the thermal acceleration of injection laser operating lifetime, and effects of surfaces on LED and laser reliability. In 1980 Mr. Hartman was appointed Supervisor of the Semiconductor Lightwave Lasers Group. Member, Sigma Xi, American Physical Society.

Franklin R. Nash, B.S. (Physics), 1955, Polytechnic Institute of New York; Ph.D. (Physics), 1962, Columbia University; AT&T Bell Laboratories, 1963—. At AT&T Bell Laboratories, Mr. Nash has been involved in research on gas lasers (modulation and mode locking), detection of subnanosecond pulses using photodiodes, nonlinear optical materials, parametric oscillators, neoclassical theory of radiation, and semiconductor lasers (mode guidance, material defects, electro-optic anomalies, and degradation mechanisms). Currently, he is engaged in the reliability assessment of long-wavelength semiconductor lasers. Member, American Physical Society, IEEE.

Selection of a Laser Reliability Assurance Strategy for a Long-Life Application

By F. R. NASH,* W. B. JOYCE,* R. L. HARTMAN,*
E. I. GORDON,† and R. W. DIXON*

(Manuscript received March 5, 1984)

We are concerned with assuring the reliability of semiconductor lasers intended for an application in which the design lifetime is long, replacement or redundancy is impossible or impractical, and the failure of even a few lasers could be disastrous. In the search for a reliability assurance strategy that will meet our objectives, we have carefully examined the well-known and widely used bathtub and lognormal approaches. Based upon our understanding of the expected aging behavior of lasers, we propose an alternative reliability assurance strategy that we believe to be an improvement over the traditional approaches. The object is not how to make reliable lasers, but rather how to confidently predict, in a timely fashion, which lasers in a given population will endure beyond the intended system lifetime. Particular emphasis is placed upon initially imposed overstress regimes that address the anticipated presence of transient modes of degradation and infant failure modes with low thermal activation energies that may be invulnerable to detection during accelerated thermal aging. Since lasers degrade gradually rather than fail suddenly, comparable emphasis is placed upon monitoring the stabilized long-term degradation rates of the survivors of the overstress regimes so as to permit lifetime predictions of individual lasers.

I. INTRODUCTION

The problem that we faced some two years ago of assuring the reliability of the lasers to be used in the SL optical fiber submarine cable may be summarized as follows:

---

* AT&T Bell Laboratories. † Lytel, Inc., Somerville, New Jersey.

Given: A state-of-the-art device, i.e., a new semiconductor laser structure composed of new materials, which had no established record of reliability, which had never seen system use, and about whose details of operation there was a paucity of knowledge;

Goal: Design a reliability assurance strategy to determine which, if any, of a given group of these lasers may be confidently installed in a system which must operate for 25 years under circumstances where redundancy is too expensive, replacement is impractical, and where the failure of only a small number of lasers would be crippling.

To take advantage of reductions in the loss and dispersion of optical fibers, a low-threshold index-guided InGaAsP/InP laser operating at a wavelength of 1.3 μm was chosen instead of the higher-threshold gain-guided GaAlAs/GaAs laser that operated in the wavelength range 0.8 to 0.9 μm. Feasibility demonstrations of the potential for adequately long life had been made for some significant fraction of manufactured InP lasers. The size of the usable fraction and a strategy for selecting it prior to installation remained to be determined. In view of the aggressive schedule for deployment (1988), it was inevitable that the emphasis be placed upon the reliability assurance of an existing population and not upon the issue of how to design for reliability in the manufacturing process. Thus screening, not product improvement or failure mode analysis, is the primary goal.

Building in reliability from the start involves the complex technological problem of utilizing the details of innumerable specific failure mechanisms to modify the laser structure, design, growth, and processing stages of manufacture. The reliability literature is dominated by device-specific, failure-mode-specific studies, which, however useful for providing feedback to the manufacturer, do not provide an organized way of thinking about generic modes of failure and schemes tailored to their discovery. Even if expense and time were not constraints, it would be impossible, in any case, to arrive at a perfected manufacturing process that produced only long-lived devices. Thus, sooner or later, the problem that must be addressed is that of warranting the reliability of a given group of lasers, with fixed initial properties, irrespective of the failure mechanisms still operative in the manufactured population.

This paper charts the search for the reliability assurance strategy that was best suited for our goal. The intended audience is not the statistician or applied mathematician, but rather the device physicist who may some day be faced with our problem in the context of a laser or another new semiconductor device. The natural starting point for our inquiry is an examination of the strategies that are already known, in particular, the bathtub and lognormal approaches that have been widely and successfully used for some time. These traditional quality
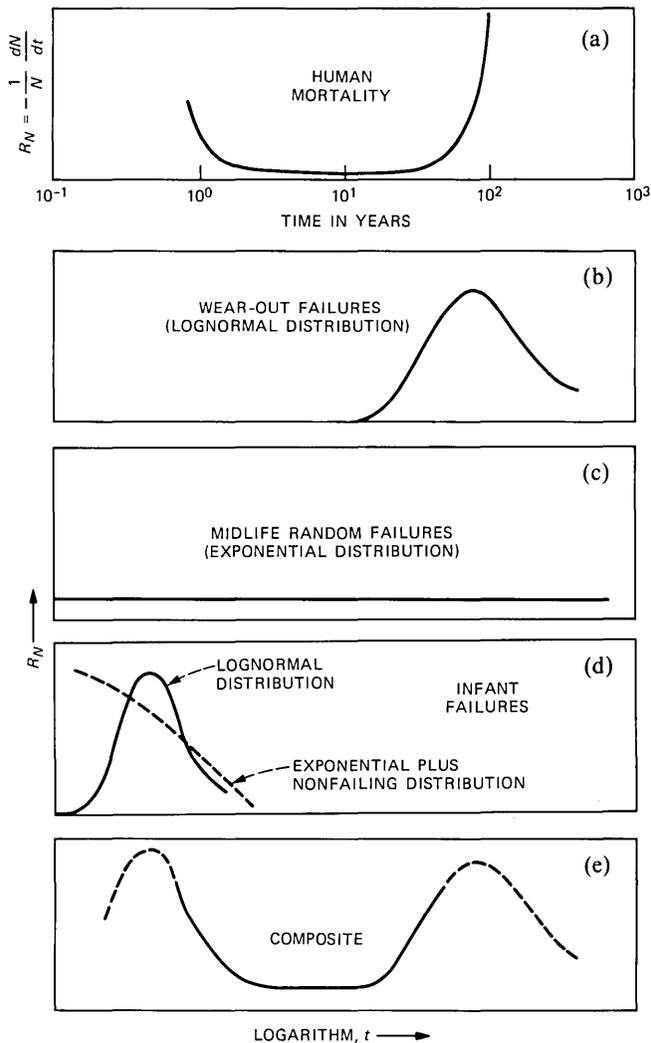
Fig. 1—(a) The hazard rate (bathtub) curve for human mortality. (b) A lognormal distribution characterizing the wear-out or long-term failures in a population of semi-conductor devices. (c) An exponential distribution which characterizes random failures, the so-called accidents in the midlives of devices. (d) Solid curve—a lognormal distribution characterizing an infant population. Dashed curve—an exponential plus a nonfailing distribution used to generate a declining hazard rate for a subset of infant failures caused by random externally imposed stress aggregations. (e) A curve which is a composite of (b), (c), and the lognormal portion of (d); beyond the bathtub part (solid curve), there is rollover of the bathtub sides (dashed curve). Only in a restricted domain of time does the semiconductor failure of (e) resemble human mortality (a).

control approaches to reliability assurance, used separately or in concert, and which rely on inferences made from the behavior of previously aged-to-failure populations, will be shown to be inadequate for our stated goal. As shall be discussed, a preferable alternative

approach, which is applicable to lasers, relies upon inferences made from the early behavior of each individual device in the very population intended for deployment.

So that a reader has the option to pick and choose among the various sections, depending upon his background in reliability assurance matters, the following "roadmap" and summary of our major points may prove useful. Section II reviews the bathtub model and a compilation of associated reliability terminology that it is important to define for purposes of making clear distinctions in subsequent discussions. Also discussed is a suggestion for a modification of the bathtub curve, which appears to be dictated by actual laser reliability data and a practical way in which to acquire bathtub data.

The traditional (human mortality) bathtub curve, a plot of the failure (hazard) rate versus time (see Fig. 1a), has a declining left-hand edge (period of infant mortality), a relatively flat bottom (period of useful life), and an increasing right-hand edge (period of inevitable failure or wear-out). By analogy, semiconductor devices are assumed to be susceptible to wear-out, but in addition some fraction of them is viewed as possessing additional failure mechanisms that cause early or infant device inoperability. If we had the traditional option of using only well proven-in devices with "nailed down" invariant manufacturing procedures and a well-established record of longevity in terrestrial applications, the bathtub reliability assurance strategy for a nonreplacement high-reliability use could be stated as follows: (A) use a mature technology that offers an intrinsic long life so that the wear-out failures (increasing right-hand side of Fig. 1a) are not present during normal (system) life, and (B) censor atypical (infant) devices (decreasing left-hand side of Fig. 1a). Simply put, the strategy is to install the survivors of the censoring process.

Competitive pressures have substantially eroded the traditional conservative approach. We are presently compelled to seriously consider the use of laser devices of an a priori unknown quality in a time frame that does not permit us to wait and see exactly what types of failures occur during some intended use. Furthermore, even were we permitted to establish some kind of reliability record, there would be no certainty that any understanding of the failures could, in feedback fashion, be translated in any timely way into corrective manufacturing changes. Our only course of action is to anticipate the worst and ask how the traditional approaches can be improved, since it is likely that the success of the entire fiber-optic undersea cable project will depend importantly upon the quality of the reliability assurance strategy employed.

Two of the potentially serious problems connected with any uncritical application of the bathtub strategy, (A) and (B) above, are exam-

ined in Section III. Using GaAlAs/GaAs laser reliability data as a guide, we show in our analysis that we must anticipate that the hazard rate curves for the infant and wear-out populations may overlap considerably so that the bathtub will not have any obvious flat-bottomed period of useful life. Thus, even if infant devices were successfully screened out, wear-out-type failures, perhaps in significant numbers, might be expected to occur during system life, and therefore (A) seems ruled out. With respect to (B), the usual procedure for promoting infant failures prior to deployment of the product is to accelerate their occurrence by operating the devices for a short period of time at an elevated temperature, the so-called elevated temperature *burn-in* screen. In Section III we emphasize the possibility, again supported by GaAlAs/GaAs failure data, that low thermal activation energy infant failure modes may exist that are invulnerable to detection in elevated temperature operation for a short duration. Therefore, it would be imprudent to rely exclusively upon a burn-in to accomplish (B).

Let us return briefly to the problem of wear-out failures occurring during system life. Assume for the moment that the wear-out failures occur suddenly and without any warning. An important addition to reliability assurance in the presence of this problem, which defeats (A) above, is the lognormal statistical approach, discussed in Section IV. It assumes that the wear-out failures are temporally distributed according to some law, in this case the lognormal. The hazard rate curve can be estimated from experimentally estimated values of the median lifetime and the variance. In a nonreplacement application, the hazard rate curve can be used to construct a redundancy scheme to allow for a certain percentage of failures during the system life. Thus in the event that the wear-out failures are unaccompanied by any detectable precursor degradation, the best reliability assurance strategy would appear to be a combination of the bathtub and lognormal approaches, to wit: (A') statistically characterize the wear-out of a population to determine the expected number of failures during the system life and then provide sufficient redundancy, and (B) censor infant devices.

An obvious and unavoidable weakness of the bathtub and lognormal approaches, in the case of sudden wear-out failures, which is shared by any sampled-population or population-inferential approach for high-reliability applications, is that conclusions about the behavior of as-yet unmanufactured devices must be based upon aging to demise some early-manufactured population. Given the inevitable inadvertent and premeditated changes in growth and processing of devices, the assumption that all populations are statistically equivalent, no matter when manufactured, may not always be correct.

Beyond our tutorial critique of the reliability assurance strategies which have been used in the past, our forwarding contribution (Section V) to improving reliability assurance comes from the recognition of three properties of lasers: (1) lasers are prone to possess low thermal activation energy modes of failure; (2) lasers may also have initial modes of degradation which eventually saturate or stabilize and which are not indicative of long-term behavior; and (3) lasers, with rare exceptions, do not fail suddenly; instead they invariably degrade slowly. Consequently, we shall describe two procedures that we believe must be incorporated into any laser reliability assurance program. The first relates to the low thermal activation energy early or infant failure mode, which is not particularly susceptible to the traditional burn-in employing temperature as the degradation accelerant or driving mechanism. It is crucial to employ considerable overstressing, as well, of the injection current and optical power in order to purge (screen out) devices which possess this potentially devastating failure mode. Overstressing of current, optical power, and temperature are also considered important to compel rapid stabilization of certain initially occurring modes of degradation in order that the activation energy and degradation rates of the long-term wear-out mode be credibly determined in a timely fashion. The second thrust is to attempt to estimate lifetimes by monitoring the thermally accelerated wear-out rate of degradation of every individual laser that survives the over-stress regimes (designed to eliminate infants and stabilize transient modes of degradation). Truncation of the surviving population will be accomplished by removing lasers whose initially observed long-term rates of degradation are too large.

Relative to the concerns of this paper, we call the reader's attention to several other papers in this volume. A successful implementation of the proposed strategy (purging, stabilization, truncation) for lasers is described by Nash et al. The truncation procedure is discussed more generally and quantitatively in papers by Joyce et al. and Eckler. Purging as a procedure for establishing the reliability of detectors is demonstrated by Saul et al.

II. THE BATHTUB MODEL AND DEFINITIONS OF SOME RELIABILITY TERMINOLOGY*

2.1 Description of the bathtub curve

In discussions of the reliability of products, it has been a common practice to plot the device hazard (failure) rate as a function of time. Imagine that a number of devices have been placed on test at $t = 0$

---

* The Glossary in the Appendix may be consulted for definitions of some reliability terminology also given in the text.

and that the failures are noted as time progresses. If, after the lapse of a time t, the expected number of survivors is N, then the rate at which devices can be expected to fail at time t is $-dN/dt$. The hazard rate is a normalized instantaneous failure rate and is defined as

$$R_N = -\frac{1}{N}\frac{dN}{dt}.$$ (1)

This can be shown[1,2] to agree with the conventional definition.[3] A slightly idealized version of a plot[4] of R_N versus time for human mortality, using data on which life insurance companies base computations of premiums, is shown schematically in Fig. 1a. This so-called bathtub curve for ages 0 to 100 years has the following interpretation. If a person survives the diseases of infancy (infant failures are associated with the decreasing hazard rate up to ~5 years), and then survives the accidents that cause demise in midlife (midlife failures are associated with the roughly constant hazard rate in the range 5 to 50 years), that person will die because of old age (wear-out failures are associated with the increasing hazard rate in the period >50 years). The useful lifetime of a surviving individual is the period up to ≈70 years.

The bathtub curve has been widely employed[1,3,5-10] as a conceptual device to model the reliability of mechanical, electrical, and semiconductor components and systems. It is intuitively obvious that the reliability of light bulbs, brake linings, tires, shoes, etc., will be described by such a curve.* Experimentally, bathtub-like hazard rate plots have been found for radar sets,[5] vacuum tubes,[11,12] resistors,[13] capacitors,[14] electronic equipment,[15] inertial guidance systems,[16] airplane engine starting systems,[3] and bipolar integrated circuits.[17†] We shall assume for the present that Fig. 1a is a plausible depiction of the hazard rate of some semiconductor devices in the course of an intended use. It is understood that none of the devices was subjected to any preoperation screening (e.g., burn-in‡ or accelerated thermal aging),

---

* Consider candles as an example. Infant failures are those candles which are made too long, too short, broken, or without wicks. Barring accidental breakage, all candles similarly made and burned are expected to live a well-defined period of time before wear-out (no more wick).

† For the first-made devices in any new and probably poorly controlled semiconductor device technology, the bathtub might not hold any water, i.e., it might not have a minimum, because of the presence of a multiplicity of failure mechanisms, each operative in different time frames, and each with a large variance in its temporal occurrence. After considerable effort using the physics-of-failure[18,19] approach (analysis of failed devices to discover the failure mechanism and feedback of this information to generate suitable controls for each mechanism by design changes or process control), a bathtub curve may result. Some infant mortality is invariably still present, and wear-out is inevitable.

‡ We define burn-in to mean normal operation at the intended use temperature. An elevated temperature burn-in will be equivalent to accelerated thermal aging.

for otherwise the infant failure population would have been largely eliminated.

2.2 Wear-out (preordained failure)—gradual or sudden

Thus far we have made one kind of distinction among device failures, i.e., when failure occurs (infant, midlife, wear-out). Another convenient *when* term is premature failure, which is inoperability prior to some desired application lifetime. This is a system-lifetime rather than a device-lifetime term. It is useful because the actual lifetime of an individual device is irrelevant if it greatly exceeds the system lifetime. Included in the premature class may be the infant, midlife, and early wear-out failures of the bathtub curve. We now wish to make another kind of distinction that relates not to when failures occur, but rather to why. We call the two classes preordained and chance.* The dominant long-term (wear-out) mechanism causes devices to degrade "deterministically at a rate preordained" but unknown at the moment of birth.[20] "The laser-to-laser variations in lifetime are the result of random slice-to-slice and laser-to-laser growth and processing variations which result in initially inequivalent lasers."[20] The distribution of the lifetimes of wear-out failures has often been modeled as being lognormal† (i.e., Gaussian in the logarithm of lifetimes) for semiconductor devices, e.g., germanium and silicon transistors,[2,18,21-23] low-noise GaAs Field-Effect Transistors (FET)s,[24] power GaAs FETs,[25] GaP Light-Emitting Diodes (LED)s,[26] GaAs LEDs,[27] and GaAs lasers.[20,28-33] The basis for this behavior is not understood,[20,34] although a proportional-effect model has been proposed.[35]

Preordained failures, of which wear-out failures are examples, may be either gradual or sudden (catastrophic); these are *how* terms. Thus, one speaks of gradual degradation in the sense in which some measurable parameter changes slowly or gracefully with time, and by monitoring that parameter a time-to-failure may be predicted.[20,24-26] The essence of the failure is a change in one or more parameters which exceed some specification. On the other hand, one may speak of catastrophic failure.[24-26] In this case the emphasis is upon the suddenness and the complete inoperativeness of the device as contrasted with a slow change that eventually exceeds some specification. Electrolytic corrosion of contacts[24] and burnout[25] of FETs are examples of sudden

---

* Chance or random failures will be discussed in Section 2.3.

† The Weibull model[3,36,37] has also achieved popularity. It is difficult to distinguish between Weibull and lognormal distribution functions; differences become significant only in the tails of the distributions, but actual observations of times-to-failure are sparse in the tails because of limited sample sizes.[38] "[T]he only practical way of progressing is to choose a simple function, test it empirically, and stick to it as long as none better has been found."[11,36]

failures. The suddenness refers to the fact that the complete failure occurred somewhere in between two times of observation; the actual failure might have occurred in a millisecond or a day. The failure was, nonetheless, still preordained, i.e., due to some initially present device weakness, but it was inconvenient or impossible to monitor the parameters that would have enabled failure to be anticipated. Even had monitoring occurred, the changes might have been too small to detect in the incubation period prior to reaching some threshold at which, e.g., a thermal runaway occurred.[25]

When the distribution of lifetimes is a Gaussian and the independent variable is the logarithm of time, the corresponding hazard rate curve is asymmetric with a positive skew.[21] Figure 1b depicts the hazard curve for a lognormal wear-out distribution where the ordinate is linear.

2.3 Midlife (chance) failure

In contrast to the preordained[20] or time-dependent-type[39] failure there is the so-called chance or random failure,[20] which has also been called event dependent.[39] The chance aspect exists in factors external rather than internal to the device. External events,* the most likely of which are electrostatic discharges[39] and random accumulations of operating stresses (e.g. current, mechanical vibration, etc.) that exceed device capacity,[1,7] can cause these failures to weak (susceptible to infant failure) and strong devices alike, and will occur even if all devices are initially equivalent in *all* respects. No preinstallation screen, burn-in, or accelerated thermal aging is capable of identifying which devices have a potential for chance failure, since all devices have the same potential. By definition, the chance failure does not display any symptoms of prior deterioration.†

By analogy to spontaneous radioactive decay, chance failure is usually associated with the Poisson or exponential distribution.[20] Thus, if at time t, the expected number of survivors in a population, whose initial number was N_0, is given by $N = N_0 \exp(- t/\tau_c)$, it follows from (1) that $R_N = \tau_c^{-1}$, where τ_c is the lifetime against chance failure. A population whose failure is governed exclusively by chance failures with constant R_N is depicted in Fig. 1c. This constant (or near constant) hazard rate is associated with the bottom of the bathtub and the so-called midlife or accidental failures.

---

* For a subcable laser system, an unlikely event would be the creation of current leakage paths by cosmic rays.

† Note that the sudden preordained failure and the sudden chance failure have in common a lack of observed precursor degradation. It is conceivable that a chance mechanical shock could be responsible for starting a gradual failure. Even in such an event, there still would have been no precursor degradation prior to the shock, so that the event is still of the chance variety.

2.4 Infant (preordained or chance) failure

So far, in the course of discussing preordained and chance failures, we have examined the bottom (midlife failures) and right-hand side (wear-out failures) of the bathtub curve, Fig. 1a. The left-hand side is associated with the so-called infant mortalities. It is not expected that the fundamental wear-out mode of degradation, presumably common to all devices, will actually control the failure of every device. On occasion, other mechanisms that are simultaneously and randomly present in the population of devices will cause early (infant) failures, substantially prior to the times-to-failure that would have occurred if wear-out had been the only degradation mechanism operative.

Unlike the wear-out mechanism, which is generally well characterized even though its manifestation is intended to be beyond the desired useful lifetime of devices according to Fig. 1a, the infant failures are poorly characterized[17-19] for several reasons: (1) the $t = 0$ point is arbitrary; (2) the infant failures often occur in such relatively short periods of time after the device turn-on that hazard rate data are difficult to acquire for each small increment of time during operation; (3) for the same reason, it is inconvenient or impossible to monitor a suitable device parameter during the lifetime to distinguish preordained from chance failures; (4) infant failures may not show up in small-sample life tests and large samples may be difficult to obtain in a rapidly evolving technology; (5) the major concern is with knowing the screening techniques required to eliminate the infant class and not with knowing the precise failure distribution; and (6) evidence[17,18] suggests that the infant failures have a number of origins, so that, for example, only a fraction may be susceptible to thermal acceleration. The activation energies associated with infant failures in an integrated circuit are typically low (0.25 to 0.40 eV).[17] The operating time duration in which infant failures occur appears to be $\sim 10^4$ hours.[17,39]

One view[17,18] of infant failures is that they represent workmanship-like defects (e.g., cracked chips, actual or incipient opens or shorts) which require some time of operation to be detected. Such devices are initially substandard* and are called weak because they are destined

---

* There is another class of substandard devices which may be called odd or eccentric. Although they are not obviously connected with premature device failure, their initial characteristics are sufficiently different from the main population to warrant rejection for reasons of eccentricity. This is one of the prices paid for assuring reliability in any demanding application. Imagine that for a particular wire bond, there is an air space over one-half of the potential attachment area. Imagine also that mechanical vibration is largely absent in the intended operating locale. A harsh vibration or pull test might very likely reveal this half-attached bond. To be sure, it might be considered wise to identify the laser as a so-called odd or substandard device, but it is far from obvious that this device would ever fail, during the cable lifetime, much less be a candidate for premature failure.

for early failure. The cause of failure would be a random accumulation of operating stress(es) (e.g., current transients) in excess of the normal, and thus would exceed the capacity of the weak, but not the strong devices. Even though the weaknesses are inherent in the devices, the failure-causing events are random and extrinsic, and hence the failure time distribution will be exponential. To derive a declining hazard rate (the left-hand edge of the bathtub curve, Fig. 1a), assume that the infant fraction of some population is governed by $N_{in} = N_{0,in}\exp(-t/\tau_{in})$, while in the relevant time frame, the remaining number (N_r) of devices do not degrade, or do so negligibly. From (1) it may be computed that

$$R_{N,in} = \frac{1}{\tau_{in}} \; \frac{1}{1 + \dfrac{N_r}{N_{0,in}} \exp(t/\tau_{in})}, \qquad (2)$$

which continually decreases from $t = 0$. Such a function, plotted against the logarithm of time, is depicted schematically in Fig. 1d as a dashed curve.

Two observations from our years of experience with the elevated temperature (70°C) burn-in of GaAlAs/GaAs lasers at normal optical outputs, however, suggest that the characterizations of some infant failures are different. First, the infant failures are most often of the preordained type; the operating current to maintain a constant output increases smoothly and significantly over a short period of time (tens of hours). Second, the times to infant failures are susceptible to characterization by a lognormal distribution.[40]* Early failure distributions found during accelerated thermal aging, and consistent with lognormal characterization, have also been found for transistors[2,18,23] and Complementary Metal-Oxide Semiconductor (CMOS) integrated circuits.[41] In summary, early or infant failures have the potential for being either of the chance or preordained types.

A hazard rate signature for a lognormally distributed infant population is shown in Fig. 1d. It displays an increasing hazard rate region, which to our knowledge has never been experimentally observed. Connected with some of the reasons, already given, concerning the poor characterization of infant mortality, there is another reason why an increasing hazard rate is unlikely to be observed even if the failure distribution were lognormal. This is related to the dispersion (σ) in the logarithm of lifetimes for such a distribution. Figure 2 shows a temporal plot of the hazard rate[21] for $\sigma = 0.5$ (good quality control)

---

* In Ref. 40, the plotted data of Ref. 20 were supplemented by 14 additional early failures, and several additional long-term lifetimes.
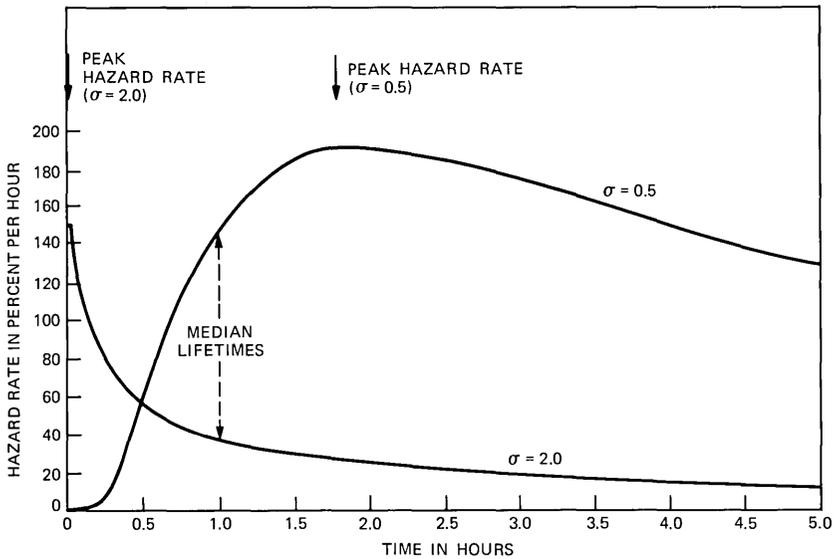
Fig. 2—Hazard rate curves for two lognormal distributions with the same median lifetime ($\tau = 1$ hour) but with different dispersions ($\sigma = 0.5$ and 2.0). If the σ is large enough, the initially increasing hazard rate cannot be observed (see Ref. 21).

and a median infant lifetime, $\tau = 1$ hour. The peak hazard rate occurs at $t_p \approx 1.8$ hours, which on the time scale shown should be easy to detect. With poorer product quality control, a lognormal infant distribution ($\sigma = 2$ and $\tau = 1$ hour) has a hazard rate that peaks at $t_p = \tau/100 = 36$ seconds, a very short period of time in which to record failures, especially with small sample sizes. Therefore, depending upon σ, hazard rate curves can take a form that is practically indistinguishable from that of a decreasing function of time.[21]

2.5 Comparison of bathtub curves for semiconductor and human mortality

A composite of the two lognormal (see Figs. 1b and d) and the exponential hazard rate curves, Fig. 1c, is shown in Fig. 1e. Bathtub-like characteristics are displayed by the practically accessible central region (solid curve). The rollover (maximum in hazard rate) at short time, if it exists, may be impractical to detect. The rollover at long times has been seen in the failures of digital bipolar integrated circuits[17] and inferred for the failures of high-quality submarine cable transistors.[2] In general, wear-out rollover will not be seen in the field use of high-quality devices because it occurs too far out in time. The rollover at long times for the lognormally distributed wear-out of semiconductor devices is a clear departure from the analogy to the wear-out of human beings, which has an always increasing hazard rate and which is typical of a normal distribution. The steady-state or low

hazard rate region (bottom of the bathtub) has been observed[17] and inferred[42] for integrated circuits. As a practical matter, the testing of semiconductor devices is usually terminated at a point (left-hand corner of Fig. 1a) where the hazard rate, though small, is continuing to decline.[19] Since it cannot fall below some random-event-determined level, it is generally taken for granted that were testing continued, an essentially constant* hazard rate would be recorded.[19] A modification of the human mortality bathtub curve (see Fig. 1a), which is similar to our Fig. 1e, has been independently proposed.[37]

2.6 Reliability assurance strategy suggested by the modified bathtub curve

We assume that some particular semiconductor device fabrication technology has been perfected to the point where a bathtub-like curve with a flat-bottomed period of useful life accurately describes the unscreened reliability of devices in field use. No matter how hard one tries to additionally perfect the design, growth, choice of materials, and processing, there will always be at least some infant mortality. Experience has also demonstrated that eventual wear-out appears inevitable for devices which were not infant failures. To the extent that the wear-out times-to-failure are credibly described by a lognormal law, then Fig. 1e, rather than Fig. 1a, is the anticipated bathtub curve; the exact shape of the infant distribution is of minor interest in constructing a reliability strategy, since regardless of its shape, the infant subpopulation must be eliminated prior to deployment.

The traditional strategy for reliability assurance, which is suggested by either bathtub curve, Fig. 1a or e, is the following. Operate (burn in) the unscreened devices in the laboratory, under the field conditions used to obtain the bathtub curve in the first instance, until the hazard rate becomes constant, or falls below some acceptable value, or becomes unmeasurably small. This burn-in procedure (accelerated perhaps by the use of an elevated temperature ambient)[43] is the usual way to censor the infant mortality subpopulation, and is widely and successfully employed.[42,43] The surviving devices are then ready for use. As the time for wear-out approaches, the devices are replaced. If replacement is not possible, then the technology must be sufficiently mature so that wear-out failures are not present during system life. This bathtub strategy does not involve characterizing the wear-out failure distribution, it only requires knowing when wear-out failures start to become important.

---

* In principle, if not in practice, an advantage in using the hazard rate, which is normalized to the instantaneous expected number (N) of survivors is that the minimum of the bathtub appears flat. If the normalization factor had been the total initial population (N_0), both the infant and chance hazard curves would be decreasing functions, making it difficult to distinguish one from the other.

(If the wear-out distribution is characterized, and if it is credibly lognormal, then an alternative strategy is specifically suggested by the rollover in Fig. 1e. The majority of the population which survived the infant screening could be aged to failure. The remaining few devices might be expected to live indefinitely if one were sufficiently far out on the declining right-hand edge of Fig. 1e.*)

2.7 Timely determination of bathtub data—accelerated thermal aging

For a mature technology whose development has produced millions of device hours of testing, data sufficient to estimate the bathtub hazard rate exist. For a new technology under a qualification constraint, not only is there an absence of bathtub data, but there is no time to acquire it directly. To appreciate the time issue, assume that only a long-term gradually occurring wear-out mode of degradation is operative and consider the subcable laser that must operate for at least 25 years = 220 kh in an $\approx 10°C$ ambient. Aging to failure at $10°C$ is out of the question. If $I(t)$ and $I(0)$ are the operating current and its initial value, respectively, and if one defines the lifetime (τ) as the first time at which $I(t) = 2I(0)$, then the lifetime is related to the assumed uniform rate of increase of I by $\tau \, dI/dt = I(0)$. Setting $\tau = 220$ kh shows the initially normalized degradation rate to be 0.45 percent per kh; if $I(0) = 30$ mA, then $dI/dt \approx 0.1$ mA/kh. This would be very difficult to measure in a laboratory-convenient time period. Even if one could age at $10°C$ for 2.5 years, there would remain the problem of predicting behavior for a time period which was ten times longer than the test period in the face of the uncertainty about whether the aging would continue to depend linearly upon time.

An indirect approach must be found to relatively quickly obtain the required data, or at least the confidence that we know what it will look like. Fortunately, semiconductor devices are susceptible to a more sophisticated approach than the brute-force testing for long periods of time at the low-stress conditions of the intended use.† The use of elevated temperatures to accelerate failure[18,22] compresses the useful life-span of a device into the time scale of a laboratory experiment. By comparing the times-to-failure at various elevated temperatures, a thermal activation energy (acceleration factor) can be determined[44] that may permit a reliable estimate‡ of device lifetimes at much lower temperatures where actual measurements would be of intolerably long duration. The implicit assumption is that a unique temperature-time

---

* This particular strategy will be discussed again at the end of Section IV.
† This was the approach taken for electron tubes (see Ref. 18).
‡ The axiomatic difficulty with this procedure is that of showing, for example, that a prediction of times-to-failure which exceed the life-span of the data analyst will indeed come true (see Ref. 2).

relationship with a single value of the activation energy is applicable for all temperatures of interest.

Accelerated thermal aging is a key ingredient of all of the reliability assurance strategies that we shall consider. If a single mode of failure is controlling the demise of all devices at some elevated temperature, then the lifetimes may be modeled by a single lognormal distribution.[2,18,21-33] The absence of any evidence of infant failures, at the accelerating temperatures, in the cited examples[2,18,21-33] may be attributed variously to the smallness of the sample size, special selection, prescreening, or to the fact that the infant failure mode, though not controlling at the accelerating temperature, would become predominant in the relevantly afflicted devices at some lower temperature.* The random selection of large samples appears to more commonly produce a bimodal distribution,[18-20,23,27] an example[20] of which is shown in Fig. 3. Although it characterized the lifetimes of GaAlAs/GaAs lasers operating at 70°C, it strongly resembles the S-shaped lifetime distributions found during the accelerated thermal aging of transistors[18,23] and CMOS intergrated circuits.[41] It is usual[18,23,41] to define the early failure group in accelerated thermal aging as the "freak" population. The remaining larger population is assumed to be controlled by a single longer-term (wear-out) mode, since the relevant failures are lognormally distributed.[18,20,23] Where the point of inflection in the (dashed) curve is sharply defined, as it is in Fig. 3, the freak and wear-out distributions overlap, only slightly, so with some confidence it may be said[18,19,23,40,41] that the freak distribution in Fig. 3 comprised 10 percent of the total population, and that if aging at 70°C had been performed for only 100 hours, the surviving population would be controlled by a single degradation mechanism.† Knowledge of the latter point is of considerable importance, since one of the main goals of accelerated thermal aging is the determination of the activation energy of the ultimately controlling long-term mode, and confidence in the value obtained is heightened if the main population's demise is controlled by a single mechanism.

Freak failures (i.e., the earliest anomalous failures in accelerated thermal aging) have been experimentally found to be lognormally

---

* The fact that elevated temperature aging might not reveal the presence of a potent low activation energy infant mortality mechanism which would be significant at the ultimate device operating temperature will be dealt with in Section 3.3.

† Considerable complexity and possible misinterpretation would exist if more than two different and temporally overlapping failure mechanisms coexisted with similar activation energies. Different combinations of median lives, dispersions (i.e., the variations in the potencies of failure mechanisms from one laser to another), contributing fractions, and activation energies will produce many forms of cumulative failure plots in accelerated thermal aging regimes, and only in the event of bimodality will the sorting out be tractable.
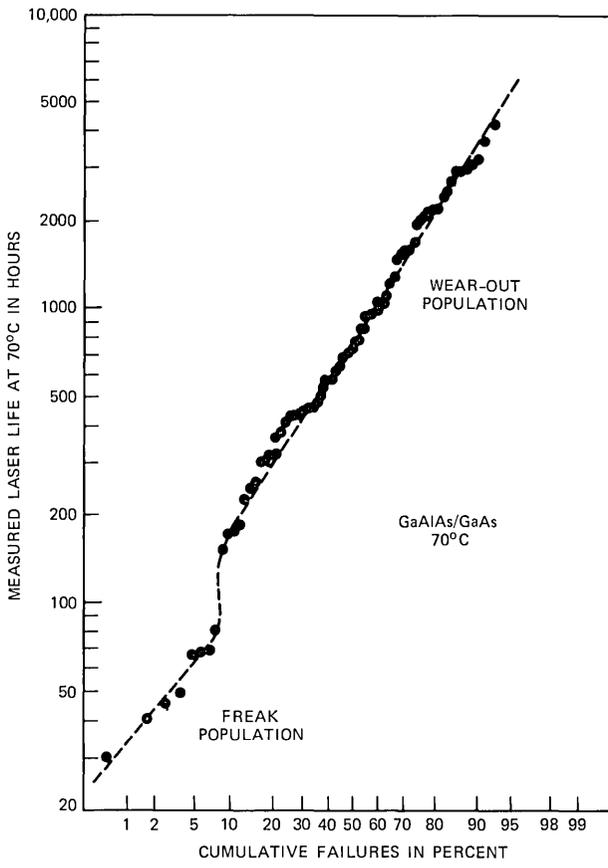
Fig. 3—Wear-out and freak distributions of the failures observed in 70°C ambient aging of (Al,Ga)As lasers (see Ref. 20).

distributed for some of the first commercially available plastic-encapsulated transistors[2] and for lasers.[40] Generally, however, not enough data are available for the unambiguous characterization of freaks, so it is often assumed[18,23,41] that such failures are lognormally distributed.

III. POTENTIAL PROBLEMS WITH THE BATHTUB STRATEGY AND THE USE OF ACCELERATED THERMAL AGING

3.1 Failure to exploit the potential of accelerated thermal aging

When only a single lognormal failure distribution results from aging studies at a single elevated temperature, accelerated thermal aging is only a characterization procedure, i.e., it permits the median lifetime (τ) and the dispersion (σ), which is the standard deviation in the logarithm of lifetimes, to be determined for what is presumptively the

wear-out mode. If failure distributions (each being single) are established at two elevated temperatures, then the activation energy of the wear-out mode may be determined as well.[44] If it is plausible to assume that this activation energy applies at the lower intended field use temperature, then τ may be calculated for this temperature, at which direct measurements are not possible. In the event that it is found that τ exceeds the system lifetime in a high-reliability application, no more than feasibility has been established, a fact that we assumed to exist prior to embarking on our quest for reliability assurance. Knowing that the wear-out mode is tolerably slow acting in some unknown subpopulation of devices is not equivalent to a confident prediction of reliability assurance because of early wear-out failures and the infant mortality problem. The latter, which must exist for some fraction of a population, has not been adequately dealt with in past studies. Alternatively put, the problem of reliability assurance in a high-reliability application is, to an important degree, the problem of identifying those devices in a population for which an early (infant) mode of degradation, rather than the simultaneously present wear-out mode, is controlling at the use temperature.

Single lognormal failure distributions (i.e., absence of any anomalous early failure distribution), established during accelerated aging, probably result from inadequate population sizes,[2,18,21-27,29,31,33] preselection,[28,30] and screening tests.[24,32] Demonstrations of anomalous early failure modes[18,20,23,31,41,45] conform to the expectation that all semiconductor devices (including lasers) are potentially susceptible to non-wear-out failures and that large test populations will be required if elevated temperature aging is to achieve optimum utilization. Aging bimodal populations at several temperatures will permit the activation energies of both modes of degradation, as well as the contributing fractions, median lifetimes and dispersions to be determined.[41] (With few exceptions,[20,40,41] statistically significant data on early anomalous failures have not been recorded.) From these data a hazard rate plot (bathtub curve) may be calculated. If the infant and wear-out distributions are very well separated, as in the ideal case of Fig. 1e, then the bathtub strategy is employed, i.e., burn in the mixed population at some elevated temperature until the infant class has been eliminated, and then install the survivors.

3.2 The problem of overlap in a bimodal failure distribution

An intended use-temperature burn-in test of a large population of nominally good devices will produce, in the absence of preselection or screening, an infant population. Wear-out will generally be unobservable since it occurs largely in the remote future. Accelerated thermal aging of semiconductor devices will significantly compress the time

scale in which wear-out becomes important, because its activation energy is universally found to be large (0.5 to 1.5 eV) and thus wear-out is usually viewed as a strongly thermally activated process. The question to be considered, if a bimodal failure distribution is found at an elevated temperature, is what form this distribution will assume at the lower use-temperature. If the two separate components of the bimodal distribution overlap significantly at the use-temperature, then the bathtub could appear "all filled in." There would be no obvious period of useful life. The traditional bathtub strategy (see Section 2.6) would not work.

The most realistic estimate about whether the bathtub strategy was likely to be applicable to our problem was thought to derive from the best available reliability data[20] on GaAlAs/GaAs lasers, as shown in Fig. 3. An important feature of the lognormal distribution of the long-term failure mode in Fig. 3 is its large dispersion, i.e., the standard deviation in the logarithm of lifetimes[20,40] is $\sigma = 1.1$ to 1.2. It is even larger ($\sigma = 1.3$ to 3.0) for other long-term laser reliability studies.[28-30] The range, $\sigma = 1$ to 3, has, in the past, also been typical[2,18,22,23,41] of thermally accelerated transistor failure distributions. The data of Fig. 3 have been analyzed[40] as a mixture of two lognormal distributions with the following results for the percentages, median lifetimes, and dispersions of the freak and wear-out populations, respectively: 15 percent, $\tau_f = 1$ hour, $\sigma_f = 1.2$; and 85 percent, $\tau_w = 680$ hours, $\sigma_w = 1.25$. By using these values and parameterized failure rate curves,[21] we can translate the lognormally distributed lifetimes[20] of Fig. 3 to the hazard rate plot of Fig. 4. The peak hazard rate for the freaks is ~100 times larger than that for the wear-out group.* The bathtub of Fig. 4 barely holds any water, i.e., it seems filled in. The prime question now is, what is the lower room or use-temperature version of Fig. 4?

Unfortunately, no reliability data from devices comparable to those of Fig. 3 exist at any temperature other than 70°C, so that no bathtub curve can be constructed for 30°C operation. To the extent to which the activation energy of the freak population is the same† as that of

---

* The reason for this follows from an order-of-magnitude estimate in which eq. (1) is applied to each population separately. The hazard rate for the freak population, where the measuring interval is the median freak lifetime, is $[(0.15\ N_T/0.85\ N_T)(\tau_f)]$, since 15 percent of the initial total number (N_T) failed, leaving a wear-out population equal to $0.85\ N_T$. The hazard rate for the wear-out population is $[(0.85\ N_T)/(0.85\ N_T)(\tau_w)]$. The ratio of the former to the latter, with $\tau_f = 1$ hour and $\tau_w = 680$ hours, is 120.

† The claim[17,18] has been made that transistor freaks have approximately the same activation energy (≈ 1.0 eV) as the main (wear-out) population. By inference the freaks are not a subset of the infant group, whose activation energies are lower.[17] Published data[41] for CMOS integrated circuits reveal activation energies for the main and freak populations to be 1.3 and 0.9 eV, respectively. The large activation energy and lognormal distribution for the freak population[41] would suggest that it was an early wear-out distribution.
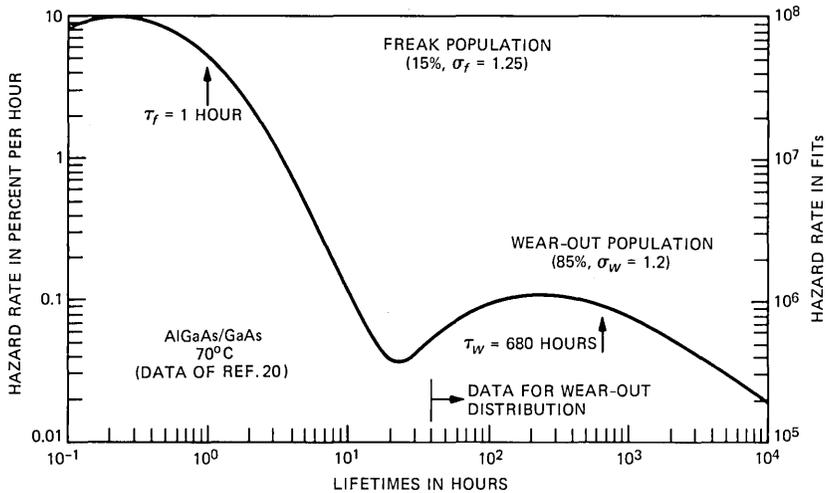
Fig. 4—The hazard rate plot corresponding to the failure distributions in Ref. 20. The contributing fractions, median lifetimes, and dispersions of the two lognormal distributions, which were necessary for the hazard rate calculation, were determined in the analysis of Ref. 40.

the wear-out group, which appears to be approximately true for certain transistors,[17,18,41] then the shape of the bathtub curve at room temperature would be approximately the same as Fig. 4. The substantial variance ($\sigma_w = 1.2$) of the wear-out group would make an early wear-out failure as important as any freak failure in an application in which even one premature failure would be disastrous. The overlap of distributions defeats any strategy for reliability assurance that is based upon accelerated thermal aging until the freaks are eliminated. No matter when the screening is terminated, that point in time will be well into the wear-out distribution, so that statistically, some fraction of the wear-out group will become an early failure group, just subsequent to any installation of the survivors. The usual bathtub strategy would therefore be inadequate in this event.

Although freak failures in the realm of transistor reliability appear to be due to high thermal activation energy mechanisms, this is not thought to be generally true in the realm of laser reliability. For example, for GaAlAs/GaAs lasers (see Figs. 3 and 4), the dominant early failure mechanism is the Dark-Line Defect (DLD).[46,47] In GaAlAs/GaAs LEDs, the DLD have been found[27,48] to be *weakly* temperature dependent, and in optical pumping of wafers, the activation energy was determined to be 0.2 to 0.3 eV.[49] (For InGaAsP/InP lasers, however, mixed results have been obtained; low activation energies [0.16 to 0.34 eV] were found for DLD[50,51] and Dark-Spot Defects [DSD],[51] while activation energies of 0.55 and >1 eV have been found[52,53] for an initial saturable mode of degradation.)

It is likely, then, that had the failure data[20] of Fig. 3 been recorded during operation at room temperature, there would have been considerably less overlap in the room-temperature analog of Fig. 4, since the wear-out mode had an activation energy $(E_a)_w = 0.7$ eV.[54] Suppose, for example, that $(E_a)_f \approx 0.2$ eV. Then the $\tau_f(70°C) = 1$ hour (see Fig. 4) would become $\tau_f(30°C) = 2.4$ hours; in contrast, $\tau_w(70°C) = 680$ hours (see Fig. 4) would become $\tau_w(30°C) = 15{,}368$ hours using $(E_a)_w = 0.7$ eV. Thus, the evidence from studies of specific failure mechanisms would suggest that the so-called freak population of Fig. 4 is just a low activation energy infant population whose hazard rate curve is essentially independent of temperature. Note, however, that even though there is considerably more temporal separation between the hazard rate curves for the infant and wear-out distributions at room temperature, the median wear-out lifetime is only ~2 years. Thus one would still expect a significant number of wear-out failures during some normal system lifetime (e.g., 10 years). The conventional bathtub strategy is not then anticipated to be applicable in our high-reliability use.

3.3 The problem of the low thermal activation energy mode that is not detected in accelerated thermal aging

Low thermal activation energy degradation mechanisms appear to play important roles in early failures of AlGaAs/GaAs and InGaAsP/InP lasers and LEDs.[27,48-51] Since temperature is not an important accelerant for these mechanisms, the possibility exists that the occurrence of these mechanisms in some fraction of a device population would be invulnerable to detection regardless of the temperatures at which accelerated aging was performed. The finally acquired bathtub data might then encourage misleadingly optimistic predictions about performance at a lower use-temperature, when, in fact, behavior was significantly controlled in midlife by an unapprehended mode.

To give one example[2] of what is perhaps the major nightmare of any reliability assurance scheme, consider the schematic distributions in Fig. 5a, which exist in an elevated temperature accelerated aging regime. Two lognormal distributions (dashed lines) are shown. Low activation energy infant failures occur in only 10 percent of the population. This infant distribution is assumed to have a dispersion larger than the long-term (wear-out) mode of degradation, and solely for the purpose of making a point, the activation energy of the infant population is assumed to be equal to zero. The times-to-failure are plotted on the ordinate. The degradation of the remainder of the population (90 percent) is controlled by a lower dispersion, thermally activated long-term mode. If the infant failure mechanism had not existed in 10 percent of the population, then the degradation of that
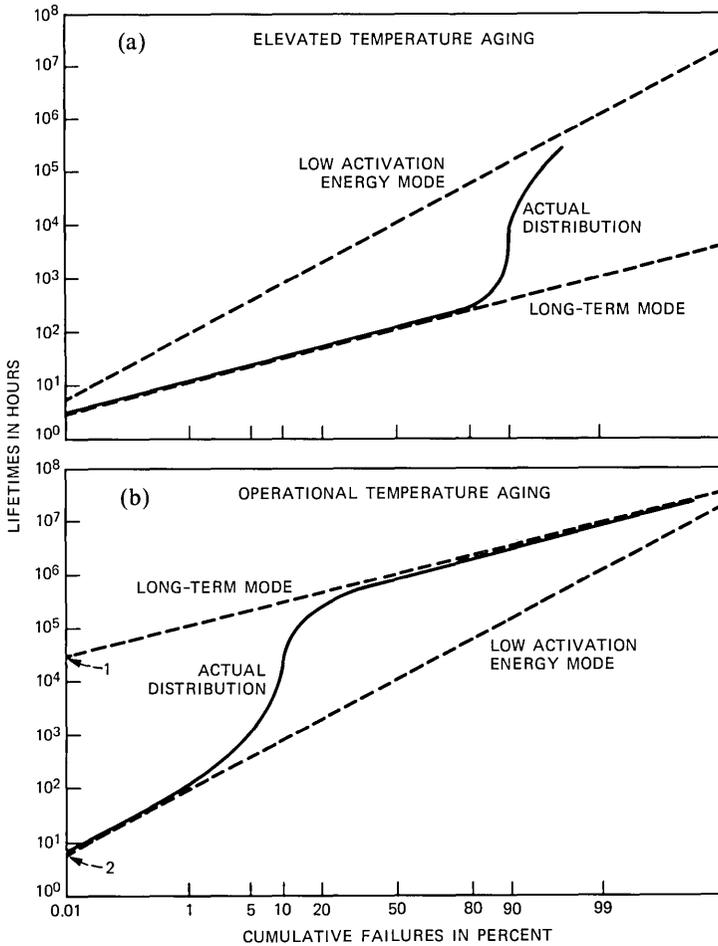
Fig. 5—(a) Hypothetical lognormal distributions (dashed lines) of failure caused by either of two mechanisms during elevated temperature aging. The short-term mode has a dispersion higher than the long-term mode, it is present in 10 percent of the devices, and it is assumed to be nonthermally activated. The failure of 90 percent of the devices is governed by the long-term, or wear-out, mode which is assumed to be thermally activated. Overall, degradation is dominated by the long-term mode beyond the 50-percent cumulative failure point. The short-term mode will remain undetected if testing is terminated at this point. The actual, or composite, distribution is shown as a solid curve. (b) At the intended use or operation temperature, the unaltered short-term mode actually dominates the earliest failures (2), even though it was predicted from (a) that the earliest failures would occur at (1).

10 percent would also have been governed by the long-term mode. A reasonable screening scenario would terminate the elevated temperature testing when the cumulative total failures had reached 50 percent or just beyond.[25] The actual distribution is shown as a solid curve. From the failure results up to the median lifetime (100 hours in the

example), it might have been concluded that a single failure mechanism, the one with the lower dispersion, was controlling.

By going to still higher temperatures, with another lot of nominally similar devices, the lower dashed curve would be downshifted to shorter times, and an activation energy would be found in the usual way.[44] The upper dashed curve would remain fixed. Extrapolation to lower operating temperatures would lead to a prediction that the first failure would occur after 10^4 hours* (see Fig. 5b). The distribution that would actually be seen if a population were aged for a long time at the intended operating temperature (solid curve in Fig. 5b) would, however, show that 10 percent of the devices failed prior to 10^4 hours.

The elevated temperature burn-in has been widely and traditionally used to compel a timely identification of infant failures prior to deployment.[43] Conceptually, at least, as we have noted, this has a mindless aspect in view of the knowledge that the infant failure modes of integrated circuits,[17] LEDs, and lasers[46,51] can have low thermal activation energies. The conceptual error is in the exclusive use of temperature as an accelerant for degradation or failure mechanisms that are relatively insensitive to temperature. Our discussion of Fig. 5 has shown that an infant mode may not be detected in an elevated temperature burn-in. Figure 5a also reveals that an elevated temperature burn-in, used as a screening procedure, may produce a complete disaster, even though it was a success in another case (e.g., the GaAlAs/GaAs data of Fig. 3). Thus, if the hypothetical lasers of Fig. 5a had been burned-in until the point of inflection of the S-shaped curve, with the view that all failed lasers were infants and all survivors were long-lived, the actual result would have been that all long-lived lasers would have been eliminated by the screen and all of the survivors would have been infant failures at the lower operation temperature.

An alternative view of the low activation mode problem appears in an Arrhenius plot. It is common[2,18] to assume that the Arrhenius law governs the dependence of median device lifetimes upon temperature (other degradation accelerants being held constant) through

$$\tau = C \exp\left(\frac{E_a}{kT}\right), \tag{3}$$

where E_a is the activation energy, k is Boltzmann's constant, and C is a constant independent of temperature.† Rearranging (3) gives

---

* This statement is true if the initial population contained 10,000 lasers, so that the first failure would occur at the 0.01-percent cumulative failure point.

† The Eyring equation in which C is a function of temperature and which includes other multiplicative exponential functions that take into account the role of other accelerants, and their interaction with temperature, has been used (see Refs. 41 and 55). For our purposes, (3) is adequate.

$$\frac{1}{T} = \frac{k}{E_a} \ln \tau - \frac{k}{E_a} \ln C, \tag{4}$$

which when plotted as $(1/T)$ versus $\ln\tau$ gives a straight line. For convenience, we desire to retain the straight line plot, but choose to label the ordinate in T; in this case the ordinate is *not* linear in temperature. A schematic plot of the median lifetime of the wear-out mechanism as a function of T is shown as a solid line in Fig. 6a. The darker portion, bounded by the two test temperatures (T_1, T_2), chosen so that failures occur in laboratory-convenient times, is the part actually determined experimentally; all else is extrapolation. Were wear-out the only degradation mechanism present, lifetimes of 10^6 hours at 10°C would be hypothetically possible.

Suppose, however, that in certain devices there also exists a degradation mechanism that has a low activation energy for thermal acceleration. Three examples of degradation behavior (differing potency) associated with such a low activation energy mechanism are shown as nearly vertical lines in Fig. 6b. These lines are not composite curves, but are drawn as though each low activation energy mechanism was the only degradation mechanism present. The presence of a mode whose potency is represented by curve a interferes with the determination of the activation energy of the wear-out mode. The determination of the activation energy for curve a would lead to predictions of unacceptably short lifetimes at the intended use-temperature. The existence of some devices, not affected by a curve a, indicates that a mode represented by curve a is not fundamental. Curve b does not prevent the determination of E_a, but the undetected presence of curve b would mean that lifetime predictions based upon extrapolation of the wear-out curve to 10°C would be too optimistic. Curve c is benign.

It is important to emphasize the fact that the presence of curve b in Fig. 9b is invulnerable to detection by any accelerated thermal aging strategy. For the temperature range in which it could be the controlling degradation mechanism, aging takes an impractically long period of time, and it consumes a considerable fraction of the useful lifetimes of good devices. Insidious low activation energy modes that can invalidate lifetime predictions based on high temperature aging have been contemplated[2,24,56] and observed.[25,57,58] In laser reliability studies, evidence[33] suggests that different degradation mechanisms can dominate at different temperatures.

For completeness, the situation pictured in Fig. 6c should be anticipated, i.e., nonfundamental modes whose activation energy exceeds that of the wear-out mode. A family of such modes with varying potency is shown as curves d, e, and f. Curve d is an early failure, curve f is without consequence, and curve e would obscure any deter-
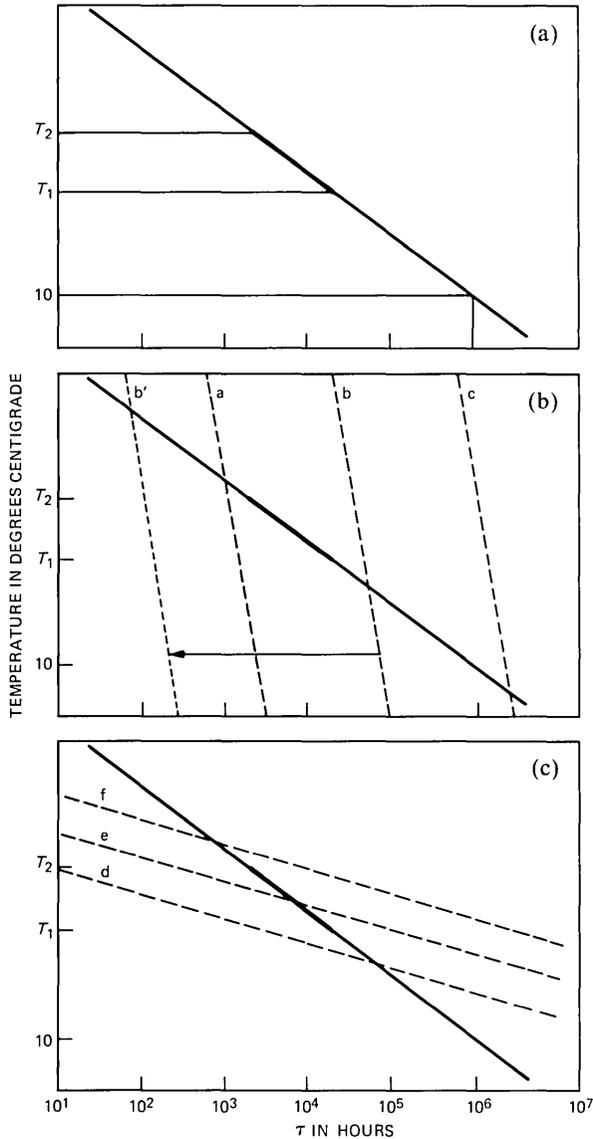
Fig. 6—(a) Hypothetical plot of the median lifetimes of wear-out-mode-dominated failures as a function of the device temperature. The actual lifetime measurements to determine the activation energy (E_a) for this thermally activated long-term mode are conducted in the temperature range $T_2 - T_1$. The remainder of the curve is established by a straight-line extrapolation of the measurements obtained between T_1 and T_2. (b) Superimposed upon (a) is a family of lifetime curves corresponding to varying degrees of potency of a weakly thermally activated nonfundamental mode, i.e., one whose activation energy is less than that of the wear-out mode of (a). The shift of line b to b' represents a selective acceleration of the nonfundamental mode relative to the wear-out mode. (c) The same as (b), except that a higher activation energy nonfundamental mode with varying degrees of potency is superimposed on the wear-out mode failure curve a.

mination of the activation energy for the long-term mode. The obvious way to prevent curve e from introducing ambiguity in establishing the activation energy is to increase the test temperature range, so that curve e becomes like curve d.

The solution to the undetected low activation energy mode problem is to find its appropriate accelerant. This will be examined later in Section 5.2.

3.4 The problem of the single lognormal wear-out failure distribution with large dispersion

Let us imagine that a mixed population has been successfully purged of the infant subpopulation, which includes all low activation energy modes that might have caused premature failure in field use. The survivors will then presumptively follow a single lognormal wear-out failure distribution. In this section we show that the consequence of a large dispersion for wear-out is that there is no period of useful lifetime for a population, the hazard rate is a decreasing function of time for all times of interest,* and hence the traditional bathtub strategy will be inapplicable.

Consider, for example, the AlGaAs/GaAs laser failure time data[29] recorded at 55°C. The dispersion was found to be $\sigma = 3.0$. It has been assumed[2,20,22,28,30,41] and experimentally[18,32] determined that σ is not a function of temperature when a single mode of degradation is controlling. Using $E_a = 0.7$ eV[29,54] for the wear-out data, the hazard rate curve at 70°C may be calculated and it appears in Fig. 7. This plot should be compared with Fig. 4, where the scales are exactly the same; Fig. 7 represents what might have been recorded if the thermal acceleration had occurred at 70°C. Although the median wear-out lifetime is larger in Fig. 7, as compared with Fig. 4, there is no practical duration of time in Fig. 7 in which the hazard rate is relatively constant, i.e., there is no period of useful life as contemplated by the flat-bottomed portion of Fig. 1e.

The operational significance of using accelerated thermal aging as a first step in a screening technique is that it enables a reasonably prompt location of the first time at which a minimum hazard rate is nearly reached. In an ideal case, this corresponds to the left-hand edge of the bottom of the bathtub, Fig. 1e. At this point, it is hoped, most of the devices are still operative and are ready for installation. By contrast, accelerated aging for the devices represented in Fig. 7 has shown that an acceptably low hazard rate may only occur when all but a few of the devices have failed.† The situation would have been

---

* A hypothetical example of this has already been given in connection with Fig. 2.
† As noted in Section 2.6, this observation suggests an alternative strategy, which will be considered at the end of Section IV.
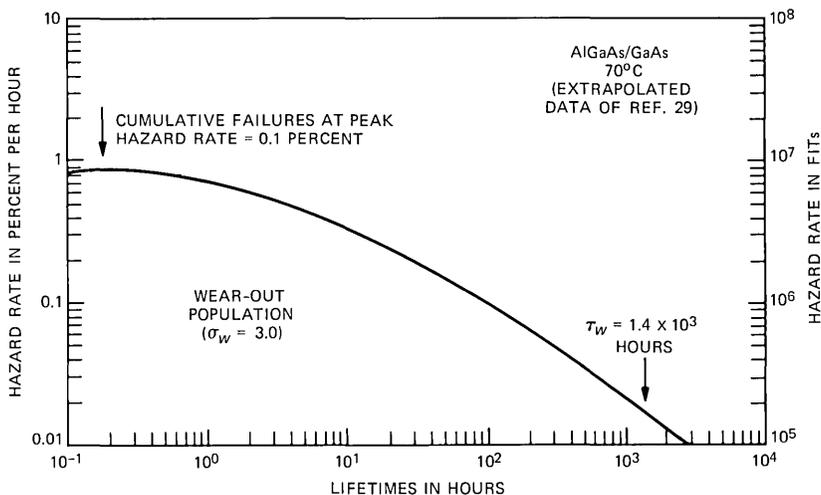
Fig. 7—In Ref. 29 an experimental lognormal failure distribution was established in a 55°C ambient for AlGaAs/GaAs lasers. If those data are extrapolated to 70°C using an activation energy equal to 0.7 eV, the corresponding hazard rate curve (shown above) may be calculated from the deduced values of median lifetime and dispersion.

no better, if instead, the failure data[29] had been accumulated at some reasonable operational temperature, e.g., 30°C, as may be seen in Fig. 8. If, for example, a failure rate of 100 FITs* is acceptable in some 30°C application, then the population represented in Fig. 8 would have to be aged for ten times the median lifetime ($\tau_w = 3 \times 10^5$ hours), at which point 75 percent of the devices would have already failed.

It seems, however, that such data[29] are indicative of poor control in growth and processing. It is much more usual to find[2,18,20-23,30] that the earliest lifetimes appearing in single lognormal failure distributions produced by accelerated thermal aging lie on the increasing portion of the hazard rate curve, e.g., Fig. 4. A particularly good example of the right-hand side of the bathtub curve comes from the 30°C aging of AlGaAs/GaAs lasers. We randomly selected nine lasers from a "tested" wafer and aged them at 3 mW per facet without any prior burn-in. Seven lasers failed, the first at 9365 hours and the seventh at 34,929 hours. Two remained operative at 45,484 hours when the aging was terminated. The lifetimes were lognormally distributed with a $\sigma = 0.54$. All seven data points appear on the increasing portion of the hazard rate curve in Fig. 9. Figures 8 and 9, which have identical

* One FIT is one failure in 10^9 device hours, so that one failure in a population of 10^4 devices operated for 1 year (8760 hours) is a failure rate equal to 11 FITs. For purposes of comparison, 10 FITs = 0.001 percent/kh.
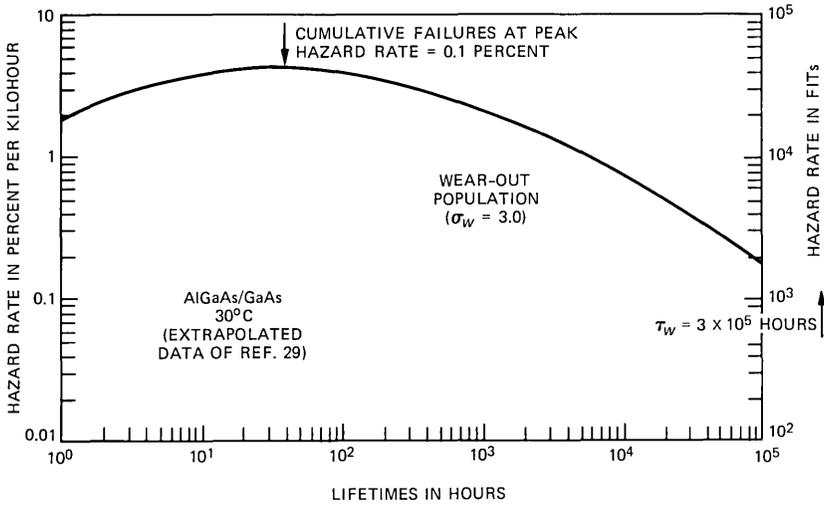
Fig. 8—If the 55°C lognormal failure data of Ref. 29 are extrapolated to 30°C using an activation energy equal to 0.7 eV, the corresponding hazard rate curve (shown above) may be calculated from the deduced values of median lifetime and dispersion.
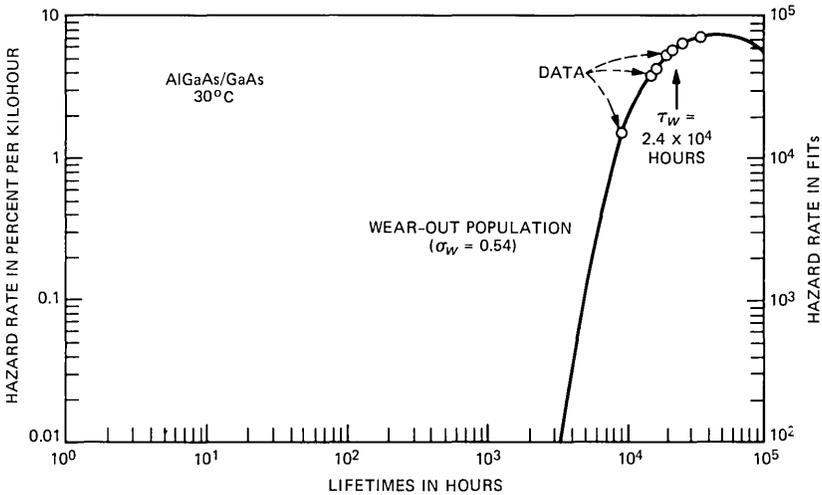
Fig. 9—Failure data (points shown) of AlGaAs/GaAs lasers at 30°C lead to a median lifetime and dispersion which enable the above shown hazard rate curve to be calculated.

scales, should be compared. Thus, there is likely to be at least some period of useful life, prior to the manifestation of wear-out, if σ is sufficiently small; the period of useful life would not appear to be large enough, in general, however, to permit an uncritical application of the bathtub strategy.

3.5 Summary of possible difficulties in the use of a bathtub strategy with an elevated temperature burn-in to censor infant failures

The potential or actual deficiencies of the conventional burn-in approach to using the bathtub strategy for reliability assurance are the following:

1. A reasonably complete distribution of the times-to-failure is required for a large population of devices so that all significant failure modes have some representation in the distribution. This might be impractical to obtain because of the long periods of aging required, even at elevated temperatures.

2. Assuming that it is possible to obtain such a distribution, it will very likely be of no use if it is not bimodal (see, e.g., Fig. 3). Not only will no obvious screening strategy be apparent if the distribution is multimodal, but it will also appear impossible to achieve a segregation of those devices whose lifetimes are governed solely by wear-out. Segregation is needed in order to determine the thermal activation energy for wear-out, which would then permit lifetime predictions at lower temperatures to be made.

3. Assuming that it is tractably bimodal, distributions at other aging temperatures are required in order to avoid the possibility that an elevated temperature burn-in will promote failure in the population that would be long-lived at a lower use-temperature, while retaining for deployment an infant failure group (see Section 3.3 and Fig. 5). If distributions of times-to-failure at low use-temperatures are required, deficiency (1) becomes even more of a problem.

4. Again, assuming tractable bimodality, any overlap of the infant and wear-out hazard rate curves at the use-temperature would mean that some wear-out failures would occur during system life.

5. Assuming that it has been shown that a burn-in at a specified temperature and for a specified duration can remove infant failures in an early-manufactured population, it will remain uncertain whether the burn-in will remain as effective when applied to subsequently manufactured populations. New device technologies evolve rapidly and changes in processing often have a significant impact on reliability.

6. Assuming that the first-established burn-in conditions remain temporally "correct," the time-to-effect may prove to be excessive. In an actual example,[52,53] ~1 kh of elevated temperature burn-in aging was required to establish a stable surviving population. Long screening times might require maintenance of a considerable number of expensive aging sockets. If the duration of the burn-in were to be arbitrarily restricted to a short duration ($\sim 10^2$ hours), disaster might follow.[52,53]

7. Finally, it is noted that long-term wear-out modes of degradation have high thermal activation energies, while infant mechanisms have lower activation energies. The only degradation accelerant available

in the conventional burn-in procedure is temperature. However important it may be to compress the life-spans of devices into laboratory-convenient time periods by using an elevated temperature burn-in, it is possible that the wear-out mechanism becomes relatively more accelerated so that a considerable fraction of the useful life of a good device, one free of aberrant failure mechanisms, may be consumed by the usual burn-in screen.

IV. THE LOGNORMAL MODEL OF WEAR-OUT—AN IMPROVEMENT OF THE BATHTUB STRATEGY

As we have seen, wear-out failures are very likely to occur during a laser system lifetime. Consequently, a conventional bathtub strategy will not be applicable, even if it was successful in eliminating the infant subpopulation. A considerable improvement in establishing long-term reliability is provided by the lognormal statistical approach[24-26,34,59,60] that seeks to characterize the wear-out failure distribution by a lognormal failure law. No procedure is provided in this scheme[24-26,34,59,60] for purging the population of those devices whose degradation or failure is controlled by non-wear-out modes. An important assumption is, therefore, that an infant subgroup either did not exist, or was eliminated by a burn-in or some other means. The starting point of this analysis is similar to the case of Section 3.4, where the conclusion was that if σ (wear-out) was too large, the simple bathtub strategy failed. The natural extension provided by the lognormal statistical approach uses the hazard rate curve computed from experimentally measured values of the median lifetime τ and σ. The confidence level of the hazard rate predictions made for early time failures, where no data may exist, follows from the application of known statistical theory.[60] One can then construct replacement or redundancy schemes that allow for significant numbers of early failures if σ is large. The uncertainties of the predictions based upon inadequately large test populations have been noted.[59]

In the event that early wear-out failures are significant enough to thwart the use of the traditional bathtub approach, as assumed in this section, then the bathtub and lognormal approaches could conceivably be combined as follows. Burn in the starting population under ordinary operating conditions until the hazard rate becomes too low to measure. Presumably at this point, the infants have been eliminated. By means of accelerated thermal aging, operate the remainder of the devices until they all fail, and characterize the presumptively single lognormal wear-out failure distribution (i.e., determine E_a, σ, and τ). Then calculate the hazard rate curve for the use-temperature and extend it back in time to make statistical predictions about early wear-out

failures, i.e., the failures that would have occurred soon after installation if there had been no infants.

The traditional bathtub strategy, which relied upon a useful life-span in which no wear-out failures occurred, does not require that the wear-out failures be characterized. When early wear-out failures can cause a problem, some kind of statistical characterization of the wear-out population becomes imperative, so that compensating redundancy or replacement schemes can be instituted. This characterization approach is the only useful procedure in the event that wear-out failures are of the sudden or catastrophic kind,[24-26] because of the absence of any warning, i.e., any recognizable precursor degradation. (In the next section we shall see that laser wear-out failures are *not* of this type and thus an alternative strategy may be employed.) The lognormal approach[24-26,34,59,60] to characterization that we have highlighted is only one in a class of sampled-population or population-inferential approaches, any one (e.g., Weibull) of which might do equally well.

Despite the improvement in reliability assurance provided by the addition of the lognormal approach, or any similar population-inferential scheme, there remain several shortcomings thereof, which we hope to circumvent by our proposal (see Section V). These drawbacks are (1) the hazard rate curve for the wear-out mode is extrapolated to times less than those at which the first actual failures of the test population would have occurred at lower use-temperatures, thus demanding a complete prior elimination of infants if the resulting predictions are to be credible; (2) the failure predictions are strictly valid, if even then, for only the aged-to-failure population, and their value will be more uncertain for a subsequent population scheduled for installation, which although nominally similar, may be quite different because of inadvertent processing variations; (3) test populations that are small by necessity reduce the confidence of the lognormal approach, partly because infant mortalities will not be apparent in the failure distribution, and partly because uncertainties in the determinations of τ and σ can have a large effect on calculated hazard rates;[59] and (4) in the event that the σ (wear-out) is large, and τ at the use-temperature is comparable to the desired system lifetime, the inevitable statistical predictions of a large early wear-out failure rate could well prevent, perhaps unnecessarily, the system installation.

Objection (4) above relates directly to Figs. 7 and 8, in which the hazard rates might be excessive for all periods of time shown. A system could be viewed as impractical because of the expenses associated with redundancy or replacement. As noted at the end of Section 2.6 and in Section 3.4, this objection to system installation might be avoided by adopting the following procedure. If characterization of wear-out failures is performed until all devices in an adequately large population

suffer failure, and if the failure law is credibly lognormal for all times, then it is always possible to increase the mean life to any extent desired by continuing to age the population until a sufficiently large number of the devices have failed.[61] Beyond the time at which the wear-out hazard rate is a maximum—in, for example, Fig. 8—the hazard rate decreases indefinitely. Thus, if one is willing to age devices for long periods of time prior to deployment, suffer failure of a major fraction of the initial population, and believe that the lognormal law is accurate in the extreme long-time tail of the distribution, any or all of which could be viewed as unacceptable, then the survivors would have a negligibly small hazard rate and be suitable for installation. As a practical matter, this would appear to be an economically unattractive way to circumvent objection (4).

V. THE PROPOSED APPROACH—STP (STABILIZATION, TRUNCATION, PURGE)

Using ideas that were anticipated by efforts on previous submarine cables,[62-65] we want to establish a plan that will permit the confident prediction about which particular lasers in a given initial population will possess lifetimes which exceed that of a system in which there is no redundancy and where replacement is impossible. It should be applicable to small initially available populations and make optimum use of existing knowledge of failure modes. Except where the scheme detects flaws whose elimination by alterations in the device fabrication is feasible in a timely fashion, the plan is of the black-box type, in that it should not rely on building absolute reliability into the device (an impossible task in any event), via the physics-of-failure approach, but should instead screen all devices intended for installation, in a selective and tailored manner, to eliminate all premature failures; i.e., all non-wear-out and all early wear-out failures. A road map of the proposed scheme is given in Table I.

5.1 Stabilization

We recognize that for lasers, certain modes of degradation (annealing) are transient in nature,[53,66] i.e., after some period of aging the degrading centers may become spatially homogenized (the defects may be gettered), and these modes may exist to some extent in all lasers. We understand that one purpose of a reliability assurance scheme is the stabilization[53,66] of transient modes. In general, devices may not be eliminated from the population by the process of stabilization. However, if in the course of stabilization, which may involve varying degrees of a temporally finite amount of degradation, certain initial properties, e.g., threshold, are caused to exceed specification, those devices will be eliminated, even though thereafter they may be capable

Table I—The stages, purposes, and procedures by which laser reliability assurance is established. The overall goal is uniformity of behavior, i.e., to obtain a population whose degradation, in the simplest case, is governed by a long-term wear-out-type mode, and which contains only the tolerably slow-acting manifestations thereof

Regimes for Establishing Laser Reliability Assurance, and Uniformity of Aging Behavior		Purposes	Procedures
Stabilization		To stabilize, in a timely manner, transient modes of annealing and degradation, which may be present to some extent in all lasers	Overstress aging (current, temperature, optical power, reverse bias, etc.)
Rejection	Purge	To eliminate, in a timely manner, potential premature failures due to all modes of degradation, other than the long-term wear-out mode.	Cosmetic inspection Initial and final L-I-V characterization Overstress aging (etc.)
	Truncation	To eliminate premature failures of the wear-out type	Rate monitoring

of adequate longevity. A procedure to compel rapid stabilization is important[53,66] so that unambiguous determinations may be made of the activation energy and degradation rates of the longer-term wear-out mode. The procedure for forcing a timely stabilization of some mode is similar to one regime of the purge, which we consider next.

5.2 Purge

We understand the purge[53,66] regime to include any method (cosmetic inspection, light-current characterization, overstress aging, etc.) by which actual or potential non-wear-out failures (flawed devices) are eliminated from the population. Emphasis is placed upon overstress aging,[53,66] and in particular upon three component concepts—harshness, selectivity, and tailoring.

The first is that of harshness. There is often present in the testing of new, initially expensive, and perhaps few in number, semiconductor devices, the "don't hurt my baby" syndrome. In our view, such an attitude is inconsistent with providing reliability assurance in a timely fashion. Well-made lasers can withstand operation at currents equal to 50 times threshold without adverse consequences.[53] Excessive stress levels, however, will destroy both potentially good and bad devices. The problem is to find the critical level. When found, stressing can be performed at slightly reduced levels. This procedure should produce the most robust class of survivors, in the shortest time. While temperature overstressing is well known,[18,24,25,27,41,48,51] current overstressing in LEDs,[27] current (or optical power) overstressing in lasers,[50,51,67] and voltage overstressing in photodiodes[68,69] have also been used with success in establishing reliability assurance by eliminating early failures. Overstressing using temperature cycling, thermal and mechanical shock, and humidity is well established in the semiconductor industry[8,41] as a means of producing a robust population.

Another concept is that of selectivity.[56,66] The physical basis for any reliability prediction is related to the fact that the failure of a semiconductor device is caused by interactions between the device and its electrical, thermal, and mechanical environment. Unbiased devices in hermetic packages generally last for an extremely long time at room temperature in a tranquil setting. The concept of a mode of degradation that is not accelerable by some means is therefore physically untenable. Thus, one function of a reliability assurance scheme is to discover degradation or failure driving mechanisms or accelerants, other than temperature, that will cause, if possible, a low thermal activation energy mode (curve b in Fig. 6b), which may be undetectable in a pure thermal acceleration regime, to become the dominant mode of degradation[56] (curve b' in Fig. 6b). The notion of *selectivity** refers to

---

* Selectivity also comprises the idea that particular attention should be paid to those stresses that are likely to be present in the operating environment.

the fact that the chosen stress will accelerate the low activation mechanism, but not accelerate the long-term, higher-activation energy mode as much. This expectation is reasonable for AlGaAs/GaAs lasers because, for example, the low thermal activation energy mechanisms are strongly current-density dependent,[27,48] while the long-term, gradual wear-out mode is very weakly current-density dependent.[48,70] Similar strong current-density dependence of DLD and DSD formations have been found for InGaAsP/InP devices;[51,71] it is reasonable to suppose that the wear-out mode may be analogously weakly current-density dependent.

A third concept involves the idea of tailoring the stresses. The approach that should be taken to identify the premature non-wear-out failure population cannot be a straightforward elevation of every thermal, environmental, electrical, and mechanical stress ever devised. An example of the conceptual inadequacy of such a procedure is the high-current, high-temperature operation of the etched planar mesa, buried heterostructure,[72] in which the current flows around the active region in a shunt path.[53] In order to overstress the active region, lower temperatures and currents must be used.[53]

The purge regime contemplates an application of harsh, selective, tailored stresses (ideally, used at levels just shy of those at which good devices would be annihilated) for only short periods of time ($\sim 10^2$ hours) in order to avoid consuming too much of the useful lifetime of devices. This regime should produce both stabilization of transient modes of annealing and degradation, and the failure of devices whose mortality is controlled by the aberrant non-wear-out mechanisms. Stabilization is a product of the overstress regime of the purge, and may occur in all lasers;[53] it was distinguished in Table I in order to make clear the difference between it and the process of rejecting failures (or selecting the survivors), which is the goal of the purge and the truncation stages.

Unlike the bathtub strategy, the purge does not require that a distribution of times-to-failure be established at an elevated, or any other, temperature. Were a lifetime distribution established, it would be irrelevant, in principle, whether the distribution were tractably bimodal or intractably multimodal. The purge strategy, which anticipates the worst, is to fashion a screening regime so strenuous that it does not require alteration as the processing technology evolves. By employing harsh levels of potential degradation accelerants, in addition to and including temperature, the duration of the purge may be kept short.

5.3 Truncation

A fact which is of great significance, and the basis for the truncation

selection scheme, is that the wear-out modes in lasers are characterized by gradual degradation rather than sudden failure, as is the case with other semiconductor devices.[24-26] This frees us from the limitations of the lognormal statistical approach, and permits the use of the rate monitoring[20,31,33,45,55,73-75] of degradation (e.g., a recording of the increase in a laser's operating current to maintain a constant output power) to estimate lifetimes of individual lasers.* When the end of life is not accessible even in accelerated thermal aging, rate monitoring is crucial. An obvious advantage of the rate-monitoring scheme exists for the case in which the wear-out lognormal distribution has a large dispersion (see Figs. 7 and 8) and only a moderate median lifetime. Rate monitoring attempts to determine as accurately as possible which survivors of the purge and stabilization procedures will be early wear-out failures and which will outlive the intended life of the system. Provided that yield was not a problem, the individual-interference approach could legitimately favor the installation of a system which the lognormal approach would have condemned. The individual-inference approach is not concerned with determining the τ and σ of the lognormal failure distribution nor with the calculation of hazard rates. It is only necessary that a sufficient number of lasers have individually projected lifetimes which exceed the reliability requirements of the system.[†]

Figure 10 is a simplified schematic in a population-inferential format of what an individual-inference reliability strategy based upon truncation hopes to accomplish. The initial population is assumed to be composed of both infant and wear-out-dominated failure modes. Accelerated aging to failure at elevated temperatures would produce the bimodal distribution, shown as the lower curve. If the same population had been subjected to purge overstressing to remove the infant population prior to aging, the middle curve, a single lognormal distribution, would have resulted. But, if in addition to overstressing for short time periods, a longer time degradation rate monitoring was subsequently imposed to remove the early wear-out failures, then the upper curve, now no longer lognormal, would have resulted. Note that while the median lifetime has not been substantially increased, there is a great

---

* Quantitative procedures for making lifetime estimates based upon observations of early wear-out degradation rates are discussed in Refs. 44 and 76.
† For applications with less than the most demanding reliability requirements (e.g., because replacement and/or redundancy are economically feasible), it may become important to minimize the cost of the device. Elaborate screening is then too expensive.[43] The focus may shift back toward the physics-of-failure approach and procedures to increase yield. The lognormal statistical approach is used even though it is possible, in principle, to monitor the gradual degradation of devices in order to be able to estimate individual failure times.
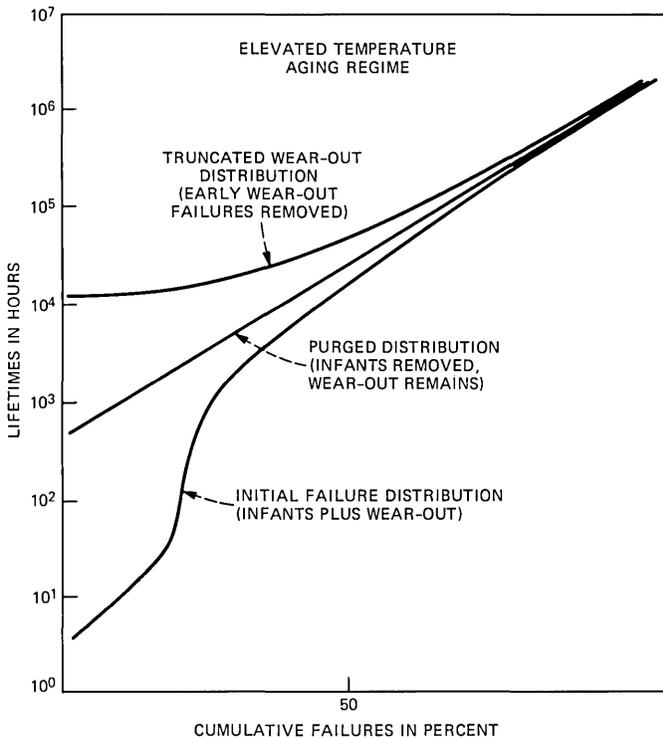
Fig. 10—The S-shaped bimodal distribution (bottom curve) contains both infant (or freak) and wear-out failures. With the infant failures removed as a result of harsh overstress testing, the single lognormal wear-out distribution (middle curve) is expected to occur. Using rate monitoring, to eliminate the early wear-out failures, it should be possible to produce the screened wear-out distribution (top curve).

reduction in the number of failures prior to $\sim 10^4$ hours, the hypothetical system lifetime at the elevated temperature. Rate monitoring, for example, during a short duration elevated temperature burn-in, is not by itself adequate as a reliability assurance strategy since it might not compel a timely stabilization of transient modes, nor might it detect sudden low thermal activation energy infant failures.

VI. SUMMARY AND CONCLUSION

It is understood that the spot checking of lasers is unacceptable as a means to assure reliability. Every laser of initial subcable quality must pass through a screening procedure that schematically consists of two active aging regimes. In the first, the various degradation accelerants (e.g., optical power, current, temperature, humidity, etc.) are elevated, whether singly or in concert, to high levels, for relatively short periods of time ($\sim 10^4$ hours). In this overstress regime, saturable modes of annealing or degradation should become stabilized and non-

wear-out controlled devices should be identifiable by their large degradation rates or actual failures. In the second regime, longer-term aging is performed, with sufficient thermal acceleration so that clear degradation (beyond experimental error) occurs in times $\sim 10^3$ hours. This rate-monitoring stage serves three purposes: (1) It enables the rates of degradation of the survivors of the harsh overstress stage to be compared with the rates of degradation of nonoverstressed lasers so it can be demonstrated that the purge did not introduce a mode of degradation which would not otherwise have been present. (2) It permits determination of rates of long-term degradation due to the wear-out mode, which is present in each laser in different degrees, and hence the extrapolated times-to-failure can be estimated, which permits early failures in the wear-out population to be identified. (3) It may serve to expose any mode of degradation with a long incubation period which is not eliminated by the purge, thus compelling the purge regime to be suitably modified where possible, so that the undetected mode becomes susceptible to identification (corrective feedback). Overall, these regimes will serve to reject some class of lasers and to sort the stabilized survivors.

We desire that the harsh overstress tests be necessary (every aspect of that test should be directed at some mode), sufficient (all possible modes are detected), and restrained (good lasers are not eliminated). We assume that all modes, except for random failures, are accelerable by some means, and hence identifiable. The challenge is to find the correct strategy for the particular device. The questions that should be addressed in any first practical implementation are (1) Does the screening eliminate all weak lasers and those prey to premature failure? (2) If the answer is yes, is too much useful life of good lasers consumed during the screening? (3) Does the purge introduce degradation not otherwise present at the intended operating temperature? (4) Does the purge eliminate lasers which would have been adequately reliable at the intended use-temperature?

The motivation for, and the assumptions of, the STP (Stabilization, Truncation, Purge) strategy may be summarized as follows:

1. The long-term reliability of all lasers is controlled by some wear-out mechanism, which may be processing- or structure-related, as opposed to material-related, and which is tolerable for the intended use even if the wear-out mechanism is unknown, or known, but not susceptible to any reduction in its potency.

2. Some lasers, not identifiable after initial characterization, are additionally afflicted with initial drifting of operating parameters and/or one or more other modes of degradation, which are randomly distributed within the population of as-received lasers, and which cause premature failure in the desired regime of operation.

3. The yield of lasers whose degradation is governed exclusively by the wear-out mechanism is inadequately large so that indiscriminate selection is not possible, especially in view of the replacement costs associated with failed devices.

4. Extensive redundancy is, however, not entirely adequate as a reliability assurance strategy because of the requisite increase in complexity and cost which is associated with sparing.

5. The operating stresses, in particular current, or current and optical power, in concert with temperature cause device failure; the shelf life is long.

6. Elevating the operating stresses will accelerate degradation, and elevation to levels just shy of those which annihilate good devices will do the best possible job of producing a robust population of survivors.

7. The duration of the overstress aging should be sufficiently long to stabilize saturable modes of degradation and annealing and to detect premature non-wear-out failure modes (especially those which have incubation periods), but sufficiently short so that enough useful lifetime remains in the survivors to meet the system requirement.

8. The goal in the preceding motivation can be optimally achieved if the modes of premature failure and stabilizable drift have dependences upon the accelerants (optical power, current, temperature, etc.) which are different from those of the wear-out mechanism, and which would permit selective acceleration.

9. Even if the nonfundamental modes of premature failure have the same dependences on the accelerants as does the wear-out mode, acceleration of aging to identify early failures is still essential, especially if they are of the sudden type.

The goals of the STP strategy are then:

1. To identify, where possible, repairable design, growth, and processing flaws that cause, or might cause, failure during the purge, thereby leading ultimately to an increase in the number of survivors.

2. To stabilize initially existing drifts in operating parameters (saturable degradation and annealing mechanisms) that give misleading indications about long-term reliability.

3. To identify and eliminate devices whose lifetimes are dominated by premature failure mechanisms, whether denominated infant, early wear-out, or otherwise.

4. To produce a class of survivors:

(a) whose degradation is governed exclusively by a tolerably slow acting long-term mechanism;

(b) whose degradation is thermally accelerable and whose thermal activation energy determination is unobscured by the presence of nonfundamental modes of degradation or drift;

(c) which possesses no mode of degradation which was introduced

solely by the purge and which would not have been present otherwise; and

(d) for which degradation rates can be unambiguously established so that reasonable predictions about the time-to-failure in the operating environment can be made for every single member of the long-lived class, unobscured by the presence of nonfundamental modes of degradation or drift.

VII. ACKNOWLEDGMENTS

REFERENCES

1. I. Bazovsky, *Reliability Theory and Practice*, Englewood Cliffs, N.J.: Prentice-Hall, 1961, Chap. 5, pp. 32–5.
2. F. H. Reynolds, "Thermally Accelerated Aging of Semiconductor Components," IEEE Proc., *62*, No. 2 (February 1974), pp. 212–22.
3. R. E. Barlow and F. Proschan, *Statistical Theory of Reliability and Life Testing: Probability Models*, New York: Holt, Rinehart, and Winston, 1975, Chap. 3, pp. 52–80.
4. E. J. Henley and H. Kumamoto, *Reliability Engineering and Risk Assessment*, Englewood Cliffs, N.J.: Prentice-Hall, 1981, Sect. 4.2.1, pp. 161–9.
5. R. R. Carhart, "A Survey of the Current Status of the Electronic Reliability Problem," Rand Corp. Res. Memo. RM-1131 (August 1953), pp. 1–134.
6. D. R. Earles and M. F. Eddins, "Reliability Physics (The Physics of Failure)," 9th Nat. Symp. Reliab. Qual. Cont. (January 1963), pp. 43–57.
7. M. L. Shooman, *Probabilistic Reliability: An Engineering Approach*, New York: McGraw-Hill, 1968, Chap. 4, pp. 159–201.
8. C. G. Peattie et al., "Elements of Semiconductor-Device Reliability," IEEE Proc., *62*, No. 2 (February 1974), pp. 149–68.
9. J. Wood, *Reliability and Degradation: Semiconductor Devices and Circuits*, M. J. Howes and D. V. Morgan, eds. New York: Wiley, 1981, Chap. 4, pp. 191–236.
10. W. Nelson, *Applied Life Data Analysis*, New York: Wiley, 1982, Chap. 2, pp. 15–102.
11. J. H. K. Kao, "A Summary of Some New Techniques on Failure Analysis," 6th Nat. Symp. Reliab. Qual. Cont. (January 1960), pp. 190–201.
12. H. P. Goode and J. H. K. Kao, "Sampling Procedures and Tables for Life and Reliability Testing Based on the Weibull Distribution (Hazard Rate Criterion)," 8th Nat. Symp. Reliab. Qual. Cont. (January 1962), pp. 37–51.
13. A. L. Floyd, "Reliability Testing," 7th Nat. Symp. Reliab. Qual. Cont. (January 1961), pp. 134–43.
14. D. Mannheim and R. Russell, "Accelerated Tests and Predicted Capacitor Life," 7th Nat. Symp. Reliab. Qual. Cont. (January 1961), pp. 253–61.
15. F. M. Gryna, "Basic Statistical Concepts of Reliability," 7th Nat. Symp. Reliab. Qual. Cont. (January 1961), pp. 41–50.
16. L. A. Weaver and M. P. Smith, "The Life Distribution and Reliability of Electromechanical Parts of an Inertial Guidance System," 8th Nat. Symp. Reliab. Qual. Cont. (January 1962), pp. 148–55.
17. D. S. Peck, "New Concerns About Integrated Circuit Reliability," IEEE Trans. Electron Dev., *ED-26*, No. 1 (January 1979), pp. 38–43.
18. D. S. Peck and C. H. Zierdt, "The Reliability of Semiconductor Devices in the Bell System," IEEE Proc., *62*, No. 2 (February 1974), pp. 185–211.
19. A. T. English and C. M. Melliar-Smith, "Reliability and Failure Mechanisms of

Electronic Materials," Ann. Rev. Mater. Sci., *8* (1978), pp. 459–95.

20. W. B. Joyce, R. W. Dixon, and R. L. Hartman, "Statistical Characterization of the Lifetimes of Continuously Operated (Al,Ga)As Double-Heterostructure Lasers," Appl. Phys. Lett., *28*, No. 11 (June 1976), pp. 684–6.

21. L. R. Goldthwaite, "Failure Rate Study for the Lognormal Lifetime Model," 7th Nat. Symp. Reliab. Qual. Cont. (January 1961), pp. 208–13.

22. G. A. Dodson and B. T. Howard, "High Stress Aging to Failure of Semiconductor Devices," 7th Nat. Symp. Reliab. Qual. Cont. (January 1961), pp. 262–72.

23. D. S. Peck, "The Analysis of Data From Accelerated Stress Tests," 9th Ann. Proc. Reliab. Phys. Symp. (March 1971), pp. 69–78.

24. J. C. Irvin and A. Loya, "Failure Mechanisms and Reliability of Low-Noise GaAs FETs," B.S.T.J., *57*, No. 8 (October 1978), pp. 2823–46.

25. A. S. Jordan, J. C. Irvin, and W. O. Schlosser, "A Large Scale Reliability Study of Burnout Failure in GaAs Power FETs," 18th Ann. Proc. Reliab. Phys. Symp. (April 1980), pp. 123–33.

26. A. S. Jordan et al., "The Reliability of 302A Numerics," B.S.T.J., *57*, No. 8 (October 1978), pp. 2983–99.

27. C. L. Zipfel et al., "Reliability of DH $Ga_{1-x}Al_xAs$ LEDs for Lightwave Communications," 19th Ann. Proc. Reliab. Phys. Symp. (April 1981), pp. 124–9.

28. R. L. Hartman, N. E. Schumaker, and R. W. Dixon, "Continuously Operated (Al,Ga)As Double-Heterostructure Lasers With 70°C Lifetimes as Long as Two Years," Appl. Phys. Lett., *31*, No. 11 (December 1977), pp. 756–9.

29. A. Thompson, "The Reliability of a Practical $Ga_{1-x}Al_xAs$ Laser Device," IEEE J. Quantum Electron., *QE-15*, No. 1 (January 1979) pp. 11–13.

30. M. Ettenberg, "A Statistical Study of the Reliability of Oxide-Defined Stripe CW Lasers of (Al,Ga)As," J. Appl. Phys., *50*, No. 3 (March 1979), pp. 1195–202.

31. A. R. Goodwin et al., "Enhanced Degradation Rates in Temperature-Sensitive $Ga_{1-x}Al_xAs$ Lasers," IEEE J. Quantum Electron., *QE-13*, No. 8 (August 1977), pp. 696–9.

32. M. Ettenberg and H. Kressel, "The Reliability of (AlGa)As CW Laser Diodes," IEEE J. Quantum Electron., *QE-16*, No. 2 (February 1980), pp. 186–96.

33. S. Ritchie et al., "The Temperature Dependence of Degradation Mechanisms in Long-Lived (GaAl)As DH Lasers," J. App. Phys., *49*, No. 6 (June 1978), pp. 3127–32.

34. A. Jordan, "A Comprehensive Review of the Lognormal Failure Distribution With Application to LED Reliability," Microelectron. Reliab., *18*, No. 3 (1978), pp. 267–79.

35. J. H. K. Kao, "Statistical Models in Mechanical Reliability," 11th Nat. Symp. Reliab. Qual. Cont. (January 1965), pp. 240–7. (See also N. R. Mann, R. E. Schafer, and N. D. Singpurwalla, *Methods for Statistical Analysis of Reliability and Life Data*, New York: Wiley, 1974, Sect. 4.6.)

36. W. Weibull, "A Statistical Distribution Function of Wide Applicability," J. Appl. Mech., *18*, No. 3 (September 1951), pp. 293–7.

37. F. Jensen and N. E. Peterson, *Burn-in*, New York: Wiley, 1982, Chap. 2, pp. 3–19.

38. R. E. Barlow and F. Proschan, *Mathematical Theory of Reliability*, New York: Wiley, 1965, Chap. 2, pp. 9–45.

39. B. Unger, "Electrostatic Discharge Failure of Semiconductor Devices," 19th Ann. Proc. Reliab. Phys. Symp. (April 1981), pp. 193–9.

40. E. B. Fowlkes, "Some Methods for Studying the Mixture of Two Normal (Lognormal) Distributions," J. Amer. Statist. Ass., *74*, No. 367 (September 1979), pp. 561–75.

41. M. Stitch et al., "Microcircuit Accelerated Testing Using High Temperature Operating Tests," IEEE Trans. Reliab., *R-24*, No. 4 (October 1975), pp. 238–50.

42. Linear IC Reliability, Western Electric Data Sheet Supplement, *LIC 1-1* (September 1981), pp. 1–5.

43. W. Kuo and Y. Kuo, "Facing the Headaches of Early Failures: A State-of-the-Art Review of Burn-in Decisions," IEEE Proc., *71*, No. 11 (November 1983), pp. 1257–66.

44. W. B. Joyce et al., "Methodology of Accelerated Aging," AT&T Tech. J., this issue.

45. H. Ishikawa et al., "Accelerated Aging Test of $Ga_{1-x}Al_xAs$ DH Lasers," J. Appl. Phys., *50*, No. 4 (April 1979), pp. 2518–22.

46. B. C. DeLoach et al., "Degradation of CW GaAs Double-Heterostructure Lasers at 300K," IEEE Proc., *61*, No. 7 (July 1973), pp. 1042–4.

47. P. Petroff and R. L. Hartman, "Defect Structure Introduced During Operation of Heterojunction GaAs Lasers," Appl. Phys. Lett., *23*, No. 8 (October 1973), pp. 469–71.

48. S. Yamakoshi et al., "Growth Mechanism of <100> Dark-Line Defects in High Radiance GaAlAs LEDs," IEDM Tech. Dig. (December 1978), pp. 642–5.

49. H. Imai et al., "Degradations of Optically-Pumped GaAlAs Double Heterostructures at Elevated Temperatures," Jap. J. Appl. Phys., *18*, No. 3 (March 1979), pp. 589–95.

50. K. Endo et al., "Rapid Degradation of InGaAsP/InP Double Heterostructure Lasers Due to <110> Dark Line Formation," Appl. Phys. Lett., *40*, No. 11 (June 1982), pp. 921–3.

51. M. Fukuda, K. Wakita, and G. Iwane, "Dark Defects in InGaAsP/InP Double Heterostructure Lasers Under Accelerated Aging," J. Appl. Phys., *54*, No. 3 (March 1983), pp. 1246–50.

52. B. W. Hakki, P. E. Fraley, and T. F. Eltringham, "1.3-μm Laser Reliability Determination for Submarine Cable Systems," AT&T Tech. J., this issue.

53. F. R. Nash et al., "Implementation of the Proposed Reliability Assurance Strategy for an InGaAsP/InP, Planar Mesa, Buried Heterostructure Laser Operating at 1.3 μm for Use in a Submarine Cable," AT&T Tech. J., this issue.

54. R. L. Hartman and R. W. Dixon, "Reliability of DH GaAs Lasers at Elevated Temperatures," Appl. Phys. Lett., *26*, No. 5 (March 1975), pp. 239–42.

55. K. Mizuishi et al., "Acceleration of the Gradual Degradation in (GaAl)As Double-Heterostructure Lasers as an Exponent of the Value of the Driving Current," J. Appl. Phys., *50*, No. 11 (November 1979), pp. 6668–74.

56. F. R. Nash and R. L. Hartman, "Accelerated Facet Erosion Formation and Degradation of (Al,Ga)As Double-Heterostructure Lasers," IEEE J. Quantum Electron., *QE-16*, No. 10 (October 1980), pp. 1022–33.

57. W. C. Ballamy and L. C. Kimerling, 'Premature Failure in Pt-GaAs IMPATTS—Recombination-Assisted Diffusion as a Failure Mechanism," IEEE Trans. Electron Dev., *ED-25*, No. 6 (June 1978), pp. 746–52.

58. A. T. English and E. Kinsbron, "Electromigration-Induced Failure by Edge Displacement in Fine-Line Aluminum—0.5% Copper Thin Film Conductors," J. Appl. Phys., *54*, No. 1 (January 1983), pp. 268–74.

59. J. C. Irvin, *GaAs FET Principles and Technology*, J. V. DiLorenzo, ed., Dedham, MA: Artech House, 1982, pp. 353–400.

60. A. S. Jordan, "Confidence Limits on the Failure Rate and a Rapid Projection Nomogram for the Lognormal Distribution," Microelectron. Reliab., *24*, No. 1 (1984), pp. 101–24.

61. G. S. Watson and W. T. Wells, "On the Possibility of Improving the Mean Useful Life of Items by Eliminating Those With Short Lives," Technometrics, *3*, No. 2 (May 1961), pp. 281–98.

62. A. J. Wahl et al., "SF Submarine Cable System: Transistors, Diodes, and Components," B.S.T.J., *49*, No. 5 (May–June 1970), pp. 683–98.

63. I. G. Abrahamson, E. T. Lee, and E. B. Slutsky, "Procedures for Selection of Semiconductor Diodes for Use in Undersea Cable Systems," IEEE Trans. Reliab., *R-21*, No. 4 (November 1972), pp. 200–6.

64. L. E. Miller, 'Reliability of Semiconductor Devices for Submarine-Cable Systems," IEEE Proc., *62*, No. 2 (February 1974), pp. 230–44.

65. W. M. Fox et al., "SG Undersea Cable System: Semiconductor Devices and Passive Components," B.S.T.J., *57*, No. 7 (September 1978), pp. 2405–34.

66. E. I. Gordon, F. R. Nash, and R. L. Hartman, "Reliability Assurance for New Technology Semiconductor Devices," IEEE Electron Dev. Lett., *EDL-4*, No. 12 (December 1983), pp. 465–6.

67. R. T. Lynch, "Effect of Screening Tests on the Lifetime Statistics of Injection Lasers," IEEE J. Quantum Electron., *QE-16*, No. 11 (November 1980), pp. 1244–7.

68. R. H. Saul, F. S. Chen, and P. W. Shumate, "Reliability of InGaAs Photodiodes for SL Applications," AT&T Tech. J., this issue.

69. R. H. Saul and F. S. Chen, "Reliability Assurance for Devices with Sudden-Failure Characteristic," IEEE Electron Dev. Lett., *EDL-4*, No. 12 (December 1983), pp. 467–8.

70. S. Yamakoshi et al., "Reliability of High Radiance InGaAsP/InP LEDs Operating in the 1.2–1.3 μm Wavelength," IEEE J. Quantum Electron., *QE-17*, No. 2 (February 1981), pp. 167–73.

71. S. Yamakoshi et al., "Degradation of High Radiance InGaAsP/InP LEDs at 1.2–1.3 μm Wavelength," IEDM Tech. Dig. (December 1979), pp. 122–5.

72. M. Hirao et al., "Fabrication and Characterization of Narrow Stripe InGaAsP/InP Buried Heterostructure Lasers," J. Appl. Phys., *51*, No. 8 (August 1980), pp. 4539–40.

73. H. Kressel, M. Ettenberg, and I. Ladany, "Accelerated Step-Temperature Aging of $Al_xGa_{1-x}As$ Heterojunction Laser Diodes," Appl. Phys. Lett., *32*, No. 5 (March 1978), pp. 305–8.
74. Y. Furukawa et al., "Accelerated Life Test of AlGaAs-GaAs DH Lasers," Jap. J. Appl. Phys., *16*, No. 8 (August 1977), pp. 1495–6.
75. H. Imai et al., "Activation Energy of Degradation in GaAlAs Double Heterostructure Laser Diodes," J. Appl. Phys., *52*, No. 5 (May 1981), pp. 3167–71.
76. A. R. Eckler, "A Statistical Approach to Laser Certification," AT&T Tech. J., this issue.

APPENDIX

Glossary of Some Reliability Terminology

A.1 When failures occur

Wear-out* failure—This failure eventually terminates the life of all[†] devices, provided that they have not failed previously from non-wear-out causes. In connection with the failure of, for example, brake linings, wear-out can be understood as using up some property until little or nothing of it remains. In semiconductor devices, where the wear-out concept has no comparably simple meaning, all wear-out failures appear to be lognormally distributed. The dispersion or variance of the lognormal wear-out distribution occurs because all devices are initially inequivalent and hence they proceed to demise at different rates. The thermal activation energies for semiconductor wear-out failures lie in the range 0.5 to 1.5 eV.

Midlife* failure—This is an accidental failure of a device during its useful life, prior to wear-out. The accidental aspect refers to devices being at the mercy of inherently unpredictable external events. Even though all of the devices are initially equivalent in all respects and contain no built-in weaknesses, some devices will fail because of random accumulations of operating stresses, e.g., current surges, due to electrostatic discharges, which exceed normal specifications. Although these failures are usually infrequent, they would be expected to be exponentially distributed in time, by analogy with radioactive decay in which all nuclei are viewed as initially equivalent. The corresponding hazard rate is a constant.

Infant* failure—A failure from a variety of possible causes, all of which constitute an initial weakness in the as-made device, which terminates the operation of a device early in its intended use. Thermal activation energies are generally low (<0.4 eV). Since the weakness is in the device, screening techniques employing overstresses are useful

---

* Wear-out, midlife, and infant are defined in the context of the bathtub curve, Fig. 1a.

† Window glass, whose purpose is to let in some light and keep out rain, never wears out. For the purposes of a discussion of semiconductor reliability, it is assumed that all biased devices will eventually fail because of diffusion of nonradiative defects to the active region, if for no other reason.

in detecting the fraction of the population potentially susceptible to such early failures. Infant failures in an unscreened population may not be recognized if the sample size is too small.

Freak failure—An early failure group observed during accelerated thermal aging which is manifested as an S-shaped distortion in the early time portion of the lognormal wear-out failure distribution. Only if the accelerated aging is performed at another elevated temperature is it possible to tell if the freak distribution is associated with the presence of a high-activation (wear-out-type) or low-activation (infant-type) mechanism.

Premature failure—A failure which occurs prior to some desired time. It can comprehend the aggregate of the infant, midlife, and the early portion of the wear-out failures.

A.2 Why failures occur

Random or chance failure—The previously defined midlife failure is a random or chance failure. The failure is event dependent. It is an external randomly occurring event, rather than a device deficiency, that causes the failure. The failure (or start of degradation) cannot be anticipated because prior to the event, no conceivably measurable device parameter would have exhibited any precursor degradation. The failure (or start thereof) is sudden. No preinstallation screen, burn-in, accelerated aging, or purge will identify devices potentially susceptible to chance failure, since all devices are equally prey. This use of the word random is rather specific, and the use in other contexts can cause confusion. Thus, wear-out failures are never random failures despite the fact that the initial inequivalence in the devices, which is responsible for the dispersion in the wear-out failure distribution, was promoted by some randomness in growth or processing. The location of devices, in a box, subject only to wear-out failure, may also be said to be random.

Preordained failure—Wear-out failures and some infant failures are examples. A preordained failure is one in which some internal mechanism causes a device to fail deterministically at a rate preordained at the moment of birth. Preordained failures are, therefore, time dependent.

A.3 How failures occur

Gradual failure—A failure that is characterized by a graceful degradation, a gradual change with time in one or more critical parameters. The essence of the failure is the exceeding of some specification; the device is often still operative (e.g., wear-out of a pair of shoes).

Sudden or catastrophic failure—A failure that gives no warning of its imminence. It may actually occur rapidly (e.g., light bulb failure),
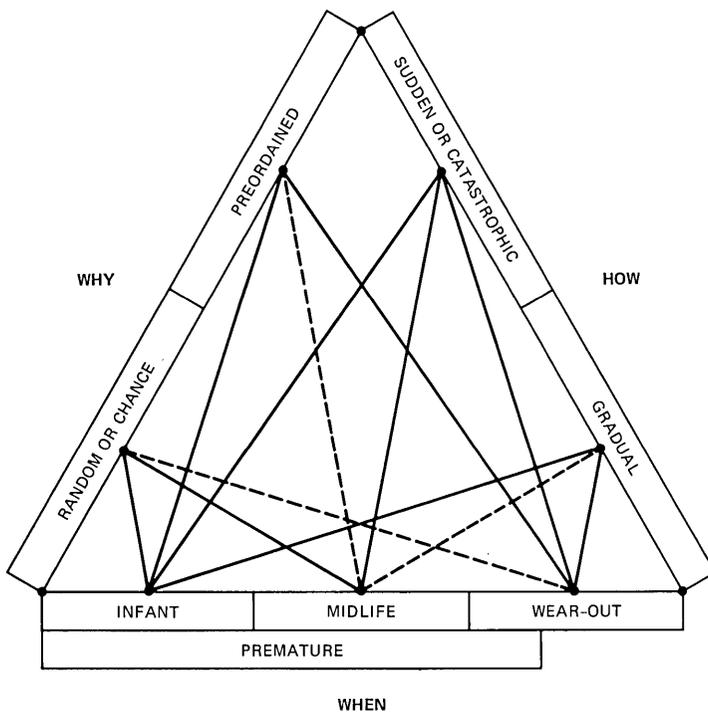
Fig. 11—Depiction of connections among terms defined in the Appendix. Solid lines denote the connections; e.g., wear-out failures can be sudden or gradual. Dashed lines denote an absence of connection; e.g., wear-out failures are not randomly provoked. The classification "premature" is a system-lifetime term; all other terms are referenced to device lifetimes.

or in the period in between times of observation. A preordained failure may appear sudden because it is impossible or inconvenient to monitor the parameters that would have enabled failure to be anticipated. Even had monitoring occurred, the changes might have been too small to detect if there was an incubation period prior to reaching a failure threshold, e.g., thermal runaway. The other aspect to the catastrophic failure is the complete inoperativeness of the device under any circumstances, subsequent to the failure, e.g., the opening of an electrical contact due to corrosion. Both random and preordained failures may be sudden.

A summary of the connections among these terms appears in Fig. 11.

AUTHORS

Richard W. Dixon, A.B. (Engineering and Applied Physics), 1958, Harvard College; M.S. and Ph.D. (Engineering and Applied Physics), Harvard Univer-

sity in 1960 and 1964, respectively; AT&T Bell Laboratories, 1965—. Mr. Dixon is currently the Director of the Lightwave Devices Laboratory of AT&T Bell Laboratories.

Eugene I. Gordon, B.S. (Physics), 1952, City College of New York; Ph.D. (Physics), 1957, The Massachusetts Institute of Technology; AT&T Bell Laboratories, 1957–83; Lytel, Inc., 1983—. Mr. Gordon has worked in the field of gas-discharge physics; microwave traveling-wave tubes; gas and GaAs injection lasers; acousto-optic modulation and deflection devices; image and display devices for the *Picturephone*® visual telephone service; and photo and electron beam lithography and packaging of silicon integrated circuits. Most recently, Mr. Gordon was Director of the Lightwave Devices Laboratory. He is the author of 43 published articles and holds 23 patents. Mr. Gordon received the IEEE 1975 Vladimir K. Zworykin Award and was elected to the National Academy of Engineering in 1979. He is past chairman of the Working Group on Imaging and Display Devices for the Office of the Director of Defense Research and Engineering and chairman of the Publications Committee for the Administrative Committee of the IEEE Electron Devices Society. Fellow, IEEE. Member, Advisory Group on Electron Devices, IEEE Publications Board, American Physical Society, Phi Beta Kappa, Sigma Xi.

Robert L. Hartman, B.A. (Mathematics), 1958, Columbia College; B.S. (Engineering Physics), M.S. and Ph.D. (Physics), Columbia University, in 1959, 1961, and 1968, respectively; AT&T Bell Laboratories, 1968—. At AT&T Bell Laboratories, Mr. Hartman is presently engaged in studying the physics of III-V compound semiconductor devices, particularly InGaAsP/InP and (Al,Ga)As-GaAs double-heterostructure injection lasers. He has contributed extensively to the understanding of the degradation mechanisms in these devices and authored or coauthored several key papers in this field, especially those involving the discovery and analysis of dark-line defects; the discovery of strain effects upon laser reliability; the attainment of long operating lifetimes in LEDs and injection lasers; the first measurement of the thermal acceleration of injection laser operating lifetime; and effects of surfaces on LED and laser reliability. In 1980 Mr. Hartman was appointed Supervisor of the Semiconductor Lightwave Lasers Group. Member, Sigma Xi, American Physical Society.

William B. Joyce, B.E.P. (Engineering Physics), 1955, Cornell University; Ph.D. (Theoretical Physics), 1966, Ohio State University; AT&T Bell Laboratories, 1966—. Mr. Joyce publishes primarily in the areas of architectural acoustics, classical mechanics and optics, history of science, semiconductor optical devices, research and development organization, and reliability. Member, American Physical Society.

Franklin R. Nash, B.S. (Physics), 1955, Polytechnic Institute of New York; Ph.D. (Physics), 1962, Columbia University; AT&T Bell Laboratories, 1963—. At AT&T Bell Laboratories, Mr. Nash has been involved in research on gas lasers (modulation and mode locking), detection of subnanosecond pulses using photodiodes, nonlinear optical materials, parametric oscillators, neoclassical theory of radiation, and semiconductor lasers (mode guidance, material defects, electro-optic anomalies, and degradation mechanisms). Currently, he is engaged in the reliability assessment of long-wavelength semiconductor lasers. Member, American Physical Society, IEEE.

Methodology of Accelerated Aging

By W. B. JOYCE,* K-Y. LIOU,* F. R. NASH,* P. R. BOSSARD,†
and R. L. HARTMAN*

(Manuscript received December 20, 1983)

Outlines are given for eight alternative black-box (i.e., input-output) methodologies that are appropriate for estimating, from external characteristics, the reliability of semiconductor lasers or other gradually degrading manufactured products with lifetimes too long to measure directly over practical time spans. These reliability estimates, which are essential for various components of such systems as submarine communication cables or satellites, are obtained from two classes of data. One class consists of the measured properties of statistically equivalent components, i.e., samples from the manufactured population, that have been operated to failure or at least to a significant degree of degradation. This degradation is often brought about in a shortened time span through the application of a temperature or other "accelerating stress" that is large compared to the operating temperature or other stress of the intended application. The other class of data is, for each component, comprised of the predeployment properties of that very component, including particularly its own predeployment degradation rate (which may also be measured under accelerating stresses). Brief consideration is given in passing to important special cases when only one of these two classes of data is available.

I. INTRODUCTION

It is usually desirable to have reasonably accurate estimates of, or at least bounds on, the reliability (expected operating life, failure rate, etc.) of a semiconductor device or package or other manufactured component before it is put into service. This is particularly true in the

---

* AT&T Bell Laboratories. † Voyager Technologies, Inc., Langhorne, Pennsylvania.

case of submarine communication cables or satellites where the replacement of a failed component is enormously expensive or even impossible. Unfortunately, the lifetime of a component often cannot be directly measured. That is, most degradation mechanisms are irreversible, and so the lifetime of a particular component cannot be measured before that very component is put into service, because measuring the life wears out the component. Furthermore, the possibility of measuring the lifetimes of statistically equivalent components under operating conditions and then inferring, in a statistical or population sense, the distribution of lifetimes of the components going into service is also sometimes ruled out simply because the median component lifetime (>100 years) and the planned system life greatly exceed any acceptable delay in the deployment of the system. Nevertheless, there is a need to obtain predictions of the life-spans of even such long-lived components. For example, the development of a submarine cable or satellite that requires the survival of 90 percent or even 99+ percent of the components for years or decades of system life is undertaken only if the median component life can be persuasively shown to greatly exceed the system life.

As a consequence of the need for estimates of component lifetimes, the semiconductor industry continues to develop procedures for estimating lifetimes by extrapolation from data obtained over relatively short periods of time. These procedures are clearly flawed by the logical and practical limitations that are typical of schemes for predicting the future, but they have, nevertheless, often turned out to be sufficiently accurate to the extent that many semiconductor components now have a credible claim to *predictable* reliability. In the present paper we formalize some elements of an accelerated-aging methodology appropriate for semiconductor lasers and other components (devices, packaged assemblies, manufactured products, etc.) that are characterized by gradual degradation and by possibly temperature- or stress-dependent operating characteristics. The methodology concentrates on estimating, rather than bounding, the reliability. Bounding, which is often sufficient for the clearly acceptable and the clearly unacceptable components, usually takes the form of a simple limiting approximation to an estimate.

While microscopic studies of the degradation physics play an important complementary role, which is often specific to each component type, the present work is devoted to the macroscopic black-box (input-output) methodology, which is more or less applicable to every component. A black-box component is described by choosing a set of *independent variables* consisting, in the case of a laser, of such macroscopic observables as temperature and optical-output power (total power in all modes). Under test conditions these variables are pre-

scribed functions of time (often constant or piecewise-constant functions), where time is also an independent variable. Other macroscopic observables, such as current, voltage, and optical power in the fundamental transverse mode, are then the *dependent variables* or *dependent observables*. The changing values of the dependent observables are the measures of degradation. *Failure* or *wear-out* is the point at which the observables have so degraded that the component no longer meets the requirements of the particular application of interest. (Thus only part of this work is applicable to components which fail suddenly* without prior influence on practically accessible observables.)

When a component goes into service, its life is, in general,[†] predicted using two classes of data: a statistical characterization of the population from which it is drawn (as inferred from measurements on samples) and observations of the predeployment rate of aging and other predeployment properties of that single component. The population data are usually "accelerated"; that is, a statistically equivalent group of components is worn out, or at least appreciably aged, in a short time period by the application of high temperature, high current, and/or other high stresses. The measured high-temperature degradation must then be extrapolated down to the system operating temperature. For the individual component's own predeployment data, and sometimes for the population data, a further extrapolation in time of the initial rate of aging is also required in addition to any temperature extrapolation. Thus, the central issue of accelerated aging and of this work is the methodology of these temperature (stress) and time extrapolations.

For carrying out the extrapolations with respect to time and temperature (or other stresses) we discuss eight (i.e., 2^3) alternative frameworks which are differentiated from each other by the following three dichotomies. First, a given framework uses either the (1) *activation* or the (2) *extrapolation* description. With the activation description, the degrading effects of various temperatures (or other accelerating stresses) are measured at a common reference temperature (or

---

* Often a failure is "sudden" only because relevant observables are not monitored. For example, a formerly sudden bond detachment may become predictable if the deterioration of the bond is inferred by adding the voltage to the monitored observables of a laser operated at constant current. Similarly, the onset of self-pulsation in a laser might become predictable if the ringing time or the peaks in the noise-power spectrum were monitored while the laser was still stable. To avoid the limitations of sudden-failure methodology, it is often desirable to make the effort to find a gradually degrading observable that is well correlated with failure.

[†] In particular cases only one class of data is available. If no degrading observable (precursor) is known (e.g., Ref. 1), or if cost considerations limit the measurements to samples, then only population data are available. If there are no spares to use as samples or if there is no time to degrade samples, then a component's life is estimated only from its own predeployment characteristics.

stress level). With the extrapolation description, the effects of each temperature or other stress are measured at that temperature or stress level. Second, each framework considered uses either the (1) *replacement* or the (2) *high-reliability* strategy. For replacement components (expected component life $\lesssim$ expected system life), such as automobile tires, where the component is likely to fail during the system (vehicle) life, the statistical distribution of lifetimes and, particularly, the mean component lifetime are of direct interest. On the other hand, for high-reliability components (as defined by the criterion that expected component life $\gg$ expected system life), such as submarine-cable lasers or the springs and contacts under the keys of a cheap hand calculator that is discarded at the first failure, it is only the early-component-life reliability that affects the system performance, and it is thus sometimes more efficient to terminate measurements at the expected system life (as possibly contracted by temperature or other degradation-accelerating stresses), and to use a description based on the degree of degradation observed over the system life. In short, with the replacement strategy one measures the distribution of lifetimes at a given degree of degradation (failure), whereas with the high-reliability strategy one measures the distribution in the degree of degradation at the end of a given time (e.g., the system life or the guarantee period).*

Third, each framework that we discuss uses either the (1) *sampled-population* or the (2) *truncation* approach. With the sampled-population approach, every deployed component from a given manufactured population has the same statistical failure description as inferred from measurements on sample components drawn from that population. With the truncation approach, each deployed component has its own statistical description based not only on samples from the population but also on its own predeployment degradation rate and its other predeployment characteristics. This permits discarding those components likely to fail early (e.g., the components that initially degrade rapidly) and results in a distribution of deployed-component lifetimes that is a truncation of the distribution given by the sampled-population approach.†

---

* As a variation on the high-reliability strategy, the data are extrapolated past the system life to failure. The lifetime distribution is then analyzed using the replacement methodology with the understanding that the results are accepted with confidence only over the system life.

† We do not discuss explicitly the case where the absence of population data requires that a component's reliability be predicted only from its own predeployment characteristics. When the observed predeployment characteristic is the degradation of an observable, this is similar to the extensively treated problem of extrapolating the motion of a comet. Such an extrapolation is practical for a comet because of the existence of a known law of motion. Thus, extrapolation accuracy is limited only by measurement error. In contrast, for a given type of semiconductor component, there is typically no parameterized law of degradation-versus-time that is precisely applicable to all, or even

This paper is organized as follows:

Following a largely qualitative summary of typical procedures and underlying assumptions (Section II), Sections III through VIII describe primarily the deterministic degradation of a single black-box component as it depends on time, temperature, and other stresses. Populations of degrading components are then considered statistically in Sections IX and X.

II. LIFETIME PREDICTION AND IMPROVEMENT

To improve the early lifetime, the median lifetime, or other reliability measures of shipped or deployed components, or to increase the confidence with which the relevant lifetime can be predicted, it is common to undertake certain reliability procedures during and/or after the manufacturing sequence. Specific procedures, or combinations of procedures, are known by such names as inspecting, performance evaluating, burning-in, accelerated burning-in, gettering, outgassing, tailored accelerated screening,[2] high-stress screening,[2,3] overstressing, purging,[4-6] stabilization,[4,6] etc., and the reader is referred to the literature for their explicit descriptions. In the present section we simply outline a framework for characterizing such reliability procedures.

Following the remarks in Ref. 7 we note that many reliability procedures, including most of those we discuss, are premised on the tacit assumption of deterministic degradation. Determinism means here that, under given external conditions, each component degrades in a manner that is a unique consequence of its initial state; i.e., the component wears out. The observed component-to-component differences in operating life and in performance are thus understood as

---

to any, components, even if the parameter values are chosen separately for each component. Thus, the inaccuracies of the "law," rather than measurement error, may be the main barrier to extrapolation. To evaluate this inaccuracy, population data on aged samples are required, as discussed in Section X, and thus we are back to the case of combined population and individual-component data.

consequences of the unintentional initial component-to-component differences that inevitably result from the limitations of any materials-preparation and manufacturing process. In contrast, the assumption of determinism is, for example, inapplicable to predictions about the subsequent imbalance between two initially equal radioactive radiation sources. Also, variations in lifetimes due to imprecisely anticipated variations in external conditions (e.g., a heat wave or a trawler attack on a submarine cable) are part of an overall system analysis, but are not considered in this work.

A given reliability procedure may either *extend* the life of the deployed components (i.e., yield a deployed population with a greater life than that of the original as-manufactured population), or *characterize* their life, or both. Thus aging samples to failure provides a characterization (e.g., estimates of the mean population lifetime, variance in lifetime, etc.), but it does not extend the lifetime of the population. On the other hand, a burn-in (predeployment operation) results in a survivor population that may have a greater lifetime than that of the original population. However, a burn-in does not, in itself, provide any verification that a lifetime extension has, in fact, occurred.

A given reliability procedure may be either *nonaccelerated* or *accelerated*. Thus, a burn-in may be carried out under the normal operating conditions of the deployed system, or it may be carried out to the same degree of degradation in a shortened time period through the application of stresses (temperature, current, humidity, etc.) exceeding those of normal operation.

A given reliability procedure is *nondestructive* or *destructive* according to whether or not there are survivors available for deployment. Aging samples to failure is destructive of the samples, whereas a burn-in or an inspection yields deployable survivors.

A nondestructive procedure is *benign* or *altering* depending on whether or not the component is changed by the procedure. There are *passive* benign procedures such as inspecting for a missing mirror coating or for a bad bond, and there are *active* benign procedures such as inspecting the voltage-current relation or other operating characteristics of a component. (The usual assumption is, of course, that the failure criterion associated with an inspection eliminates a subset of the components with a shorter expected life and leaves a population of survivors for deployment with a longer expected life.) In contrast to benign methods, procedures such as burning-in, annealing, outgassing, gettering, or stabilizing typically alter the components, for example, by causing partial wear-out.

Most lifetime-affecting procedures are intended to produce a deployed population for which the relevant lifetime is longer than that of the original population. This can be achieved in any of three ways

according to whether the lifetime of the deployed components is (1) shorter than, (2) approximately the same as, or (3) longer than the lifetime of those very same components (same subset of the original population) before the procedure was carried out. For example, a successful burn-in yields deployed survivors with a longer relevant life than that of the original population. Even so, the lifetime of each survivor is shortened by the burn-in period. In contrast, active or passive inspection has little or no effect on the lifetime of the components that survive the inspection. Finally, certain procedures, such as the addition of a getter (agent for impurity removal) or a high-temperature controlled-environment outgassing, may actually increase the life of most or all survivors, whether or not the survivor yield is 100 percent. (Once the efficacy of such a procedure has been established, the procedure may well become just one more step in manufacturing; i.e., the lifetime distribution of the as-manufactured population would then be correspondingly redefined upward.)

Some of these distinctions can be quantified as we now show by considering an example that is hypothetical except for the real data points of Fig. 1. The example is based on the replacement strategy: the data represent the distribution of lifetimes for a fixed degree of degradation (failure), but an analogous example could equally well be constructed using the high-reliability strategy, i.e., the data would represent the distribution of the degradation* for a fixed time (e.g., the system life or guarantee period). Figure 1 is the experiment of Ref. 7, as completed after the publication of Ref. 7. It shows the lifetime τ (for producing a given level of optical output power) for GaAs lasers of a particular design and manufacturing sequence.

The data of Fig. 1 were obtained with aging at 70°C. To say that (increasing) temperature is an *accelerating stress* implies, in the broad sense of acceleration, that aging at a higher temperature yields a distribution that falls below that shown in Fig. 1. In the narrower sense that we will use, valid acceleration of a population implies that the two distributions differ from each other only in a vertical displacement in $\ln\tau$. Figure 2a shows a valid accelerating stress, whereas Figs. 2b and c show stresses that are accelerating only in the broad sense. In all three cases the medians $<\tau>$ and $<\ln\tau>$ are changed by the stress, but in Figs. 2b and c the shape of the distribution, and hence, such other properties as σ, the standard deviation in $\ln\tau$, are also changed. Once the existence of a valid accelerating stress has been established by demonstrating a vertical shift between the measured

---

* For example, if L_f were the failure (minimum useable) value of the optical output power and L_i were the initial value at a common current, then one possible candidate for the ordinate in Fig. 1 under the replacement strategy would be the lifetime estimate $(L_f - L_i)t/\Delta L$ [or just $(L_f - L_i)/\Delta L$] rather than τ, where ΔL is the output change in a fixed time t.
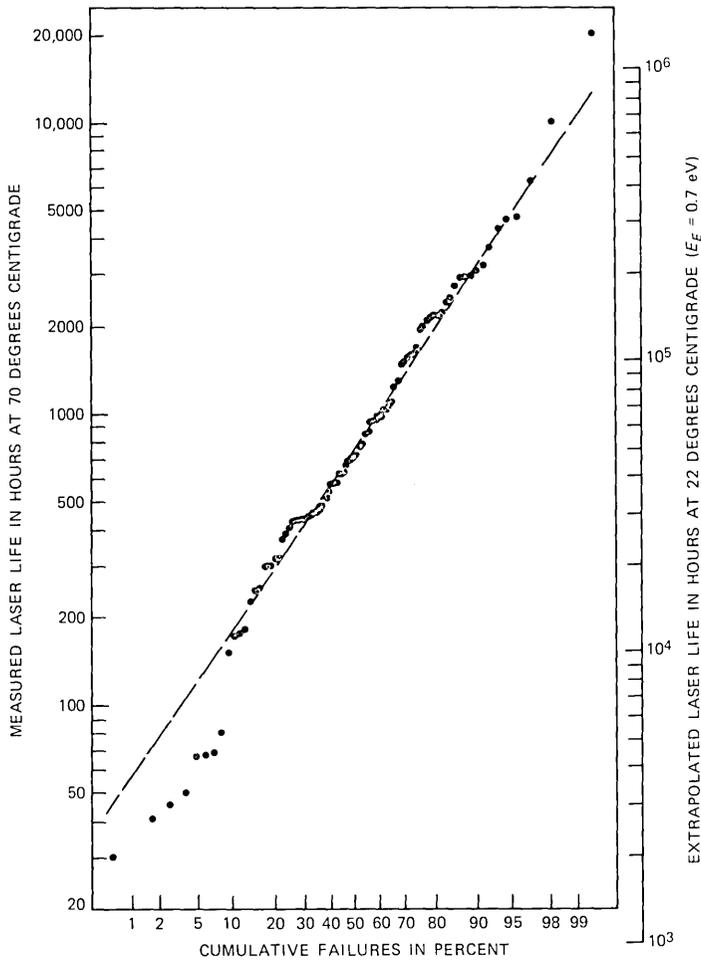
Fig. 1—Cumulative distribution on lognormal paper of AlGaAs laser lifetimes as measured at 70°C (left ordinate) and as extrapolated to 22°C with the Arrhenius relation (right ordinate).

distributions at different stress levels (e.g., two temperatures), then further data need only be taken at the higher level, and then the distribution at the lower level can be inferred by relabeling the ordinate in the manner of the right-hand scale in Fig. 1. An additional possible implication of a valid acceleration is that the lifetime has been shortened by the same factor for every component; i.e., the clock runs faster by a universal acceleration factor. This further implication should, however, be verified by step stressing; that is, by aging samples individually at two or more stress levels to see that they have, to within experimental error, a common acceleration factor, rather than
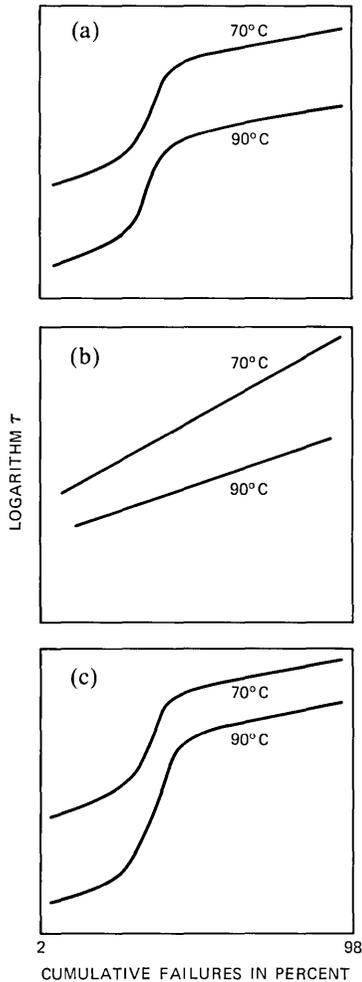
Fig. 2—Cumulative distribution of the lifetime τ with isothermal aging at 70°C and at 90°C. The ordinate is the logarithm of τ and the abscissa is an unspecified distribution. (a) Temperature is a valid accelerant; i.e., the clock runs faster by the same acceleration factor for all components, and the curves are vertically displaced. (b) Temperature is an invalid accelerant that accelerates different components differently, but conformance to the distribution (linearity of the curve) is preserved (see Fig. 5). (c) Temperature is a separately valid accelerant for each of the two subsets of the population.

different acceleration factors that are simply poorly correlated with the lifetimes.

Occasionally, accelerating stresses that yield the results of Figs. 2b or c can be utilized by invoking an acceleration factor that varies with time during aging or varies from component to component. For example, there may be two failure mechanisms (infant and normal), each with its own constant acceleration factor as in Fig. 2c, where the

curves become parallel on both sides of the bend. Often, the data are insufficient to justify high confidence in predictions based on a variable-acceleration-factor strategy. However, for example, it may be possible to screen out the components with the infant mechanism and leave a survivor population with a common acceleration factor. Alternately, as in Fig. 2b, one can sometimes scale both the median and the standard deviation. To give a physical example of how such a situation could arise we will later show, in conjunction with Fig. 5, that in some cases a single degradation mechanism can have an acceleration factor that varies from component to component.

After the first 10 percent or so of the lasers in Fig. 1 have failed, the dashed-line fit shows that remaining failures are well approximated by the lognormal distribution (the logarithm of the lifetime is normally distributed). Thus, in a replacement application where the mean and standard deviation of the lifetime are of paramount significance, it might be sufficient to use lognormal statistics to predict component reliability.

On the other hand, as is often observed in component lifetime distributions, the first few percent of the data fall significantly below the lognormal line, and lognormal statistics would be much too optimistic in a high-reliability application where only the first few percent of the failures were of importance. One high-reliability strategy would then be the use of a more realistic distribution, and indeed the data of Fig. 1 have been well fit by a double lognormal distribution.[8] This is a self-consistent approach, but not necessarily an appealing approach because of the high initial failure rate $-N^{-1}dN/dt$ (N = number of surviving lasers) arising from the characteristic lower-than-lognormal lifetimes of the early failures.

An alternative approach, assuming that all known practical methods for improving the lifetime of the lasers have already been incorporated into the manufacture, is to accept a yield reduction and to try to somehow pick out the longer-lived lasers. Suppose for definiteness that the best 40 percent ($y = 0.4$) of the lasers can be used for a particular application. What method should be used to select the 40 percent? The brute-force approach is a burn-in. Using the dash-line lognormal fit (median life = $\tau_m = 750$ hours; standard deviation in $\ln\tau = \sigma = 1.1$) shows that the 40-percent yield, 60-percent cumulative-failure point ($c = 0.6$) corresponds to $\tau = 991$ hours, i.e., a 40-percent yield would occur after about 991 hours of burn-in. To construct the post-burn-in distribution (the $B_{y=0.4}$ curve in Fig. 3), we note, for example, that the 80-percent point on the original distribution (point i in Fig. 3) becomes the 50-percent point of the survivor distribution ($i \rightarrow j$) and that its life is reduced by the 991-hours burn-in ($j \rightarrow k = 991$ hours). By this means the original lognormal curve (LN) is mapped into the burn-in curve ($B_{y=0.4}$). (Because the first 60 percent of the
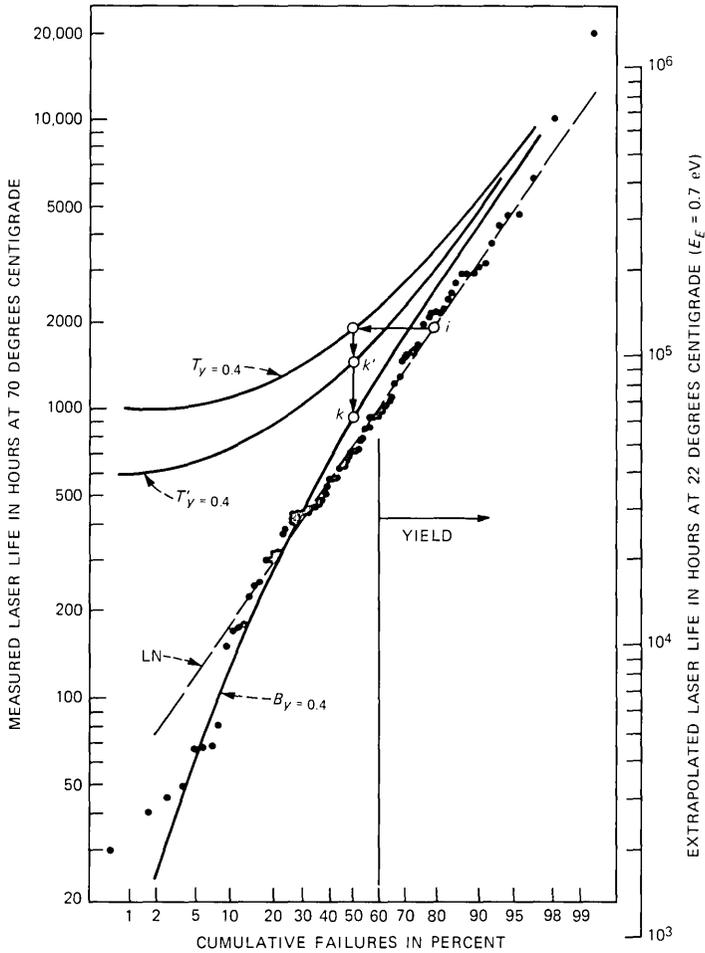
Fig. 3—LN is the lognormal distribution of Fig. 1. $B_{y=0.4}$ is the 40-percent yield burned-in survivors from LN (991-hour burn-in). $T_{y=0.4}$ is the 40-percent yield truncation of LN. $T'_{y=0.4}$ is the 400-hour burn-in (stabilization plus monitor times) 40-percent yield truncation of LN.

original lasers do not affect the $B_{y=0.4}$ curve, it does not matter that the first 10 percent or so of them fall below the LN curve.)

Several conclusions can be drawn from a comparison of the B and LN curves: After about 25-percent failure, the B curve is longer-lived than the original data, e.g., the median is raised from 750 to 900 hours. Thus the burn-in is of some value for a replacement application. Below $c = 25$ percent, the B curve is marginally distinct from the original data; i.e., the burn-in strategy costs time and money, wastes 60 percent of the population, and offers no early-life improvement for a high-reliability application.

The lognormal distribution is not invariant to a burn-in; that is, the straight LN line results in a curved B line. The lack of invariance of the lognormal distribution under many such yield-reducing procedures and the common presence of partial yield at each stage of manufacture imply that the lognormal distribution may often be only an approximation to the actual lifetime distribution. In particular, the fall-off of the B curve in Fig. 3 at low cumulative failure comes not from any physical mechanism that would cause additional early (infant) failures, but rather from a burn-in or from any manufacturing stress that subtracts a common number of hours from all devices in an otherwise lognormal distribution. Thus in particular cases other distributions, such as the double lognormal, the exponential, or the Weibull distribution, or burned-in or truncated modifications of these distributions, may prove more accurate.

If the burn-in strategy were pursued, the large number of hours required (991) means that an accelerated burn-in would likely be sought. Measurement of the associated acceleration factor is not normally a difficulty; that is, the accelerating stress (current, humidity, vibration, increased temperature, etc.) is simply applied until 60 percent of the lasers have failed. The problem is verification that a burn-in has, in fact, occurred (as distinguished, for example, from the precipitation of a new otherwise irrelevant failure mechanism which might yield a better or worse distribution than a burn-in). Evidence for or against the occurrence of an accelerated burn-in comes from aging to failure a sample of the survivors of the tentative accelerated burn-in and comparing their lifetime distribution to the B curve.

If it were decided not to use a burn-in to eliminate *all* of the 60 percent, it still might prove desirable to eliminate some of the 60 percent by using a shorter burn-in. The merits of such an approach could be assessed by similarly constructing the B_y curves for smaller values of y. In practice a burn-in more often proves useful for eliminating only a small fraction of the components.

Ideally it would somehow be possible to identify the first 60 percent to fail without, however, using up any of the life of the other 40 percent. The surviving distribution (curve $T_{y=0.4}$ in Fig. 3) is, in mathematical terms, a truncation of the original distribution, i.e., point i becomes point j, etc., as before, except that now there is no burn-in-time penalty ($j \rightarrow k = 0$). Clearly, the truncated curve is greatly superior to both the original and the burned-in curves, particularly for high-reliability applications where the early failures are important. (Incidentally, the curvature in the truncated distribution shows that the lognormal distribution is not invariant to a truncation procedure.)

Inspection of the static or dynamic (operating) properties of the components is an appealing method for eliminating early failing de-

vices because inspections do not use up any of the life of the survivors. Also, no life is consumed for those components that fail the inspection, and thus they are available for other, less demanding applications. On the other hand, in contrast to a burn-in that automatically eliminates the early failing components, an inspection criterion may be imperfectly correlated with device life. The quality of a given inspection criterion is judged by comparing the lifetime distribution of the inspection survivors with the T_y curve constructed for the yield of that inspection. To the extent that the survivor curve falls below the T_y curve, the inspection criterion is rejecting at least some longer-lived components and passing some shorter-lived ones. As with burn-ins, known inspections often prove useful only for eliminating a modest fraction of the population.

Another truncation strategy, which is incorporated, for example, in the purge[4-6] and the tailored-accelerating-screen[2] procedures, is premised on the following physical model: All components eventually fail because all possess a common wear-out mechanism. In addition, certain flawed components are also afflicted with a second failure mechanism having a shorter life under normal operating conditions. Evidence suggesting this two-mechanism model is the so-called S-shape of the data in Fig. 1 that corresponds to a high failure rate in the early failing branch of the S. (However, as noted above in conjunction with the B curve, a fall-off of the early failing end of the distribution is no proof of a second mechanism because a fall-off can occur for other reasons.) In the case of the data of Fig. 1 there is physical evidence for two mechanisms; that is, postmortem inspections of failed lasers of this type show that early failing lasers typically exhibit dark-line defects[9] in their active layers whereas longer-lived lasers usually fail without developing inhomogeneities in their spontaneous emission patterns viewed through windows in their contacts.[10] Figure 2c shows another form of evidence for two mechanisms. The two distributions are vertically displaced (same slope) on either branch of the S curve, but the acceleration factor (vertical displacement) is different for the right branch (longer-lived mechanism) than for the left branch (second mechanism).

Where there is a shorter-life mechanism present in some components it may be possible to find a *selective* stress which accelerates this mechanism to component failure without, however, appreciably accelerating the common mechanism.[2,4-6] For example, a briefly applied reverse-bias current may be virtually harmless to a uniform laser, but the current may be so concentrated into the defect of a laser with a dark spot or line that the current-voltage characteristic of the laser with the defect is altered or destroyed almost instantly. The justification for a claim that selectivity has permitted an improvement over a

cumulative distribution obtainable from a burn-in is, again, a comparison with the T_y and B_y curves for the relevant yield, i.e., the data should at least fall above the B_y curve. (If the data were to fall above the T_y curve as well, then the selective stress is also affecting the unafflicted survivors in some way that increases their life.) In contrast to a burn-in, the selective-stress method is relatively inflexible with respect to yield; that is, if only 10 percent of the lasers are flawed with a second mechanism, then a selective stress cannot go very far toward identifying the 60 percent with the shortest lives. In common with a burn-in and in contrast to an inspection, a selective stress often destroys all but the surviving components. However, it is certainly possible in principle to develop a selective test (inspection) which identifies the second mechanism nondestructively.

Another important approximation to the truncation ideal is the *projection method* or rate monitoring in which each component is aged for a short while during which its degradation is monitored. Then the observed degradation is projected into an estimated life for that component. (Alternatively, with the high-reliability strategy it is the degree of degradation at the end of the system life that is obtained by projection.) The deployed survivors are, in the case of our example, the 40 percent with the longest projected lives. As with other procedures, the monitor-period degradation may be conducted on an accelerated basis.

A particularly convincing demonstration of the effectiveness of the projection method was given by English et al., who predicted the life of titanium-gold thin-film conductors from the predeployment rate of increase of each thin film's resistance.[11] The method has also proved useful for semiconductor lasers since the early days of reliability studies on such devices. In the case of Ref. 7, the projection of a laser's life, defined as the ability to emit some minimum optical-output power, was based on the early rate at which the drive current had to be increased in order to maintain that power. Ninety lasers were then aged to death and a comparison of their projected and actual lifetimes justified the conclusion that "a typical laser degrades fairly steadily toward a time of death which can be reasonably accurately estimated from the laser's early degradation rate."[7]

The confidence in the projection method can be assessed, and the accuracy of the method can be improved, if some samples from the population are aged, under normal or accelerating stresses, through the projection period (i.e., aged to failure with the replacement strategy or aged through the system life with the high-reliability strategy). This is discussed in Section X.

The accuracy of the projection method can also be improved if the population is first made more homogeneous, for example, by selectively

eliminating components with a second, shorter-lived failure mechanism or by inspecting for manufacturing defects.

Finally, the projection method is sometimes improved if the components are *stabilized*[4,6] in advance of the monitoring period on which the projection is based. The main idea of stabilization is to drive to saturation or to anneal any initial transients in the degradation rate. For example, a degrading surface current that is in parallel with the primary drive current of a device will distort the measurement of the degradation of the primary current. This surface current might reach a saturation value in a relatively short period of time under normal operation, or it might be possible to selectively accelerate the surface current to saturation through the temporary introduction of a selective stress (e.g., a particular atmosphere) that has little effect on the mechanism of the primary-current degradation. A difference between stabilization and the selective acceleration of a second mechanism of a flawed device is that in only the former case is the mechanism of a type that can be driven to saturation without causing device failure.*

The projection method may be regarded as only a variation on the inspection method because both methods are based on a correlation between some predeployment measurement and the subsequent life of the same component. The projection method differs from the inspection method because it uses up some of the component's life. For example, if stabilization plus the monitor period were equivalent to four hundred hours of operation, then, for a 40-percent yield, the expected distribution in the case of perfect lifetime prediction is the $T'_{y=0.4}$ curve in Fig. 3 ($j \rightarrow k' = 400$ hours). Unless the stabilization process happened to actually improve the individual component's life, lifetime experiments on samples from the 40 percent with the longest predicted lives would be expected to yield a distribution which fell somewhere between the $T'_{y=0.4}$ curve (perfect correlation between the projected and actual lifetime) and the burn-in curve $B_{y=0.716}$ (no correlation between the initial degradation rate and the actual lifetimes). (A yield $y = 71.6$-percent corresponds to a 400-hour burn-in; this curve is not shown in Fig. 3.) Failure to exceed the B_y curve suggests that the monitored degrading observable is uncorrelated with the actual failure mechanism and that any improvement over the original distribution arises simply from the burn-in aspect of the stabilization and monitor period.

---

* The term stabilization is also sometimes used to mean simply waiting until the degradation rate has become sufficiently low. This arises, for example, when the degradation rate is described by a bounding function (such a linear time dependence extending the last observed tangent to a sublinear curve) rather than by an actual fit; i.e., in the context of bounding the life rather than of estimating it.

Certain procedures that we have outlined involve extrapolations with respect to time and temperature or other stresses. In the remainder of this paper we discuss these extrapolations in more detail. Sometimes we will assume, for example, that the acceleration factor for a given component does not change over its life, or that all components in the population have the same acceleration factor, etc. In each such case it is understood that appropriate screening procedures, such as inspection, stabilization, or selective acceleration, have already been identified and carried out as necessary in order to ensure the applicability of the assumption in question to the survivors.

III. ACTIVATION FACTOR

Following the discussion in footnote 7 of Ref. 7, we develop in this section the first of two procedures for quantifying the acceleration that results from operation at enhanced stress levels. The treatment is based on four temperatures, some of which may be identical in particular cases:

T_a = the higher aging temperature
$T_{a'}$ = the lower aging temperature
T_r = the reference temperature where measurements are made
T_s = the system temperature for deployed components.

In conventional terminology a fundamental assumption underlying typical predictions about a given type of device is that an *accelerating stress* (temperature, humidity, radiation, current, electric field, mechanical vibration, etc.) or combination of stresses can be found for that component which, when applied at levels in excess of the intended application, brings about, in a shortened length of time, the *same* process and degree of degradation that occurs in the intended application. In a macroscopic theory the degradation that a device or package undergoes is measured by the changing value of at least one macroscopic dependent observable I. For example, I could be chosen as the threshold current i of a semiconductor laser (or the current at some other value of the optical power), and I is then observed to degrade (increase) more rapidly as the temperature, or current, or other stress is increased. As another example, I is the coupling efficiency of a transmitter (fraction of the optical power from the laser that enters the fiber mode).

A macroscopic observable is, however, a potentially ambiguous measure of degradation if, for a given amount of degradation, its value also depends on the temperature (or on whatever stress is used for acceleration). For example, the threshold current of a semiconductor laser is typically strongly temperature dependent. Similarly, in addition to the irreversible degradation in time of the coupling efficiency

due, say, to solder creep, there may also be a separate reversible dependence of the coupling efficiency on temperature due to a mismatch of the thermal-expansion coefficients of the components of the structure supporting the laser and the fiber. The simplest way to remove this additional dependence, and to achieve thereby a unique relation between the amount of degradation and the value of the observable, is to make all measurements at one reference temperature even though the aging is carried out at two or more higher temperatures. Thus, we make the notion of acceleration precise as follows. As shown in Fig. 4, for the case where the stress is temperature T, a device is raised to the higher aging temperature T_a (path 1–2), aged for time Δt_a (path 2–3), and then brought back* to the lower aging temperature $T_{a'}$ (path 3–4) where it is found (horizontal-line comparison 4–5 in Fig. 4b) to have experienced the same degradation (same change in I) that would have occurred (solid-line path 1–5 in Fig. 4b) in time $\Delta t_{a'}$ at $T_{a'}$. The acceleration factor or *activation factor* m between T_a and $T_{a'}$ is then defined by[7]

$$m = \Delta t_{a'}/\Delta t_a, \tag{1}$$

i.e., aging at T_a is equivalent, in this particular sense, to making the clock run m times faster than it runs at $T_{a'}$.

Equation (1) should be contrasted with other commonly advanced definitions of acceleration. In eq. (2) of Peck and Zierdt,[3] for example, acceleration is defined in terms of a scaling of the dependent observable I rather than of time. Thus their activation factor, which they represent as the exponential of an activation energy, would change if I were redefined as the logarithm of, or as some other function of, the threshold current rather than as the threshold current i. Since i and, say, $\ln i$ or $i^2 \ln T$ are equally good observables, it is not easy to say which, if any, of their corresponding activation energies is more closely associated with any energy associated with a microscopic degradation process, nor is it otherwise easy to establish any unique value for this Peck-Zierdt activation factor or activation energy. With the null-technique definition of eq. (1) the activation factor does not change with these redefinitions of I. Furthermore, in eq. (1) the effect of aging at both T_a and $T_{a'}$ is expressed in terms of measurements made only at $T_{a'}$; thus, we also contrast eq. (1) with those definitions of m (or of the activation energy) in which the effect of aging at each temperature is expressed in terms of measurements made at that temperature. In

---

* In practice the component may, or may not, actually be brought back to the lower stress level. For example, when eq. (22) is known to apply, the consequences of temperature lowering are inferred theoretically as in eqs. (28) through (31). Alternately, one may find experimentally a condition on I or T_a that corresponds to the condition on I which defines Δt_r at T_r.
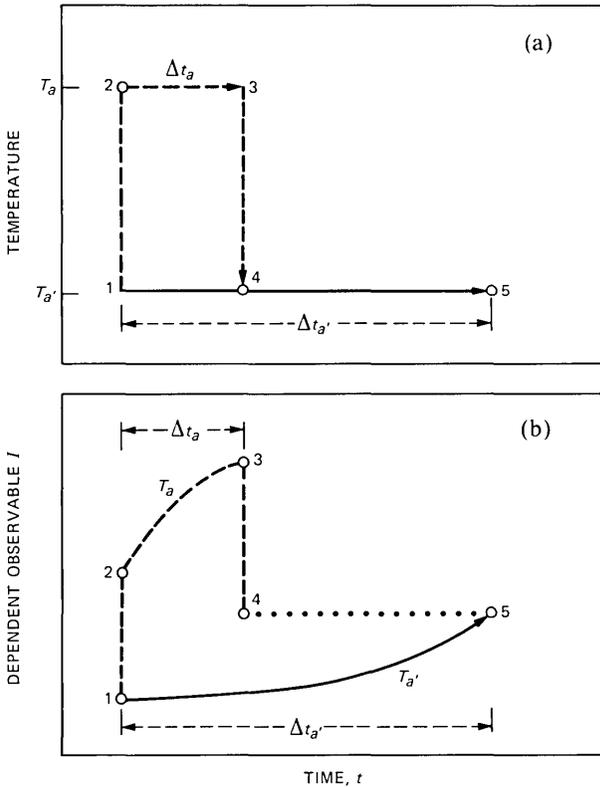
Fig. 4—The change (degradation) of an observable I (e.g., threshold current of a laser) as a function of time (t) and temperature (T). (a) The paths in time-temperature space. (b) The time dependences of the observable I.

addition to yielding an m that varies with the definition of I, these two-temperature-data definitions yield an m that almost always varies from device to device because of device-to-device differences in the temperature dependence of I or in the end-of-life criterion. (See Section V for an example based on a power-supply limitation.) By avoiding the additional arbitrariness associated with measurements at two temperatures, the definition of eq. (1) seeks to narrow the device-to-device differences in m and, hence, to uncover a single population value for m if such a single value exists. A final advantage of eq. (1) arises from the fact that m exists even when I is undefined at T_a, as, for example, when I is the threshold current (or the current at some higher value of the optical-output power) and T_a exceeds the maximum temperature for lasing.

It is useful to extend eq. (1) in two ways. Firstly, for graphical simplicity, Fig. 4 compares aging at T_a and $T_{a'}$ on the basis of measurements made at a reference temperature T_r that is equal to $T_{a'}$.

Clearly the equality of aging at T_a and $T_{a'}$ (path 4–5 in Fig. 4b) can be shown from measurements made at any reference temperature T_r. Thus, if T_a and $T_{a'}$ are the aging temperatures, then eq. (1) and the value of m are independent of the reference (measurement) temperature T_r. Secondly, it is often convenient to think in terms of rates rather than times. The reciprocal of the elapsed time at a given aging temperature is proportional to the degradation rate (clock rate) R at that temperature. Thus, eq. (1) becomes

$$m = \Delta t_{a'}/\Delta t_a = R_a/R_{a'}. \tag{2}$$

The last equality of eq. (2) expresses m (a two-temperature function) as the ratio of two one-temperature degradation rates. Because the R's are defined in terms of a ratio, they are indeterminate with respect to a common multiplicate constant; that is, we have defined relative rates at two aging temperatures, not an absolute rate at each temperature. For an example see eqs. (28) through (30).

When the degradation is reversible (e.g., can be annealed out) the measurement of m is conceptually straightforward. Otherwise, m can be measured on a population basis by comparing the degradation times or rates for two statistically equivalent populations aged at two temperatures (*isothermal aging*)[2,3,6,12] or, on an individual-device basis, by partially aging a single device alternately at two stress levels (*step stressing*)[3,6,12,13,14] and then scaling the time intervals at T_a by that m that yields a smooth $I(t)$ curve, all measurements of I being made at T_r. (Any transients associated with the temperature change are often small enough to be confidently removed in the analysis with short backward extrapolations of the post-transient data.)

With either approach for measuring m it is necessarily assumed that the lower aging temperature $T_{a'}$ is chosen to be high enough so that a measurably large amount of degradation occurs at $T_{a'}$ in the time available. This means that in some applications it is possible to choose $T_{a'}$ as the system operating temperature T_s, while in other cases one is forced to choose $T_{a'}$ above T_s, and determination of m between T_a and T_s also requires an extrapolation (see Section IV).

Most of the parameters in this work can be similarly measured on a population basis or on an individual-device basis, and an extrapolation may, or may not, be required. Because we do not treat the optimization of experimental procedures, we will not discuss the measurement of each parameter in detail.

In practice the methodology often reduces to fitting the data as well as possible with a linear degradation model (I changes linearly in time at constant T), i.e.,

$$m = \frac{R_a}{R_{a'}} = \frac{\partial I_r/\partial t_a}{\partial I_r/\partial t_{a'}}. \tag{3}$$

In eq. (3), $\partial I_r / \partial t_a$ is the isothermal degradation rate when dt_a is the elapsed time at T_a and ∂I_r is the resulting change in I as measured at T_r, that is, for small Δt_a and for $T_r = T_{a'}$,

$$\frac{\partial I_r}{\partial t_a} \approx \frac{I_4 - I_1}{\Delta t_a} \neq \frac{I_3 - I_2}{\Delta t_a}, \tag{4}$$

with I_4 and I_1 as in Fig. 4b. Thus, eq. (3) becomes

$$m = \frac{(I_4 - I_1)/\Delta t_a}{(I_5 - I_1)/\Delta t_{a'}} = \frac{\Delta t_{a'}}{\Delta t_a}, \tag{5}$$

which agrees with eq. (2) and thereby justifies eq. (3). The invariance of m, when I is changed from i to some $g(i, T)$ such as $\ln i$ or $i^2 \ln T$ can also be seen as follows. If I is replaced by $g(I, T)$ in eq. (3), i.e., if ∂I is replaced by $(\partial g / \partial I) \partial I$, then the $\partial g / \partial I$ of the numerator cancels that of the denominator because both are evaluated at the same I and at the same temperature (Fig. 4b).

Within a population of devices that are presumably degrading by a common mechanism, there invariably appear the expected device-to-device differences in the degradation rate that result from initially inequivalent devices. In addition there also sometimes appear less welcome device-to-device differences in the activation factor (as measured on an individual-device basis by step stressing). There thus arises the question of whether these differences in the activation factor should be assigned to measurement error (in which case a common population description of the activation factor should be used for each device) or whether real measurable device-to-device differences in m are to be expected (in which case reliability prediction for each device could sometimes be improved by measuring that device's own activation factor before putting it into service). Most of the literature tacitly assumes that any such differences are only experimental error. Our position on this question is that either case is possible, the choice depending upon the specifics of the mechanism. To justify this position we cite a mechanism of each type. On the one hand, for example, the temperature dependence of the degradation could be dominated by an energy level associated with the migration of an impurity or defect, and this level might well be negligibly affected by growth and manufacturing variations; i.e., to well within experimental error the level is essentially a material constant. On the other hand, consider, as in Fig. 5, a double-heterostructure device where the migration of an impurity or defect D, which crosses a p-type ternary-composition confining layer of $Al_x Ga_{1-x} As$, is driven by the energy of hole-electron recombination (recombination-aided diffusion).[15] The number of electrons which escape from the active layer and become available for recombination as minority carriers in the confining layer has a Boltzmann-factor dependence upon the conduction-band energy step E_1 and hence
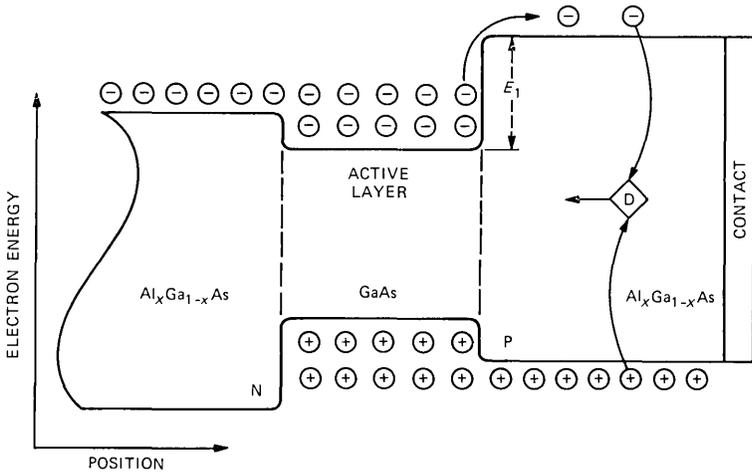
Fig. 5—Recombination-aided diffusion of a defect or impurity D, as controlled by the number of electrons which surmount the conduction-band barrier E_1. The magnitude of E_1 is determined by the composition x of the p-type $Al_xGa_{1-x}As$ confining layer.

upon the composition x of the confining layer.[16] Because x can vary inadvertently from wafer to wafer, and even across a given wafer, apparently meaningful wafer-to-wafer[12,17] and device-to-device differences in the activation factor do not necessarily rule out the possibility of a single degradation mechanism.

For a concrete and simple hypothetical example of Fig. 5, assume that under lasing conditions the active-layer carrier concentration is the same for every laser and for every value of the temperature T and optical power P. Then the carrier concentration just inside the confining layer is $n = \exp(-E_1/kT)$. Assume that there is some observable I for which the recombination-aided degradation rate is $R \sim n \sim \exp(-E_1/kT)$ and the lifetime is $\tau \sim 1/R \sim \exp(E_1/kT)$. Then if the laser-to-laser differences in E_1 are the main causes of the lifetime differences and if E_1 is normally distributed among the lasers, it follows that τ is lognormally distributed except at the shortest of lifetimes (small E_1) where Boltzmann statistics and small-signal modeling break down. Further, Arrhenius' rule is obeyed separately for each laser with E_1 the activation energy. Finally, if δE_1 is the standard deviation in the step heights, then σ in $\ln\tau$ is given by a $\sigma = \delta E_1/kT$ that is temperature dependent as shown in Fig. 2b. This sort of temperature-dependent σ is likely to occur often. Thus, we justify the common usage of a temperature-independent σ as follows: δE_1 is sufficiently small that the dependence is weak over the relevant temperature range (which is small compared to the Kelvin temperature). Alternatively, certain empirical procedures automatically deduce "the" activation energy or factor over the relevant part of the distribution, e.g., the replacement strategy normally infers "the" activation energy at the

median of the distribution, and the high-replacement strategy automatically finds "the" activation energy at the early part of the degradation history.

For a more elaborate example let the degradation rate R be proportional to the product ns with n as before and s the concentration of the impurities D in an excited state with an excitation energy E_2; i.e., $s \sim \exp(-E_2/kT)$. Then $\ln\tau \sim E/kT$ where $E = E_1 + E_2$, and the variance σ in $\ln\tau$ is now given by $\sigma = \delta E/kT = (\delta E_1 + \delta E_2)/kT$. If the variance δE_2 arises from weak impurity-impurity interactions of a low-concentration impurity, then δE might be negligibly small and thus further justify the usual temperature-independent-σ approximation which underlies the valid-acceleration methodology.

It may, of course, turn out that aging at T_a yields a different form of degradation than aging at the system temperature T_s. For example, the device or its bond may melt at T_a, or a strongly temperature-dependent degradation mechanism,[5,18] which is of secondary significance at T_s, may become dominant at T_a. Then temperature, or at least the value T_a, is simply not a valid accelerating stress because it does not induce the *same*, or even approximately the same, physical degradation process as aging at T_s. For such reasons, the maximum available activation factor may be less than desirable, and the available time may not be sufficient for aging all the way to failure even at the highest valid stress level. Similarly, excessive current[19] or optical power[20] is known to introduce new degradation mechanisms in lasers. Incidentally, the suggestion of a different form of aging (invalid acceleration) occurs when two dependent observables are measured on the same device, and they yield unequal activation factors.

It may also turn out in some cases that the value of m depends upon the time interval Δt_a and the starting time. Then m is appropriately measured over the time interval of interest, which is typically the system life or the device life. Alternately, as in eq. (3), one considers the instantaneous activation factor $m = dt_{a'}/dt_a$, as approximated by measurements over short intervals.

It is common and preferable for the shape of the aging curve $I(t)$ at T_a (line 2–3 in Fig. 4b) to be similar to the shape at $T_{a'}$ (line 1–5 in Fig. 4b). However, Fig. 4b emphasizes the fact that the two shapes are not necessarily similar. For example, a physically remote, and possibly even aging, shunt path might conduct part of the current at T_a, but none at $T_{a'}$, T_r, and T_s. Temperature might still be a valid accelerating stress for the degradation process of interest, although not necessarily an easy one to extrapolate, if the condition of the shunt path does not affect the observables at T_r and T_s. For another example of dissimilar shapes at different temperatures see the discussion in the paragraph following that with eq. (14).

In order to obtain a measurably large amount of degradation in the available time, it is often necessary that the lowest aging temperature $T_{a'}$ be chosen above the system temperature T_s. Consequently, the activation factor m is used primarily as basis for extrapolation to T_s (see Section IV). However, we emphasize that the measured m is a directly useful quantity when $T_{a'}$ and T_r can be taken equal to T_s. For example, consider the commonly encountered case where device-to-device and wafer-to-wafer differences in m, if any, are negligibly small in their effects compared to the corresponding differences in the device lifetimes. Then it would be appropriate to use a common value for the m of every device, with m estimated from a protracted measurement of $\Delta t_{a'}$ or $R_{a'}$ on samples, plus a brief corresponding measurement of Δt_a or R_a on a statistically equivalent sample. Thereafter, the population degradation rate or lifetime with operation at $T_r = T_{a'}$ might well be usefully estimated for devices from each successively grown wafer or successively processed batch by using eq. (2) and only a (brief) measurement of Δt_a or R_a for sample devices from that wafer or batch.

Needless to say, m depends upon the operating conditions at T_a. For example, if a laser is operated at T_a with the same current as at $T_{a'}$, the value of m will, in general, differ from the value when the operation at T_a is at the same optical power as at $T_{a'}$. The point is that in both cases m may well turn out to be meaningfully, and even usefully, defined. For example, temperature T, current i, and optical power P are all potential accelerants for laser degradation, and it is not usually possible to change T while holding both i and P constant (unless, for example, one goes to such extremes as changing the mirror reflectivity in the case of a bulk-degradation mechanism). For instance, if P is held constant, then $m(T_{a'}, T_a)$ tacitly includes some acceleration due to a changing i. The choice among holding i, or P, or some function of i, P, and T constant while T changes is usually made on some practical basis, as exemplified in the following three paragraphs.

Aging at constant P is commonly used when the cause of degradation is unknown. Because i increases with T at constant P, a comparatively larger activation factor results, and neither i nor P loses influence by becoming negligibly small. In contrast, aging at constant i causes P to decrease with increasing T. This yields a comparatively smaller activation factor for given temperatures. Further, if P were the dominant cause, the activation factor could be small, or negative, or virtually impossible to extrapolate.

On the other hand, if i rather than P is known to be the dominant cause of degradation, then constant-i aging has the advantage of permitting very large activation factors through the use of temperatures well above the maximum temperature for lasing.

Choices other than constant i and constant P have advantages. For

example, both constant-P and constant-i data are often taken, and it is then difficult to compare the results because of the differing activation factors at common temperature differences. Also it may be inefficient to generate two sets of data (constant i and constant P) at temperatures $T_{a'}$ below the maximum temperature for lasing. If the current at fixed optical power follows an empirical rule $i(T)$—e.g., $i = i_0 e^{T/T_0}$, where $i_0 = i(T = T_0)$—over much of the lasing temperature range of interest, then aging at $i(T)$ at each T ages at constant P over much of the lasing range, and then attempts the extrapolation beyond the lasing range using a method that is as rational as the constant-i method.

Constant-T aging is also a possibility in principle since increased i and P both accelerated degradation to some extent.

One might try to apportion the cause of degradation between i and P. For example, below threshold, the degradation rate is modeled as a function of i, and this dependence is then extrapolated above threshold as a function of that part of the current that does not appear as optical power. Any observed above-threshold degradation that exceeds the extrapolated value is then attributed to, and modeled as a function of, the optical power P. This is experimentally difficult to carry out, and there is no assurance of success. For example, if the degradation were actually controlled by the active-layer carrier concentration, or by the current in one shunt path driven by the active-layer carrier concentration, then the apportionment between i and P would be misleading at best.

As a practical matter, confidence in a particular acceleration scheme over a particular temperature (stress) range comes largely from an observed smooth steadily increasing dependence of m on T_a over that range.

In summary, we have considered $I = f(T_r, T_a, t_a)$, which is the dependence on time t_a of the degradation of an observable I for a device or package which is isothermally aged at T_a, but which is periodically and briefly cooled for a series of measurements at a reference temperature T_r. Here f is an interpolated function that passes through the I values measured at T_r. If T (temperature or other stress) is a valid *accelerating stress*, then I is of the particular form

$$I = f(T_r, mt_a),\tag{6}$$

with an *activation factor* m given by

$$m = R(T_a)/R(T_r) = t_r/t_a.\tag{7}$$

In eqs. (6) and (7), t_a is the elapsed time (at T_a) and $R(T)$ is the so-called clock rate or degradation rate at T. A realistic degradation model of the type of eq. (6), where there is both a benign transient

and a wear-out mechanism, appears in Ref. 21. Equation (14) is a simple hypothetical example of eq. (6).

IV. ACTIVATION ENERGY

A reliability prediction for a device drawn from a population is often based, at least partially, on measurements of the degradation of sample devices from that population. For each sample device one would typically like to measure the isothermal degradation at the system temperature of at least one observable I. That is, one would like to observe the function

$$I = f_s(t_s) \tag{8}$$

at certain times or at all times t_s over the interval $0 \leq t_s \leq t'$, where t_s is the elapsed time at the system temperature T_s, and t' is at least as long as the lesser of the system life and the device life. In the usual case, however, the available measurement time is much less than t'. Then f must be inferred by an extrapolation that consists, in general, of two parts.

First the shape of f at T_s is found by the method of Fig. 6. That is, if temperature is a valid accelerating stress, the device is aged iso- thermally along the solid line labeled T_a in Fig. 6. Periodically and momentarily the device is cooled (dotted line) to $T_s = T_r$ where I is
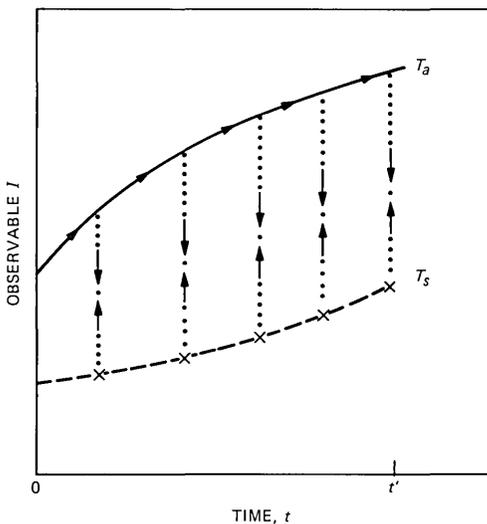
Fig. 6—The change (degradation) of an observable I which is aged isothermally (solid line) at T_a. Periodically and momentarily the component is dropped to temperature T_s and returned to T_a (down and up each dotted line) to yield an interpolated curve (dashed line) between the data points (×'s) at T_s.

measured (×'s in Fig. 6). Interpolation between the data points yields the dashed line of Fig. 6, which can be written in the form

$$I = F_s(t_a). \tag{9}$$

F_s in eq. (9) is an empirical function which gives the shape of $I(t)$ for measurements at $T_r = T_s$ and aging at T_a. Often F_s is then fitted with an analytical function.

Secondly, the inferred rate at which the shape would have been traced out, had the aging temperature been T_s, is found from eq. (7) by extrapolating the activation factor. Often it is shown empirically (at a series of temperatures T_a, $T_{a'}$, $T_{a''}$, $\cdots$ that are all high enough to induce appreciable degradation in the available time) or else it is just assumed that, to within the measurement error, $R(T)$ is of the Arrhenius form

$$R(T) = R_\infty e^{-E_A/kT}, \tag{10}$$

where E_A (the *activation energy*) and $R_\infty = R(T = \infty)$ are parameters. As indicated in the example associated with Fig. 5, the physical origin of the Arrhenius form is normally the Boltzmann distribution (which dominates the weaker algebraic dependence that usually also enters the degradation rate as prefactors to the Boltzmann exponential). Extrapolating, i.e., assuming that eq. (10) applies down to T_s, causes eq. (7) to become

$$m = e^{-E_A(1/kT_a - 1/kT_s)} = t_s/t_a. \tag{11}$$

Thus f_s of eq. (8) is given by

$$I = f_s(t_s) = F_s(t_s/m) \tag{12}$$

with m from eq. (11). Specific properties then follow immediately from eq. (12); for example, the time t_s^* when I reaches a particular value I^* (e.g., failure to meet system requirements) is

$$t_s^* = mF^{-1}(I^*). \tag{13}$$

The procedure of eqs. (9) to (13) assumes that measurements can be made at the system temperature $(T_r = T_s)$ even though there may not be enough time for perceptible aging at T_s. However, when it is undesirable (e.g., thermal-cycling-induced strain and degradation effects occur) or inconvenient to periodically drop the device temperature all the way down to T_s, a theoretical extrapolation of a parameterized degradation model is used instead of the empirical procedure of eq. (9). As a simple example, let I be the threshold current i of a hypothetical laser. Suppose that i depends exponentially on temperature at $t = 0$ and suppose that at elevated temperatures i increases linearly in time, doubling in a period that depends upon the aging

temperature but not upon the reference (measurement) temperature. Then, if the Arrhenius rule is valid, eq. (6) takes the specific form of eq. (14),

$$I = Ce^{T_r/T_0}(1 + R_\infty e^{-E_A/kT_a}t_a)$$

$$= Ce^{T_r/T_0}(1 + R_\infty e^{-E_A/kT_r}mt_a), \tag{14}$$

where C, T_0, R_∞, and E_A are parameters, the parameter T_0 being the temperature range over which I changes by a factor of e at $t = 0$. If E_A is determined from population measurements, then the parameter R_∞ can be found from the initial degradation rate at some elevated T_a. (See Section VIII if E_A must be found on an individual-device basis.) Extrapolation consists of invoking eq. (11) and assuming that eq. (14) is valid for T_r equal to T_s. Equation (8) is finally obtained by setting T_r equal to T_s and t_a equal to t_s in eq. (14).

In summary, if measurements can be made at the system temperature T_s, then the shape of $I(t)$ can be measured at T_s and we need only extrapolate the rate at which the $I(t)$ curve is traced out. If measurements at T_s are not possible, then the shape must also be inferred by extrapolation. Sometimes the extrapolation of the shape presents no problem, as when I degrades linearly in time. In other applications such an extrapolation is virtually impossible, and measurements must be made at the system temperature. For example, if I is the optical-power coupling efficiency[22] between a laser and a fiber, then I may or may not pass through a maximum according to whether the long-term creep of the alignment does or does not pass through a point of better alignment than the initial position.[23] Because of the changed alignment caused by differential thermal expansion, a maximum in $I(t)$ under isothermal aging may occur only over a limited temperature range, i.e., the temperature dependence of the function $I(t)$ can be too complicated to extrapolate with confidence.

Most of the comments on the activation factor that were given in the previous section are, of course, equally applicable to the activation energy. Thus, if the activation factor varies from device to device, so does the activation energy. Where other authors have defined their R or m differently than we have, so will the value of their E_A or m differ from ours.

The Arrhenius rule is not unique; for example, variations on the Eyring rate such as[24,25]

$$R(T) \sim T^{-1}e^{-E_A/kT} \tag{15}$$

or eq. (35) are sometimes endorsed. If alignment degradation is caused by material creep, then the Arrhenius rule is replaced by the empirical temperature dependence of creep for that material, as moderated by

any temperature dependence of the driving stress that may arise, for example, from imperfectly compensated differential thermal expansion. Also temperature is just an example here. For a different stress, or for a combination of stresses, a similar discussion applies except that R may have a different functional form. As an arbitrary example, recent experiments[26] suggest that under constant optical power the dark-line-defect mechanism of InGaAsP laser degradation may approximately scale according to the following two-parameter (E_A and J_0) dependence on the two stresses T and J,

$$R(T, J) \sim e^{J/J_0 - E_A/kT}, \tag{16}$$

where J is the junction current density. As another example the degradation of Ti-Au thin films by electromigration scales approximately as $R \sim J^4$ where the stress J is the current density.[11]

V. EXTRAPOLATION FACTOR AND ENERGY

In the preceding two sections the effect of a valid accelerating stress (or stresses) is characterized in terms of an activation factor and an activation energy. Following footnote 7 of Ref. 7 again, we give here an alternative characterization of the same accelerating stress (say, temperature) in terms of an *extrapolation factor* and an *extrapolation energy*. Both characterizations involve a comparison based on aging at two temperatures (T_a and $T_{a'}$), but with the extrapolation characterization the effect of each temperature is represented in terms of measurements made *at that temperature*. Specifically, at least one observable rate $\mathscr{R}$ is defined. The main criterion for the choice of $\mathscr{R}$ is that it act as some measure of a device's life, and thus the units of $1/\mathscr{R}$ normally contain time to the first power. For example, $\mathscr{R}$ might be the reciprocal of the lifetime t^* when an observable I no longer meets a system specification—cf. eq. (13)—or $\mathscr{R}$ might be the predeployment isothermal degradation rate $\partial I/\partial t$, where I is the current of a laser at threshold or the current at some other value of the optical power.

In analogy to eq. (2), an *extrapolation factor* m_E is defined by

$$m_E(T_a, T_{a'}) = \mathscr{R}_a/\mathscr{R}_{a'}, \tag{17}$$

where, in contrast to eq. (2), $\mathscr{R}_a = \mathscr{R}(T_a)$ is the rate when both aging and measurements are at T_a. If $\mathscr{R}$ is the initial or instantaneous rate, the form of eq. (3) would again be applicable

$$m_E = \frac{\mathscr{R}_a}{\mathscr{R}_{a'}} = \frac{\partial I_a/\partial t_a}{\partial I_{a'}/\partial t_{a'}} \tag{18}$$

except now, in contrast to eq. (3), it is understood that $\partial I_a/\partial t_a$ is the

isothermal degradation rate with both aging and measurement at T_a. That is, for the case $T_r = T_a$ of Fig. 2, one has

$$\frac{\partial I_a}{\partial t_a} \approx \frac{I_3 - I_2}{\Delta t_a}, \tag{19}$$

which contrasts with eq. (4).

For extrapolation one shows empirically, over some higher-temperature range, or theoretically that $\mathscr{R}$ is of the Arrhenius,

$$\mathscr{R}(T) = \mathscr{R}_\infty e^{-E_E/kT}, \tag{20}$$

Eyring, or other (e.g., eq. [35]) form where the parameter E_E is called the *extrapolation energy*.[7] If, for example, $1/\mathscr{R}$ is the lifetime t^*, and t^* is the quantity of interest, then setting T equal to T_s in eq. (20) completes the extrapolation. Alternatively, $\mathscr{R}$ may appear in some analytic isothermal model

$$I = g(T, \mathscr{R}t), \tag{21}$$

for which E_E and any other parameters are determined by data fitting at higher temperatures. Equation (21) is then extrapolated by assuming validity for $T = T_s$ and $t = t_s$.

We emphasize that m_E and E_E are rather arbitrary quantities. For example, suppose that the operational lifetime of each device from a homogeneous population with a range of initial currents corresponds to the time t^* when that device draws, at the system temperature T_s, a current i equal to the maximum output (100 mA) of its power supply, and suppose that a particular device has an initial current i_0 of 50 mA. For that device the failure criterion at T_s can be expressed in various ways including $i = 100$, $i - i_0 = 50$, or $i = 2i_0$. Reversible (no aging) heating of the device to the accelerating temperature T_a raises its initial current to 70 mA. Do we define the device's lifetime at T_a by the condition $i = 100$, $i = 120$ (i.e., $i - i_0 = 50$), or $i = 140$ (i.e., $i = 2i_0$)? Perhaps each of these definitions would yield a smooth temperature dependence for $1/\mathscr{R}(T) = t^*(T)$ which could be extrapolated, but clearly each definition would have its own E_E. Furthermore, if two devices (with unequal initial currents) had the same E_A, they would typically have unequal E_E's with most of these definitions. Thus one cannot speak of "the" extrapolation energy as if it were on an equal footing with the activation energy. Rather, there is, in general, a separate extrapolation energy for each definition of each observable. Even with a given definition of the observable and a common degradation mechanism, one typically expects each device to have its own extrapolation energy.

The above arbitrariness can be resolved if there exists a criterion at T_a that corresponds, for every device (every initial current), to 100 mA

at T_s. (See Section VI for an example.) Often, however, such a criterion cannot be found, except, possibly, in a rather approximate sense. Alternately one can, of course, periodically cool the device to see when it draws 100 mA at T_s, but then one has simply abandoned the extrapolation method in favor of the activation method.

In the case where the threshold current i of a laser is the dependent observable, and $\mathscr{R}$ is the initial isothermal degradation rate $\partial I/\partial t$, arbitrariness results from the fact that I could be chosen as i, or as i^2, or as $A(T) + B(T)\ln i$, etc. Sometimes, this ambiguity can be partially resolved with the requirement that $\mathscr{R}$ scale time rather than i, i.e., by finding an $I = f(i)$ which changes linearly in time for all devices in the population. Thus, if isothermal degradation is of the form $i \sim \exp(\mathscr{R}t)$, then I is chosen to be a linear function of $\ln i$, as in the third choice, above. In practice, however, experimental error, limited observation time, temperature dependence of f, device-to-device differences, and small differences between initial and final values often make it difficult to show that such an $f(i)$ exists. Even when $f(i)$ can be found, E_E may vary, in the case of a laser, according to whether i is chosen as the threshold current ($P \approx 0$) or the current at some larger value of the stimulated optical power P.

In summary, m_E and the extrapolation energy E_E are model- and observable-dependent parameters which typically have no recognizable physical significance. Nevertheless, they often form the basis of useful extrapolations when used self-consistently with their defining models. In contrast to the activation method, where multiple dependent observables on a common device are expected to have a common activation energy for a valid accelerating stress, the extrapolation method typically yields a different extrapolation energy for each dependent observable.

VI. AN EXAMPLE

It is instructive to take a simple example and contrast the activation and acceleration energies. Assume that the threshold current i, or the current at some other optical power, of an idealized laser conforms to the isochronal relation

$$i(t, T) = i(t, T_r)e^{(T-T_r)/T_0}. \tag{22}$$

That is, at any time, i depends exponentially on T with a characteristic temperature interval T_0 that is unaffected by the aging history. (The accuracy of eq. [22] over a useful temperature range and the approximate invariance of T_0 during aging have been reported for certain types of lasers.[6,12])

Without specifying the time dependence of the isothermal aging law, we define a possibly-time-dependent $b_{a'}$ by the requirement that

isothermal aging along the solid-line path 1–5 in Fig. 4 causes $i_1 = i(t = 0, T_{a'})$ to evolve into i_5, i.e., $b_{a'}i_1$ is defined as the isothermal increase in i_1,

$$i_5 = i_1 + b_{a'}i_1. \tag{23}$$

If $b_a i_2$ is similarly defined as $i_3 - i_2$, i.e.,

$$i_3 = i_2 + b_a i_2, \tag{24}$$

then it follows from eqs. (22) and (24) that aging around the dashed-line path 1–2–3–4 evolves i_1 into i_4 given by

$$i_4 = i_1 + b_a i_1. \tag{25}$$

But $i_4 = i_5$ (Fig. 4b), and hence from eqs. (23) and (25)

$$b_{a'} = b_a. \tag{26}$$

That is, regardless of the form of the isothermal degradation law, equal amounts of degradation at T_a and $T_{a'}$ correspond to equal percent changes in the isothermal currents.

To find the instantaneous activation factor m, write eqs. (23), (26), and (24), respectively, as

$$b_{a'} = \frac{1}{i_{a'}} di_{a'} = \frac{\partial \ln i_{a'}}{\partial t_{a'}} dt_{a'}$$

$$= b_a = \frac{\partial \ln i_a}{\partial t_a} dt_a, \tag{27}$$

where $\partial \ln i_a/\partial t_a$ is the isothermal degradation rate of $\ln i_a$, with aging and measurement at T_a. From eqs. (1) and (27)

$$m = dt_{a'}/dt_a = \frac{\partial \ln i_a \partial t_a}{\partial \ln i_{a'}/\partial t_a} = \frac{e^{-T_a/T_0}\partial i_a/\partial t_a}{e^{-T_{a'}/T_0}\partial i_{a'}/\partial t_{a'}}, \tag{28}$$

where eq. (22) was used for the last equality. Following eq. (2) we interpret m as the ratio of two one-temperature rates R. Thus from eqs. (2) and (28), two of the many correct expressions for R are the isothermal rates

$$R(T) = \partial \ln i/\partial t \tag{29}$$

and

$$R(T) = e^{-T/T_0}\partial i/\partial t, \tag{30}$$

where aging and measurement are both at temperature T. If R obeys the Arrhenius rule, eq. (10), then

$$E_A = k \frac{\ln(R_a/R_{a'})}{T_{a'}^{-1} - T_a^{-1}}, \tag{31}$$

where $R_a = R(T_a)$. An important point here is that eq. (22) has been used to avoid the necessity for the periodic temperature lowering of Fig. 2 even though T_0 may vary appreciably from laser to laser; that is, m, E_A, and R are expressed in terms of rates measured at the aging temperatures.

This example may be equivalently interpreted as follows. Equations (29) and (30) show that if the observable I is chosen as $\ln i$ or as $e^{-T/T_0}i$ (where i is the threshold current at T), then the extrapolation energy associated with the isothermal rate $\mathscr{R} = \partial I/\partial t$ is equal to the activation energy.

If, however, the observable I had been arbitrarily chosen, say as i, then measurement of the isothermal degradation rate

$$\mathscr{R} = \partial i/\partial t \tag{32}$$

at the two temperatures T_a and $T_{a'}$ and application of the Arrhenius rule would have lead to the following value for the extrapolation energy

$$E_E = k \frac{\ln(\mathscr{R}_a/\mathscr{R}_{a'})}{T_{a'}^{-1} - T_a^{-1}}$$

$$= E_A + kT_aT_{a'}/T_0 \tag{33}$$

$$= E_A + 0.14 \text{ eV}, \tag{34}$$

where eqs. (29), (31), and (32) were used. The example of eq. (34) is based on the values $T_{a'} = 333\text{K}$, $T_a = 343\text{K}$, and $T_0 = 70\text{K}$. Equation (34) shows that the quoted value of an extrapolation energy can depend significantly on the choice of I. Also, if T_0 varies appreciably from laser to laser, as it often does, then eq. (33) shows that at least one of the energies E_A and E_E must also vary appreciably from laser to laser.

Although it would not be apparent from aging at only two temperatures, it is clearly true that $R(T)$ and $\mathscr{R}(T) = \partial i/\partial t$ cannot both obey the Arrhenius rule. If, for example, $R(T)$ conforms to Arrhenius, then the corresponding temperature scaling rule for $\mathscr{R}$ is

$$\mathscr{R}(T) = \partial i/\partial t \sim e^{-E_A/kT+T/T_0}. \tag{35}$$

This suggests strongly that the application of the Arrhenius rule should be justified in each context by measurement of $\mathscr{R}(T)$ or $R(T)$ at a number of temperatures. Alternately, the definition of I (and hence of $\mathscr{R} = \partial I/\partial t$) could be determined by the requirement that the Arrhenius (or other) rule be applicable to $\mathscr{R}$.

Parenthetically we note from eq. (22) that a useful alternative form of eq. (28) is

$$m = \left(\frac{1}{i_a'}\frac{\partial i_a}{\partial t_a}\right) \bigg/ \left(\frac{1}{i_{a'}'}\frac{\partial i_{a'}}{\partial t_{a'}}\right), \tag{36}$$

where i'_a and $i'_{a'}$ are the currents evaluated at a common t' (any convenient time such as the moment of temperature change in a step stress) while $\partial i_a / \partial t_a$ and $\partial i_{a'} / \partial t_{a'}$ are evaluated at one or more other common times t.

VII. COMPARISON OF ACTIVATION AND EXTRAPOLATION METHODS

We compare the activation method (Sections III and IV) and the extrapolation method (Section V) by noting that each has advantages over the other. For example, if the quantity of interest is the lifetime τ defined as the time t^* when a laser will no longer put out a given amount of stimulated optical power P despite arbitrary adjustments of the drive current, and if the stimulated power plays an important role in the degradation process, then temperature acceleration of t^* is not possible by the activation method. The reason is that t^*_r at the reference temperature $T_r = T_s$ corresponds to an aging time t^*_a beyond the time when the laser will lase at the accelerating temperature T_a. (Typically, a laser that has died at T_a will still lase when lowered to T_r.) But, by assumption, operation without optical power is not relevant, and so the activation method cannot provide acceleration of t^*. (On the other hand, the extrapolation method may well yield a useful extrapolation of that lifetime τ defined as the time t^+ ($\neq t^*$) at each temperature when power P can no longer be obtained at that temperature.[27]) The activation method can also be ruled out if temperature cycling is unacceptable (because of transients or unwanted cycling degradation) or inconvenient.

By contrast, the coupling efficiency or a low-optical-power laser for which the current density, J, but not the stimulated optical power P, is significant to degradation can undergo accelerated aging above the maximum temperature for lasing by the activation method. Typically, however, there is no practical degrading observable to monitor above the maximum temperature for lasing, and thus the extrapolation method is inapplicable. Also, as suggested by the dissimilar shapes of Fig. 4b or of Fig. 6, the extrapolation method is of limited value if processes (e.g., shunt-path effects) are operative at T_a but not at the system temperature T_s. Finally, the arbitrariness of the numerical values of an extrapolation energy is objectionable to investigators seeking connections with microscopic degradation mechanisms.

Although the activation energy is less arbitrary than the extrapolation energy, neither energy can claim to be any sort of fundamental constant. For example if I is the threshold current i, and if degradation consists of dark-line-defect formation,[9] then I increases because of additional nonradiative recombination in the dark line. Furthermore, to overcome the optical loss caused by the dark line the quasi-Fermi level separation must be increased, and this increases the radiative

and nonradiative recombination even in the line-free regions of the laser. The increased separation also increases both the minority carrier leakage over the double-heterostructure barrier[16] and the shunt current around a buried active layer.[6,28] Each of these four contributions to the increase in i may have its own temperature dependence. Thus, except for fortunate cases where a single response dominates, any macroscopic black-box activation energy necessarily reflects multiple microscopic processes. In so far as these processes are not multiplicative in their effect on I, they also induce curvature in an Arrhenius plot ($\ln R$ versus $1/T$).

A further arbitrariness in any black-box activation or extrapolation energy arises from imprecision in the value of "the" temperature. In the case of dark-line laser degradation[9] a line causes pronounced, and typically unmeasured, local heating. Even with a uniform mode of degradation[10] it is believed that large temperature gradients exist within the active layer.[29]

If temperature T and some other stress such as the current density J are both accelerants, and if the acceleration rate is not a product function $[R(T, J) \neq f(T)g(J)]$, then when only T is varied the deduced value of E_E may depend implicitly upon the fixed value of J.

In summary, despite their limitations the activation method and the extrapolation method are each useful under conditions when the other is inapplicable. Thus, we conclude that both methods will survive indefinitely.

VIII. DIFFERENTIAL FORMS

In the reliability literature theoretical models of device degradation are usually presented in terms of isothermal-aging expressions of the form

$$I = f(t, T), \tag{37}$$

where I is an observable measured at T after aging at T from $t = 0$ to $t = t$. (If $T_a = T_r$, then eq. [14] becomes an example of eq. [37].) In this section we emphasize the limitations of eq. (37) and advance alternatives.

The main problem with eq. (37) is its incompleteness. For example, knowledge of $f(t, T)$ for every t and T is not sufficient information to calculate even the activation factor m. This can be seen as follows. At $t = 0$ one can move up the path 1–2 in Fig. 4a ($I_1 \rightarrow I_2$ in Fig. 4b) by raising the value of T in eq. (37) to T_a. Path 2–3 ($I_2 \rightarrow I_3$) can then be followed by increasing t from $t = 0$ to $t = t_a$ in eq. (37). But eq. (37) is insufficient to follow path 3–4, i.e., insufficient to deduce I_4. That is, if the values $t = t_a$ and $T = T_{a'}$ corresponding to I_4 are inserted into eq. (37), this (isothermal) equation yields that I_4' that results from tra-

versing the isothermal solid-line path 1–4 in Fig. 4a, not the I_4 that results from the dashed path 1–2–3–4. Alternately, we could try to validate eq. (37) down the path 3–4 by asserting that $I_4' = I_4$. But we are then saying that the value of I_4 does not depend upon its history (i.e., does not depend upon the route in $t - T$ space leading to a given end point). In such a case "accelerated" aging (route 1–2–3–4) actually accomplishes no acceleration because it yields the same result as normal aging (solid-line route 1–4 in Fig. 4a). Such path-independent (nonaccelerating) stresses may well exist, but they are not the stresses of interest in accelerated-aging theory. Thus, while eq. (37) is useful for the special case of isothermal aging from a particular starting time (i.e., a particular starting physical state of degradation), we abandon eq. (37) as a basis for formulating degradation models.

In principle the sequence of horizontal and vertical paths in $t - T$ space that describe step stressing could be followed analytically if the isothermal relation, eq. (37), were supplemented with an isochronal relation like eq. (22). Both relations must, however, be valid for arbitrary starting time and temperature, i.e., after an arbitrary aging history. (Expressions of the class of eq. [37] in the literature are often valid only for aging from a particular state of degradation at $t = 0$.) Because of the complexity arising from the arbitrary starting time and because curved paths (e.g., eq. [47]) are of interest, we shall not follow this approach.

The simplest path-dependent model for the observable I is the *differential form*

$$dI = \mathscr{R}dt + \theta dT, \qquad (38)$$

where all variables in eq. (38) are evaluated at the same time and temperature. Sometimes it is sufficient to assume that the partials $\mathscr{R}$ and θ are *state functions*, i.e., functions which depend only on the present state (T, I) of the system. However, we allow for an explicit time dependence and write

$$\mathscr{R} = \mathscr{R}(t, T, I) = \partial I/\partial t \qquad (39)$$

$$\theta = \theta(t, T, I) = \partial I/\partial T. \qquad (40)$$

Such an explicit time dependence arises, for example, when moisture diffuses into a hermetically sealed can at a rate that is independent of the aging history $T(t)$ of the enclosed device. (In some such cases the explicit time dependence of θ and $\mathscr{R}$ could be removed by adding humidity as an additional independent variable.)

As an example if I is the threshold current i and if

$$\theta = I/T_0, \qquad (41)$$

then the isochronal ($dt = 0$) integration of eq. (38) yields eq. (22). If

$$\mathscr{R} = B(T)I, \tag{42}$$

then the isothermal $(dT = 0)$ integration of eq. (38) yields $I = I_i \exp(Bt)$, where initially $I_i = I(t = 0; T)$. The linear isothermal law $I = I_i + B(T)t$ (e.g., the $T_r = T_s$ case of eq. [14]) can be represented by

$$\mathscr{R} = B(T), \tag{43}$$

in eq. (38).

To find I at any point (t, T) along the curved solid-line route $T(t)$ in Fig. 7, express the route in differential form as $dT = (dT/dt)dt$ so that eq. 38 becomes

$$dI = (\mathscr{R} + \theta dT/dt)dt. \tag{44}$$

Integration of eq. (44) yields the path-dependent generalization of eq. (37) that we seek. If the route is given as $t(T)$, then

$$dI = (\theta + \mathscr{R}dt/dT)dT. \tag{45}$$

In practice the path is often given implicitly. Consider, for example, the commonly occurring runaway curved path that results when, for convenience, a so-called "isothermal" aging experiment is actually performed at a constant heat-sink temperature rather than at a constant temperature T of the relevant region within a device (e.g., the active layer of a laser). If $\mathscr{R}$ and θ are state functions and if the temperature T of the active layer of a laser operated at constant optical power is a known function $g(I)$ of the degrading (increasing) drive current I and of the constant heat-sink temperature,* then eq. (38) can be written

$$dI = \mathscr{R}[g(I), I]dt + \theta[g(I), I](dg/dI)dI, \tag{46}$$

which separates into

$$\frac{1 - \theta dg/dI}{\mathscr{R}} dI = dt \tag{47}$$

and reduces evaluation of $I(t)$ and of the curved-path $T(t) = g[I(t)]$ to quadrature.

If eq. (44) or (45) yields a path-independent integral, i.e., eq. (37), then eq. (38) is an exact differential and $\mathscr{R}(t, T, I)$ and $\theta(t, T, I)$ satisfy the following test:

$$\theta \partial \mathscr{R}/\partial I + \partial \mathscr{R}/\partial t = \mathscr{R}\partial \theta/\partial I + \partial \theta/\partial t. \tag{48}$$

For example, insertion of eqs. (41) and (43) into eq. (48) and integra-

---

* In the usual case the thermal resistance has the constant value Ω. Then $T = g(I) = I^2\Omega/2 + T_{hs}$, where T_{hs} is the heat-sink temperature.

tion of the resulting equation for B reveal that temperature is not an accelerating stress if

$$B(T) = \text{const.} \times e^{T/T_0}. \tag{49}$$

A change of variable from I to I', i.e.,

$$I = g(t, T, I'), \tag{50}$$

is sometimes useful for simplifying $\mathscr{R}$ or θ. Then eq. (38) becomes

$$dI' = \mathscr{R}'dt + \theta'dT, \tag{51}$$

where

$$\mathscr{R}'(t, T, I') = \frac{\mathscr{R}(t, T, g) - \partial g/\partial t}{\partial g/\partial I'} \tag{52}$$

and

$$\theta'(t, T, I') = \frac{\theta(t, T, g) - \partial g/\partial T}{\partial g/\partial I'}. \tag{53}$$

As a trivial example of eq. (50), recall that in Section VI the variable change to $I' = \ln i$ from $I = i = \exp I' = g(I')$ transformed $\theta = I/T_0$ into $\theta' = T_0^{-1}$. A more interesting example occurs in the case of Fig. 7,
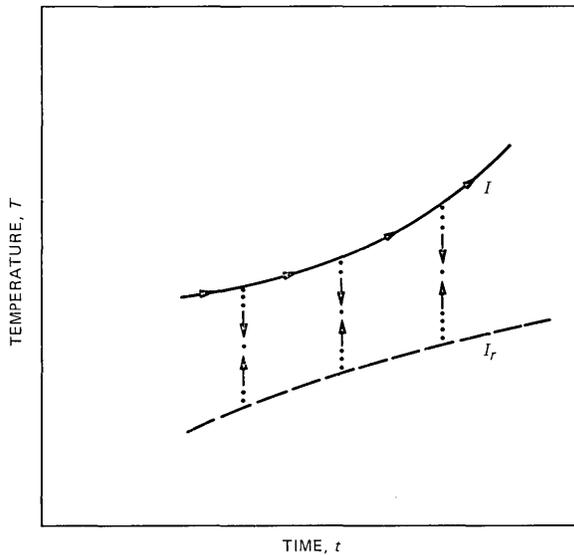
TIME, t

Fig. 7—Corresponding to the evolving value of an observable I, as a component is aged along the solid-line route in time-temperature $(t - T)$ space, is the evolving reduced observable I_r which is measured if the component is periodically and momentarily dropped to the dashed-line path and returned to the solid-line path (down and up each dotted line).

where I is the degrading value of an observable which is aged along the solid-line path (e.g., an isothermal or eq. [47]) in $t - T$ space and $I' = I_r$ is the *reduced observable*, i.e., the value of I when the component is momentarily and periodically cooled to the dashed-line reference path in $t - T$ space. In this case g of eq. (50) is, from eq. (38), the integral of

$$dI = \theta(t, T, I)dT, \tag{54}$$

with integration up the vertical dotted path (between the limits I_r and I) in Fig. 7. For example, if $I = i$ and $\theta = I/T_0$, then integration of eq. (54) shows that eq. (50) becomes eq. (22). If the dashed curve in Fig. 7 were a horizontal line (at T_r), then $\mathscr{R}'$ is the R of Section III, and eq. (51) reduces to

$$dI_r = Rdt.$$

Thus, if $\theta = I/T_0$, eq. (52) becomes

$$R = \mathscr{R}' = \mathscr{R}e^{-(T-T_r)/T_0} \tag{55}$$

in agreement with eqs. (30) and (32) (allowing for the arbitrary choice in eq. [30] of the multiplicative constant in R).

For multiple dependent observables (e.g., $I_1 =$ current, $I_2 =$ voltage, $I_3 =$ peak wavelength, $I_4 =$ thermal resistance, $I_5 =$ strain, $I_6 =$ optical power in the fundamental mode, $I_7 = \cdots$, etc.) and multiple independent observables (e.g., $T_1 =$ temperature, $T_2 =$ total optical power, $T_3 =$ humidity, etc.), one understands θ as a matrix while I, $\mathscr{R}$, and T are vectors in eq. (38). In principle the differential form is completely general with θ and $\mathscr{R}$ as state functions, because if there are enough observables to completely specify the state of a classical system, then the laws of physics can be written as first-order differential equations with state-function coefficients. If, for a set of independent observables, one finds a set of dependent observables such that eq. (38) is an adequately accurate description with θ and $\mathscr{R}$ state functions, then the observables in question may be said to form a *complete macroscopic set*.[30]

Instead of increasing the number of dependent observables, one may seek a complete description by going to higher-order differential equations or by including an integral over the history $T(t)$. We will not consider such generalizations any further here except to note that when a complete theoretical model in two (t and T) or more independent variables appears to be quite complex, then the empirical one-variable reduced-observable method of eqs. (6), (12), and (54) becomes particularly appealing.

IX. SAMPLED-POPULATION METHODS

The essential idea of sampled-population methods is that the characterization of components going into service is based upon aging properties measured on a statistically equivalent sample rather than on any aging measurements on the very components going into service. We outline here two approaches that are differentiated from each other by the fact that one approach is based on aging to failure, while the other approach is based on aging only through the time period of interest (system life, guarantee period, etc.).

Aging to failure, which is particularly appropriate for components which are likely to be replaced (mean component life $\lesssim$ system life), or which fail suddenly[1] (no continuously degrading observable can be found), has been treated exhaustively in the semiconductor reliability literature.[2,3,13,31-33] Typically a sample set of components from the population is aged isothermally to failure at an elevated value of the temperature or other stress, and the observed lifetimes τ (as defined by either the activation or extrapolation method) are then used to estimate the parameters in an analytical function $F(\tau)$ that is assumed to describe the distribution of lifetimes in the population. Good fits to the data are often obtained with the exponential,[34] Weibull,[34] or lognormal distributions[34] or with mixtures of these as in the double-lognormal distribution.[8] Also, we believe that the burned-in and truncated forms of these distributions will prove applicable, as discussed in conjunction with Fig. 3. Because it is so often used and because many of its properties cannot be expressed in terms of a finite number of elementary functions, the lognormal distribution (ln τ is normally distributed) has had its implications developed in particularly great detail.[31-33]

When the temperature (or other stress) is believed to be a valid accelerating stress, as after the selective elimination of flawed components having a secondary failure mechanism, it is customary to measure the isothermal mean life $\langle \tau \rangle$ (or $\langle \ln \tau \rangle$) at two or more temperatures (using separate statistically equivalent samples at each temperature). Then the rate $1/\langle \tau \rangle$ is extrapolated by assuming that it follows the Arrhenius rule (eq. [10]) with a single population value for E_A. The standard deviation in ln τ is assumed to be temperature-independent, as follows from eq. (6) when E_A has a common population value. When ln τ is the ordinate, this amounts to a vertical displace-ment of the distribution (Fig. 2a) or to relabeling the ordinate, as shown on the right-hand side of Fig. 1. Validly accelerated distributions other than the lognormal are similarly scaled by noting from eq. (6) that temperature extrapolation is simply a scaling of time. If, as discussed in conjunction with Figs. 2b and 5, the activation energy varies from component to component, then the variance may also

depend upon the temperature. Analogous procedures are commonly used with the extrapolation-energy method, although the justifying rationales, if any, are necessarily based on specific models of degradation or specific observations (because eq. [6] is not applicable and because it is rarely easy to argue a priori that all components should have a common extrapolation energy).

The limitations of cost or of available time and the long life of some components, even under accelerated-aging conditions, often prevent aging more than a small fraction of the sample components to failure. Under such circumstances the method just outlined may still be feasible, in principle, if the lifetimes of the longer-lived components can be estimated on the basis of an established correlation with one or more early component properties. For example, significant correlation between lifetime and initial degradation rate has been reported for some types of lasers[7] and thin-film conductors.[11] Initial properties other than an aging rate may also exhibit correlation with lifetime, as in Refs. 35 and 36.

For a high-reliability component (mean component life $\gg$ system life) the cost and time for aging a sample to even the mean life can easily exceed, by more than an order of magnitude, the cost and time for aging only through the *effective system life* (= the *temperature-contracted system life*, i.e., the system life divided by the activation or extrapolation factor). This expenditure of time and money for aging past the effective system life may not be justified if the consequences of practical interest depend only on the expected number of failures during the system life. In short, as an alternative to measuring the distribution of times for a given amount of degradation (i.e., failure), it is sometimes preferable to measure the distribution in the amount of degradation for a given time. (That time may be the effective system life, or a longer time with interpolation back to the effective system life if the degradation is still in the noise after only the effective system life, or that time may simply be the available time in a crash program with an extrapolation out to the effective system life.)

Figure 8 is an example which contrasts the two approaches of this section. Consider a population of hypothetical lasers. A laser has a predeployment threshold current i_0, and the isothermal threshold current i increases approximately linearly in time up to a failure value i_f, which is defined as the lesser of $2i_0$ (a compensating-circuit limitation) and 100 mA (a power-supply limitation). Under these conditions the observable I, defined by

$$I = \frac{i - i_0}{i - i_f},\tag{56}$$

has convenient properties; i.e., I increases approximately linearly from
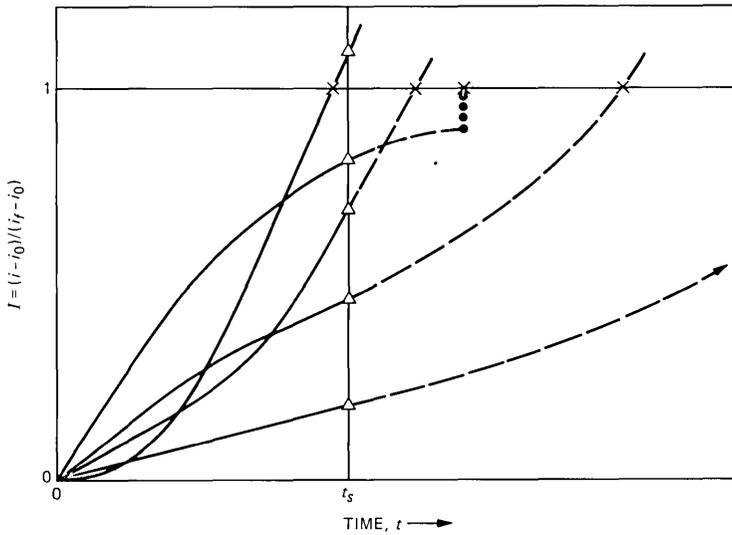
Fig. 8—Idealized isothermal degradation curves for semiconductor lasers. Here i_0 and i_f are the initial and failure values of the threshold current (or the current at some other value of the optical power). The effective system life is t_s.

$I_0 = 0$ to $I = 1$ (failure) for every laser regardless of its i_0 and its isothermal degradation rate $\partial i/\partial t$. Also, as exploited in Section X, $(\partial I_0/\partial t)^{-1}$ is the initial estimate of the component life τ; thus $\partial t/\partial I_0$ has statistical properties which are similar to those of τ, and a similar methodology can be used. Each curve in Fig. 8 represents the observed degradation of one laser. For the lifetime-distribution method of Fig. 1, the relevant data points are the ×'s in Fig. 8 that fall along the failure line ($I = 1$). For the high-reliability-component method the data points are the triangles which fall on the effective-system-life line ($t = t_s$), i.e., the values $I_s = I(t = t_s)$. For the latter method it is not necessary to generate the dashed-line portion of the data. In a typical case the distribution of $1/I_s$ values is roughly similar to the distribution of lifetime (τ) values and can be treated similarly to the treatment of the cumulative distribution $F(\tau)$. The similarity in the distributions of τ and $1/I_s$ need not be overly close, however, because the objective here is to estimate the probability of failure before t_s rather than to estimate later properties such as the mean lifetime $\langle \tau \rangle$. (Alternatively, if data are available at intermediate times and if the $I(t)$ curves are fairly smooth, the dashed sections of the curves can be estimated by extrapolation from the solid sections. Then the high-reliability data can be analyzed as if they were replacement data. The inaccuracy in the ×'s at large t may not be serious if they prove to be a smooth extension of the early ×'s and if the implications of the analysis are restricted to the range $0 < t < t_s$).

The two methods of this system have an important attribute that should be emphasized. They are useful methods whether or not the various components in the sample age with a common functional dependence on time. (Of course, when all components do have a common parameterized functional dependence on time—e.g., eq. [14] —and when the number of parameters is not too large for confident estimates, then standard statistical methods provide a third approach in which the observed degradation of the sample components is used to estimate the distribution of parameters over the population of components.)

X. TRUNCATION

Sampled-population methodology is sometimes too crude because it ignores important predictors of a given component's life. As previously noted these predictors include the predeployment degradation rate[7,11] and other initial properties of that component.[35,36] Utilization of this additional class of data implies that each component has its own statistical description including, possibly, its own acceleration factor or activation energy. If the step-stress measurement error is larger than the component-to-component difference in the acceleration factor, then the best estimate of the acceleration factor or activation energy for each component is usually the common population value inferred from isothermal aging of sample populations at different temperatures. If the converse is true, then the best estimate for a component is usually its own value inferred from step stressing that component before deployment. Hereafter we discuss the truncation method in terms of data at the system temperature. Part or all of these data may, of course, actually be inferred, in the manner previously discussed, by extrapolation from accelerated measurements at elevated stress levels.

We consider first the case of replacement components characterized by a cumulative lifetime distribution $F(\tau)$. Ideally, the initial degradation rate is an exact predictor of a component's life, and one simply uses the initial rate to identify and reject all components with lifetimes less than some value τ'. If $F(\tau)$ is the cumulative distribution (failed fraction at time equal to τ) and if $S(\tau) = 1 - F(\tau)$ is the surviving fraction (fraction of the components with lifetimes longer than τ) of the population, then after *truncation* (discarding of the short-lived components) the distribution $F' = 1 - S'$ of the retained components is given by

$$S'(\tau) = 1 - F'(\tau) = 1 \qquad \tau \leq \tau'$$
$$= S(\tau)/S(\tau'), \qquad \tau \geq \tau', \qquad (57)$$

as exemplified in the $T_{y=0.4}$ curve of Fig. 3.

Equation (57) is too idealized for most applications because it implies that 100-percent reliability can be achieved simply by choosing τ' greater than the system life. To be more realistic one must first measure the correlation between the initial degradation rate and the lifetime, as has been done, for example, in an important experiment by English et al.,[11] who studied, for Ti-Au thin-film conductors, the correlation between the film lifetime τ and the initial rate of degradation (increase) in the resistance of the conductor. From the data points in their Fig. 7 (~150 conductors), it appears that 16 percent of their conductors had failed by 400 hours, while the final 16 percent failed after 2500 hours. Thus as a very rough estimate we find that the standard deviation in $\ln \tau$ would be $\sigma \approx (1/2) \ln(2500/400) = 0.92$ if the conductors were deployed on a sampled-population basis. However, a similar analysis of their figure shows that σ of the quantity $\ln\tau - \ln\tau_{est}$ is approximately 0.75 for a conductor with an estimated lifetime τ_{est} inferred from its initial degradation rate. This means that the early-failing components could be identified and eliminated with some success. For example, for their data, the time of the earliest failures is raised by about three orders of magnitude (from ~11 hours to ~10^3 hours) if the 60 percent of the conductors of their actual population with the highest initial degradation rates are discarded (cf. Fig. 3).

To quantify the preceding let $\rho(R_0, \tau)d\tau$ be the conditional probability of lifetime between τ and $\tau + d\tau$ for a component with initial (i.e., post-stabilization, predeployment) degradation rate R_0 (or $\mathscr{R}_0$), and let $P(R_0)dR_0$ be the fraction of the components with initial rates between R_0 and $R_0 + dR_0$. If components with initial rates greater than R_0' are discarded, then the (truncated) distribution $F'(\tau)$ of the components retained for deployment is given by

$$dF'(\tau)/d\tau = \int_{-\infty}^{R'} \rho(R_0, \tau) P(R_0)dR_0 \left/ \int_{-\infty}^{R'} P(R_0)dR_0 \right. \quad (58)$$

For perfect correlation, i.e., $\tau = g(R_0)$ and $\rho = \delta[g(R_0) - \tau]$, eq. (58) reduces to eq. (57). Here $\delta[\ ]$ is the delta function.

The success of predeployment inspections implies that it is sometimes possible to establish a correlation between lifetime and some initial property other than the degradation rate, such as the initial value of the degrading observable I, or the thermal resistance of a laser bond, or the thermally induced nonlinearity in the current-voltage characteristic of a nominally linear circuit element,[35] or the stripe width[36] (in the case of incipient formation of a nonlinear "kink" in the light-current characteristic of a laser), etc.[37] The form of eq. (58) is still applicable with R_0 replaced by the parameter that is

correlated with lifetime. For components that fail suddenly[1] (no degrading observable) such other initial properties are, of course, the only predictors of individual-component life.

In principle, ρ is defined to take advantage of all observed correlations, e.g., $\rho(i_0, \partial i_0/\partial t, \tau)d\tau$ is conditional upon both the predeployment threshold current i_0 and the predeployment isothermal degradation rate $\partial i_0/\partial t$ of a laser. In some situations it is feasible to measure multiple correlations, as in Ref. 37. Often, however, limited data or small homogeneous populations do not permit the functional form of ρ nor the multiple correlation and cross-correlation parameters to be established with much confidence. As a compromise in such cases one may stay with eq. (58) while redefining R as any combination of i_0 and $\partial i_0/\partial t$ which maximizes the correlation. For example, the definition given by eq. (56) or the definition

$$I = \frac{\ln(i/i_0)}{\ln(i_f/i_0)} \tag{59}$$

may be superior to the definition $I = i$ or $I = i/i_0$ when maximizing the correlation between the lifetime τ and the initial value of the isothermal degradation rate $R = \partial I_r/\partial t_a$ or $\mathscr{R} = \partial I_a/\partial t_a$. (Equation [59] replaces eq. [56] when i increases approximately exponentially in time. Then $(\partial I_0/\partial t)^{-1}$ is still the initial estimate of component life.)

In a high-reliability application, where it is inefficient or even impossible to age beyond the effective system life, similar considerations apply. In a typical procedure, samples are aged through the system life, perhaps on an accelerated basis, and a best-fit analytical model is then sought for the functional form of $i(t)$. Because measurement error typically leads to low confidence when many parameters appear, a fairly simple function is normally chosen; for example, $i = be^{ct}$ with the parameters b and c fitted separately for each component. Usually no such simple function will precisely model the behavior of every, or even any, component, and thus extrapolations based on the model function (e.g., on $i = be^{ct}$) would be somewhat inaccurate even if there were no measurement errors.* It is convenient, although not necessary, to choose an I which increases linearly in time, e.g., eq. (59) in the case of $i = be^{ct}$. The behavior of I during the monitor period t_0 is then extrapolated (projected) through the effective system life (dashed line in Fig. 9). This is compared with the actual behavior

---

* Even if all devices degrade by a common mechanism (i.e., after perfect purging) with a common activation energy, the functional form of the isothermal aging curve, $I(t)$, could be very complicated and difficult to parameterize for the whole population. For example, if clusters of impurities or defects exist initially at different surface regions of a laser, $I(t)$ might be expected to exhibit a staircase character as the successive diffusion fronts from the clusters reach the active layer.
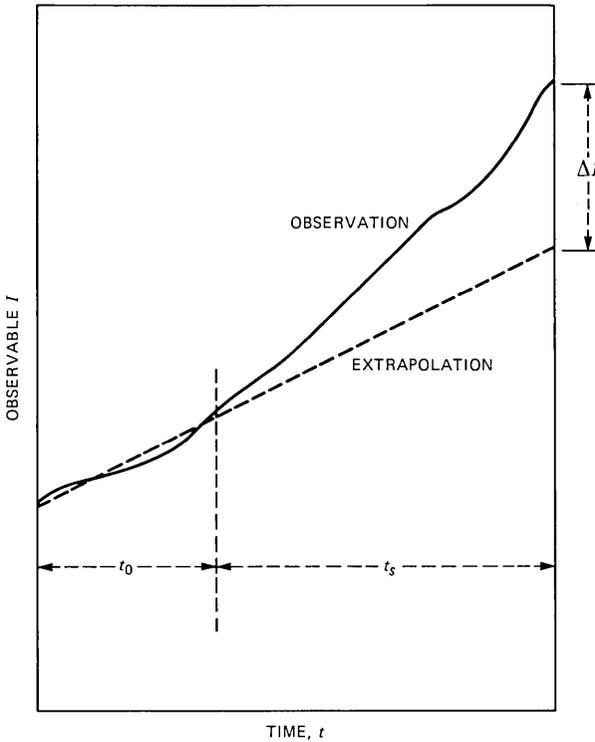
Fig. 9—After an observation or monitor period t_0, the observable I is extrapolated (dashed line) through the effective system life t_s. Then ΔI is the error between the subsequently observed value (solid line) and the extrapolated (projected) value.

through the system life t_s (solid line) to determine the error ΔI between the actual and the expected values of I. An analytical $\rho(R, I)$ or $\rho(R, I, t)$ is fitted to the ΔI data, where $\rho(R_0, I)dI$ is the probability that I will fall between I and $I + dI$ at the end of t_s given a mean rate R_0 during t_0. A scatter plot of the expected I versus the observed I is often useful for choosing the functional form of ρ. Then, in analogy to eq. (58), ρ is used to generate the distribution $F'(I)$ at the end of the system life. (I replaces τ in eq. [58].) In this case, however, ρ may include a finite fraction at $I = \infty$, (or, alternatively, the triangles are extrapolated to produce surrogate ✕'s for $t > t_s$, and an $F(\tau)$, valid primarily for $t \leq t_s$, is fitted to the ✕'s which fall on both sides of t_s). In contrast to familiar procedures for extrapolating parameterized functions (e.g., extrapolation of an elliptical orbit), the empirical distribution $\rho(R_0, I)$ has the important property that it accounts not only for the implications of measurement error but also for the fact that there may not exist any parameterized function $i(t)$ that is precisely applicable to all, or even to any, of the components in the

population. If the measurement error can be separately evaluated, its contribution can be subtracted from $\rho(R, I)$ to reveal that part of ρ which arises from the imperfect applicability of the empirical model function. This is of interest for such purposes as assessing the potential benefits of more accurate instrumentation. This projection procedure is further developed in Ref. 38. (Of course, in the fortunate case that a single parameterized function $i(t)$ accurately describes the entire population, then, as an alternate approach, one may use standard methods to estimate from the data the distributions of the parameters.)

XI. SUMMARY

A conceptual framework and quantitative procedures were presented for the accelerated aging and reliability predictions applicable to continuously degrading components.

XII. ACKNOWLEDGMENTS

We would like to acknowledge many stimulating discussions with E. I. Gordon, B. W. Hakki, J. H. Rowen, and M. Tortorella.

REFERENCES

1. A. S. Jordan, J. C. Irvin, and W. O. Schlosser, "A Large Scale Reliability Study of Burnout Failure in GaAs Power FETs," 18th Ann. Proc. Reliab. Phys. Symp. (April 1980), pp. 123–33.
2. A. T. English and C. M. Melliar-Smith, "Reliability and Failure Mechanisms of Electronic Materials," Ann. Rev. Mater. Sci., 8 (1978), pp. 459–95.
3. D. S. Peck and C. H. Zierdt, Jr., "The Reliability of Semiconductor Devices in the Bell System," Proc. IEEE, 62, No. 2 (February 1974), pp. 185–211.
4. E. I. Gordon, F. R. Nash, and R. L. Hartman, "Purging; A Reliability Assurance Technique for New Technology Semiconductor Devices," IEEE Electron. Lett., EDL-4, No. 12 (December 1983), pp. 465–6.
5. F. R. Nash et al., "Selection of a Laser Reliability Assurance Strategy for a Long-Life Application," AT&T Tech. J., this issue.
6. F. R. Nash et al., "Laser Reliability Assurance Strategies—Part II: Implementation of the Proposed Reliability Assurance for an InGaAsP/InP, Planar Mesa Buried Heterostructure Laser Operating at 1.3 μm for Use in a Submarine Cable," AT&T Tech. J., this issue.
7. W. B. Joyce, R. W. Dixon, and R. L. Hartman, "Statistical Characterization of the Lifetimes of Continuously Operated (Al,Ga)As Double-Heterostructure Lasers," Appl. Phys. Lett., 28, No. 11 (June 1976), pp. 684–6.
8. E. B. Fowlkes, "Some Methods for Studying the Mixtures of Two Normal (Lognormal) Distributions," J. Amer. Statist. Ass., 74, No. 367 (September 1979), pp. 561–75.
9. B. C. DeLoach, Jr., et al., "Degradation of CW GaAs Double-Heterojunction Lasers at 300K," Proc. IEEE, 61, No. 7 (July 1973), pp. 1042–4.
10. R. W. Dixon and R. L. Hartman, "Accelerated Aging and a Uniform Mode of Degradation in (Al,Ga)As Double-Heterostructure Lasers," J. Appl. Phys., 48, No. 8 (August 1977), pp. 3225–9.
11. A. T. English, K. L. Tai, and P. A. Turner, "Electromigration of Ti-Au Thin-Film Conductors at 180° C," J. Appl. Phys., 45, No. 9 (September 1974), pp. 3757–67.
12. B. W. Hakki, P. E. Fraley, and T. F. Eltringham, "1.3-μm Laser Reliability Determination for Submarine Cable Systems," AT&T Tech. J., this issue.
13. F. H. Reynolds, "Thermally Accelerated Aging of Semiconductor Components," Proc. IEEE, 62, No. 2 (February 1974), pp. 212–21.

14. H. Kressel, M. Ettenberg, and I. Ladany, "Accelerated Step-Temperature Aging of $Al_xGa_{1-x}As$ Heterojunction Laser Diodes," Appl. Phys. Lett., *32*, No. 5 (March 1978), pp. 305–8.

15. D. V. Lang, "Recombination-Enhanced Reactions in Semiconductors," Ann. Rev. Mater. Sci., *12* (1982), pp. 377–400.

16. N. Dutta, "Calculated Temperature Dependence of Threshold Current of GaAs-$Al_xGa_{1-x}As$ Double Heterostructure Lasers," J. Appl. Phys., *52*, No. 1 (January 1981), pp. 70–3.

17. H. Ishikawa, T. Fujiwara, K. Fujiwara, M. Morimoto, and M. Takusagawa, "Accelerated Aging Test of $Ga_{1-x}Al_xAs$ DH Lasers," J. Appl. Phys., *50*, No. 4 (April 1979), pp. 2518–22.

18. H. Temkin, C. L. Zipfel, and V. G. Keramidas, "High Temperature Degradation of InGaAsP/InP Light Emitting Diodes," J. Appl. Phys., *52*, No. 8 (August 1981), pp. 5377–80.

19. T. Ikegami et al., "Stress Tests on 1.3 μm Buried-Heterostructure Laser Diode," Electron. Lett., *19*, No. 8 (April 1983), pp. 282–3.

20. B. W. Hakki and F. R. Nash, "Catastrophic Failure in GaAs Double-Heterostructure Injection Lasers," J. Appl. Phys., *45*, No. 9 (September 1974), pp. 3907–12.

21. C. L. Zipfel et al., "Competing Processes in Long Term Accelerated Aging of Double Heterostructure $Ga_{1-x}Al_xAs$ Light Emitting Diodes," J. Appl. Phys., *53*, No. 3 (March 1982), pp. 1781–6.

22. W. B. Joyce and B. C. DeLoach, "Alignment of Gaussian Beams," Appl. Opt., *23* (December 1, 1984), pp. 4187–96.

23. R. T. Ku and A. E. Bakanowski, unpublished work.

24. M. Stitch, G. M. Johnson, B. P. Kirk, and B. J. Brauer, "Microcircuit Accelerated Testing Using High Temperature Operating Tests," IEEE Trans. Reliab., *R-24*, No. 4 (October 1975), pp. 238–50.

25. K. Mizuishi et al., "Acceleration of the Gradual Degradation in (GaAl)As Double-Heterostructure Lasers as an Exponent of the Value of the Driving Current," J. Appl. Phys., *50*, No. 11 (November 1979), pp. 6668–74.

26. K. Endo et al., "Rapid Degradation of InGaAsP/InP Double Heterostructure Lasers Due to ⟨110⟩ Dark Line Defect Formation," Appl. Phys. Lett., *40*, No. 11 (June 1982) pp. 921–3.

27. R. L. Hartman and R. W. Dixon, "Reliability of DH GaAs Lasers at Elevated Temperatures," Appl. Phys. Lett., *26*, No. 5 (March 1975), pp. 239–42.

28. P. D. Wright, W. B. Joyce, and D. C. Craft, "Electrical Derivative Characteristics of InGaAsP Buried Heterostructure Lasers," J. Appl. Phys., *53*, No. 3 (March 1982), pp. 1364–72.

29. W. B. Joyce and R. W. Dixon, "Thermal Resistance of Heterostructure Lasers," J. Appl. Phys., *46*, No. 2 (February 1975), pp. 855–62, Fig. 5.

30. M. S. Green, "Markoff Random Processes and the Statistical Mechanics of Time-Dependent Phenomena," J. Chem. Phys., *20*, No. 8, (August 1952), pp. 1281–95.

31. S. S. Cheng, "Optimal Replacement Rate of Devices With Lognormal Failure Distributions," IEEE Trans. Reliab., *R-26*, No. 3 (August 1977), pp. 174–8.

32. A. S. Jordan, "A Comprehensive Review of the Lognormal Distribution With Application to LED Reliability," Microelectron. Reliab., *18*, No. 3 (1978), pp. 267–9.

33. A. S. Jordan, "Confidence Limits on the Failure Rate and a Rapid Projection Nomogram for the Lognormal Distribution," Microelectron. Reliab., *24*, No. 1 (1984), pp. 101–24.

34. N. L. Johnson and S. Kotz, *Continuous Univariate Distributions*, Vol. 1, Boston: Houghton Mifflin, 1970.

35. E. Kinsbron, C. M. Melliar-Smith, and A. T. English, "Failure of Small Thin-Film Conductors Due to High Current-Density Pulses," IEEE Trans. Electron. Dev., *ED-26* (January 1979), pp. 22–6.

36. D. A. Ackerman, R. M. Camarda, and R. L. Hartman, "Latent Kink Prediction via Nondestructive Stripe Width Determination," unpublished work.

37. K. Mizuishi, "Statistical Analysis of Aging-Induced Degradation (or Lifetime) Variations in (Al,Ga)As/GaAs Double-Heterostructure Lasers," IEEE J. Quantum Electron., *QE-19* (March 1983), pp. 457–65.

38. A. R. Eckler, "A Statistical Approach to Laser Certification," AT&T Tech. J., this issue.

AUTHORS

Peter R. Bossard, Ph.D. (Physics), 1979, Temple University; Bell Laboratories, 1979–83. Mr. Bossard was a member of the technical staff in the Device Reliability Group. During this time he worked on various experimental and theoretical projects designed to increase the reliability of solid state devices. In February 1984 he left Bell Communications Research, Inc. to become president of Voyager Technologies, Inc., a company that designs and manufactures instrumentation and distributed computing systems for real-time quality control in the semiconductor manufacturing environment.

Robert L. Hartman, B.A. (Mathematics), 1958, Columbia College; B.S. (Engineering Physics), M.S. and Ph.D. (Physics), Columbia University, in 1959, 1961, and 1968, respectively; AT&T Bell Laboratories, 1968—. At AT&T Bell Laboratories, Mr. Hartman is presently engaged in studying the physics of III-V compound semiconductor devices, particularly InGaAsP/InP and (Al,Ga)As-GaAs double-heterostructure injection lasers. He has contributed extensively to the understanding of the degradation mechanisms in these devices and authored or coauthored several key papers in this field, especially those involving the discovery and analysis of dark-line defects, the discovery of strain effects upon laser reliability, the attainment of long operating lifetimes in LEDs and injection lasers, the first measurement of the thermal acceleration of injection laser operating lifetime, and effects of surfaces on LED and laser reliability. In 1980 Mr. Hartman was appointed Supervisor of the Semiconductor Lightwave Lasers Group. Member, Sigma Xi, American Physical Society.

William B. Joyce, B.E.P. (Engineering Physics), 1955, Cornell University; Ph.D. (Theoretical Physics), 1966, Ohio State University; AT&T Bell Laboratories, 1966—. Mr. Joyce publishes primarily in the areas of architectural acoustics, classical mechanics and optics, history of science, semiconductor optical devices, research and development organization, and reliability. Member, American Physical Society.

Franklin R. Nash, B.S. (Physics), 1955, Polytechnic Institute of New York; Ph.D. (Physics), 1962, Columbia University; AT&T Bell Laboratories, 1963—. At AT&T Bell Laboratories, Mr. Nash has been involved in research on gas lasers (modulation and mode locking), detection of subnanosecond pulses using photodiodes, nonlinear optical materials, parametic oscillators, neoclassical theory of radiation, and semiconductor lasers (mode guidance, material defects, electro-optic anomalies, and degradation mechanisms). Currently, he is engaged in the reliability assessment of long-wavelength semiconductor lasers. Member, American Physical Society, IEEE.

A Statistical Approach to Laser Certification

By A. R. ECKLER*

(Manuscript received April 2, 1984)

This paper proposes a statistical methodology for certifying the longevity of individual lasers to be used in an undersea communications cable. In general, the strategy is to extrapolate the slope of a degradation parameter such as i (the current necessary to maintain a given level of optical-output power) when the laser is subjected to elevated temperatures. This extrapolation is made from time t_1 (when the test terminates) to time t_p (the system life, as contracted in time by an Arrhenius relationship). An important part of the process is the calculation of the variability of the estimated current value at time t_p, used to estimate the probability that it will exceed a critical failure level i_f; lasers that exceed a predetermined probability will be rejected. The methodology presented here must, in the absence of actual data on lasers to be used in an undersea communications cable, be tentative with respect to details; modifications may well be necessary to accommodate idiosyncrasies of the manufacturing process and the testing equipment.

I. INTRODUCTION

In fabricating a system such as an undersea communication cable, where the cost of replacing a failed laser package (or other gradually degrading component) is very great, it is not enough to use lasers from a production run that have normal operating characteristics. Rather, each laser's longevity must be certified by placing it on an accelerated life test for a time t_p and then using these data (operating characteristics at a sequence of observation times) to estimate its longevity. This laser is used in the cable only if its estimated longevity exceeds

---

* AT&T Bell Laboratories.

a certain criterion (expressed as a probability that a degradation parameter i, such as the current necessary to maintain 5 mW of optical-output power, will not exceed some failure value i_f before the end of the planned system life).

To simplify the formalism, we choose a degradation parameter I that has, to the extent possible, the following two properties:[1] (1) I should increase linearly in time, (2) the ranking of the lasers in order of increasing initial degradation rate dI/dt should also be the ranking in order of decreasing estimated lifetime. For example, suppose that i typically increases approximately exponentially in time. Let i_f be the failure value of i, which may well vary from laser to laser. (For a hypothetical example, a laser with a high initial value i_o fails when i reaches the maximum output of the power supply i_m, while a laser with an i_o less than $(1/2)\ i_m$ fails at $i_f = 2i_o$ due to the limitation of a compensating circuit.) Then an appropriate choice for the reduced current I is[1]

$$I = \frac{\ln(i/i_o)}{\ln(i_f/i_o)},\tag{1}$$

and every laser current degrades from its initial value $I = 0$ to its failure value $I_f = 1$.

To obtain a significant amount of degradation in the time span available for certification, the aging process is speeded up by operation at an elevated temperature T_1 rather than by operation at the system temperature T_s. If aging for a time t_1 at T_1 causes the same degradation as aging for a time t_s at T_s, and if the Arrhenius relation is applicable, then

$$t_1/t_s = e^{(T_1^{-1} - T_s^{-1})E/k},\tag{2}$$

where E is a parameter called the activation energy and K is Boltzmann's constant. In the following, two cases are considered according to whether E varies significantly from laser to laser (as determined from step stressing), or whether a common value for E may be used for all lasers.

Let $t_1 - t_p$ be the planned system life, as contracted by temperature [eq. (2)] at T_1. The predicted behavior of the purged[2,3] and stabilized[3] lasers vying for incorporation in the cable is based on a linear extrapolation of the data collected during time t_p at T_1. To establish a correction factor to the predicted behavior, one places a number (m) of expendable calibration lasers (not to be used in the cable) from the same population on accelerated life test for a time t_1 to establish a correction factor for the extrapolation of the vying lasers' behavior. In the following, we consider in detail the evaluation of this correction factor.

II. THE STANDARD MODEL

It is assumed that measurements are taken of I at various times between $t = 0$ and $t = t_p$ for the n lasers vying for certification, and between $t = 0$ and $t = t_1$ for the m calibration lasers; a sequence of times, t, $t + \Delta t$, $t + 2\Delta t$, $\cdots$, is the easiest to analyze. Although the lasers are aged at T_1, the measured values of I are obtained at T_s by momentarily cooling the lasers to T_s for each measurement. Alternately, the values of I at T_s may be inferred from measurements of i at T_1 if there is a known algorithm for converting these i values to their corresponding values at T_s. (For example, for certain types of lasers it has been shown that the temperature dependence of i is well modeled by the empirical relationship $i \approx \exp(T/T_o)$ over a usefully wide temperature range, where T_o is a parameter characterizing the temperature dependence.[3])

The certification of an individual laser is carried out as follows. Let I denote the estimate of the current at time t_s if the laser has been continuously operated at temperature T_s. It is estimated by the certification equation

$$I = I_o + t_1 b_{T_1} + K, \qquad (3)$$

where t_1 is a constant derived from the assumed value of E, I_o is the value of the current measured at time $t = 0$ and temperature T_s, b_{T_1} is the estimated slope of a least-squares straight-line fit to the certification measurements on the laser (from $t = 0$ to $t = t_p$), and K is a correction factor obtained from the calibration lasers that have been operated at temperature T_1 up to time t_1 (see Fig. 1). In the discussion below, it is assumed that I_o, b_{T_1}, and K are uncorrelated random variables. The validity of this assumption can be tested by looking at the three-dimensional scatter plot of these variables for the set of calibration lasers. If correlation exists, it may be necessary to rewrite the equation above in a form in which the variables are uncorrelated, such as $I = I_o + b_{T_1}(t_1 + K)$ or $I = I_o[1 + b_{T_1}(t_1 + K)]$.

To estimate K, (1) calculate for each calibration laser the quantity d_i, the difference between the actual current I_t measured at time t_1, and an extrapolated current I_e determined by a least-squares straight-line fit to the calibration measurements on the laser from $t = 0$ to $t = t_p$ only; and (2) calculate $K = \sum d_i / m$ over all the m calibration lasers.

Not all of the calibration lasers will survive until time t_1; some will catastrophically fail at earlier times. In the case of eq. (1), *catastrophic failure* means that no I in the experimentally accessible range $0 \leqslant I \leqslant I_m > 1$ will yield a 5 mW output. These times should be recorded, as it will be necessary to use them to calculate the probability of catastrophic failure as a function of the variability of E. (This variability
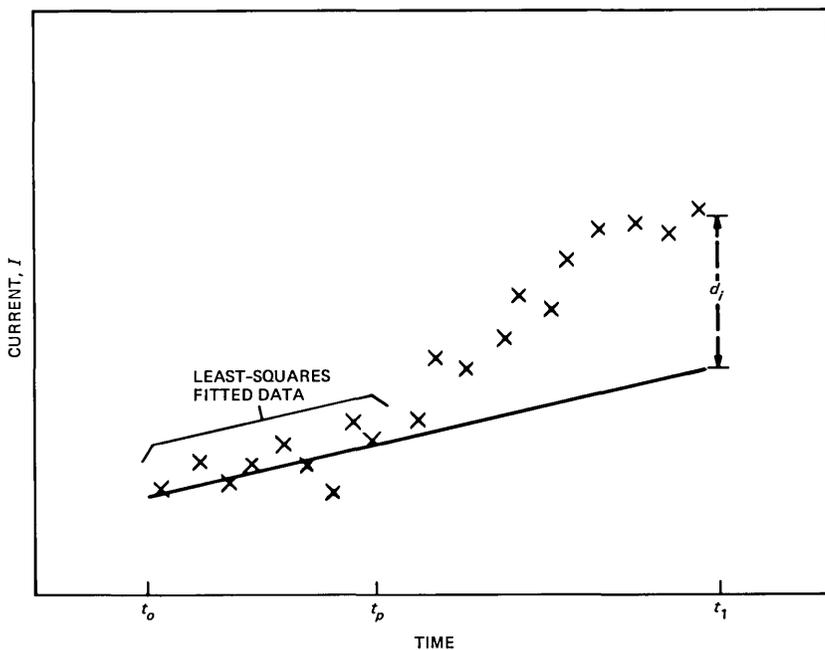
Fig. 1—Determination of correction factor d_i for an individual laser (run at elevated temperature T_1).

may be caused either by measurement error, or by inherent differences in E from laser to laser.)

Having removed the catastrophic failures, it is now necessary to ascertain the variability of the value, I, calculated for a laser surviving to time t_1. One cannot necessarily assume that the Central Limit Theorem will apply, leading to a Gaussian distribution of I (and, therefore, requiring only that the mean and variance of I be known). Instead, it will be necessary to use the calibration data to actually determine the distribution function followed by I, taken as a function of the underlying b_{T_1} (rate of degradation) of a laser, the variability of E (if it is significant), and the variability of the estimated b_{T_1} (related to the number of observations of the current, the times of these observations, and the residual variability of a single observation). Once the distribution function of I is known, then it is easy to calculate the probability for each certification laser that its current at time t_1 is greater than the critical level I_f, given that I has been calculated from the certification equation.

Let us now look at the method of calculating a distribution function for I in more detail, on a term-by-term basis. It is impossible to calculate this distribution function analytically in the general case with E variable, and even in the unlikely case that E is known exactly,

it is possible only if the distribution of the d_i and the residual error of a single observation of current are both Gaussian. Instead, the distribution function must be simulated, and an empirical multiparameter distribution (say, a gamma) fitted to the simulated data. Specifically, we assume that the distribution function of E is known (or, perhaps, approximated by a Gaussian distribution with known mean and variance). The distribution of b_{T_1} is uniquely determined by the expression

$$\sum (x_i - \bar{x})y_i / \sum (x_i - \bar{x})^2, \tag{4}$$

where $x_1, \cdots, x_n$ are the times of observation between t_o and t_p, and $y_1, y_2 \cdots y_n$ are the corresponding current measurements. If the distribution of the measurement errors of the y_i are known, then the distribution of b_{T_1} is uniquely specified; for example, if the y_i measurement errors are independent observations from a common Gaussian population, then b_{T_1} also has a Gaussian distribution with a certain mean and variance. In fact, as long as n is not too small (say, $n \geqslant 10$), b_{T_1} will approach a Gaussian distribution even if the individual measurement errors are independent random variables from a non-Gaussian population. The essential point is that the distribution of b_{T_1} is specifiable as soon as the mean and variance of the measurement errors are given.

Putting all this together, one draws a random variable from the E distribution, converts it to a value of t_1 by the acceleration equation given earlier, and multiplies this by a second random variable, which is the sum of (an assumed) b_{T_1} and a value drawn from a Gaussian distribution with a known variance. One then adds to this a third random variable, drawn from a distribution function that has been previously fitted to the d_i; the sum of these yields a single value of the distribution of $I - I_o$ (I_o is assumed to be known, or with a very small error, but if necessary this could be folded in as well). It is worth noting in passing that the distribution function of the d_i may depend upon the value of the slope b_{T_1}; if so, a family of d_i distributions corresponding to different slopes must be inferred from the calibration data. In practice, this family might be approximated by a simple empirical b_{T_1}-dependent scaling of the d distribution. For example, the empirical distribution might turn out to be only weakly correlated, if at all, with b_{T_1} or with t_1.

One repeats this sampling experiment many times (say, 100 or more) to build up enough data to make a distributional fit to the simulation data. Note that it is necessary to do this for various assumed values of three parameters—the variance of b_{T_1}, b_{T_1} itself, and the variance of E—obtaining a three-dimensional family of $I - I_o$ distributions. The appropriate member of the family is then used when

certifying each laser (which will have its own unique b_{T_1} and, possibly variance of b_{T_1}).

From the proper distribution function of $I - I_o$, the probability that the certified laser will have an I exceeding I_f is at once calculated. To this, of course, must be added the probability of catastrophic failure, obtained from an integration of the calibration laser failure data taken over the variability of t_1 (determined from the variability of E, as indicated above). If necessary, this probability can be readily obtained by another simulation procedure.

In setting up the calibration studies, note that one must run the lasers for a time t_1 corresponding to the upper 5-percent (say) value of E.

III. ACKNOWLEDGMENT

REFERENCES

1. W. B. Joyce et al., "Methodology of Accelerated Aging," AT&T Tech. J., this issue.
2. F. R. Nash et al., "Selection of a Laser Reliability Assurance Strategy for a Long-Life Application," AT&T Tech. J., this issue.
3. F. R. Nash et al., "Implementation of the Proposed Reliability Strategy for an InGaAsP/InP, Planar Mesa, Buried Heterostructure Laser Operating at 1.3 μm for Use in a Submarine Cable," AT&T Tech. J., this issue.

AUTHOR

A. Ross Eckler, B.A. (Mathematics), 1950, Swarthmore; Ph.D. (Statistics), 1954, Princeton University; AT&T Bell Laboratories, 1954–85. At AT&T Bell Laboratories Mr. Eckler was engaged in statistical consulting for military systems until 1971, and telephone systems thereafter. He was Head of the Electronics Technology Assessment Department. He coauthored with S. A. Burr the text *Mathematical Models of Target Coverage and Missile Allocation* (1972). Member, Sigma Xi, American Statistical Association.

1.3-μm Laser Reliability Determination for Submarine Cable Systems

By B. W. HAKKI, P. E. FRALEY, and T. F. ELTRINGHAM*

(Manuscript received June 25, 1984)

To meet the stringent requirements of a submarine cable system, our 1.3-μm laser prequalification program has two objectives—first, to define the testing methodology that will accurately evaluate the potential reliability of the laser; and second, to obtain a preliminary indication of laser reliability on which the system configuration can be designed. Our testing methodology involves a combination of step-temperature, step-power, and isothermal testing over the temperature interval between 10 and 80°C and power levels of 1 to 5 mW per facet. Our results show that the long-term degradation process is thermally accelerated, with a median activation energy of 1 eV and a standard deviation of 0.13 eV. By using these activation energies, in conjunction with our measurements of degradation rates, we can project laser performance to 10°C, i.e., system operating temperature. It is estimated that the median time to failure for "light bulb" operation at 10°C is over 2×10^7 hours; and with 98-percent probability it is greater than 5×10^6 hours. Hence, when viewed strictly in terms of light bulb sources of stimulated power, these 1.3-μm lasers have adequate life. In addition, other potential operational malfunctions are being investigated, and they do not seem to change our basic conclusion about the usefulness of these 1.3-μm lasers for submarine cable application.

I. INTRODUCTION

The submarine cable system requires a unique approach to reliability. Each component is subjected to an investigation that (preferably) uncovers the basic failure mechanism(s) and certainly quantifies its

---

projected failure probability over the system life. The higher the expected failure rate of a type of component, the greater is the need to accurately predict individual component lives. The reasons are obviously economic: A high failure rate may require a system configuration that utilizes redundancy and increases the initial cost. Furthermore, accurate projections of failures affect system planning to avoid costly undersea repairs. Therefore, the tie between reliability and cost manifests itself in the submarine cable system as in no terrestrial system.

The Buried-Heterostructure (BH), InGaAsP, 1.3-μm laser* considered for use in the optical submarine cable system is the product of a new technology.[1-3] Hence, the primary objective of this prequalification program is to make a preliminary determination of the extent to which the reliability of this 1.3-μm laser satisfies the stringent requirements of the submarine cable system. The program concentrates on the projection of 1.3-μm laser performance to 10°C, which is the system operating temperature. To accomplish this, a significant portion of our effort is devoted to the following: (1) the determination of activation energy(s) that governs the thermally accelerated mechanism(s); (2) the aging law that describes the time evolution of current at a specific temperature; (3) the projection of performance to the system operating temperature of 10°C; and, concurrently, (4) the identification of the dominant failure mechanism(s) that will determine the life of the laser.

Our laser prequalification program is diversified in such a manner as to provide extensive information about the various facets of laser performance. Thus, we use two generically different testing methodologies: step stress and isothermal.[4] The two methods provide supportive, complementary, and corroborative information that is essential to a better understanding of laser reliability. For example, isothermal tests provide useful information about the time dependence of the aging law, end-of-life criteria, activation energy, and data that validate the assumptions made in step-stress analysis. On the other hand, step-stress tests provide an efficient measure (in time and number of devices) of the activation energy, degradation rates, and the dependence of these rates on optical power. These two programs will be described in detail, and test results will be given on 262 unpurged[5] lasers, wherein the accumulated time (as of February 1984) is ~2.4 × 10^6 device hours. (The lasers were comprised of early vintage [1980-2] devices that had seen two types of burn-in screen. The

---

* The laser chip investigated is made by Hitachi Ltd. and meets AT&T Bell Laboratories specifications. The buried active region is nominally 1 to 1.7 μm wide, 0.1 to 0.3 μm thick, and 300 μm long. The laser is subsequently packaged within AT&T.

majority of lasers were burned in at 50°C/5 mW/100h, at constant current, and met the condition that the power change was less than 10 percent. The rest of the devices were burned in at 60°C/5 mW/ 100h, at constant power, and satisfied the condition that the drive current increased by less than 5 percent.) The projections given here are based on the assumption that there will be no unexpected device failures, such as those due to weakly temperature-activated mechanisms. Data needed to estimate the sudden failure rate of the lasers are also collected in this prequalification program. A separate purge program is expected to eliminate such devices.[5]

This paper is organized as follows:

II. 1.3-μm LASER RELIABILITY

2.1 Failure modes

A laser can fail either as a "light bulb" or operationally in a transmitter. A light bulb failure is defined as the inability of the laser to deliver the requisite amount of stimulated power (e.g., 5 mW/facet) within the constraints of the current available for the transmitter. On the other hand, there are more stringent operational requirements imposed by the transmitter. These pertain to the ability of the laser to provide the quality of optical signal that is required to maintain an error rate below a certain value (e.g., 10^{-9} per bit).[6] Light bulb failure, due to an increase in the drive current at a fixed optical power, can be due to an increase in threshold of the active region, a decrease in the differential quantum efficiency, an increase in extraneous leakage currents,[5] or change in the front/back power ratio. All these parameters are measured, as described in detail in Appendix A. Most of the

present published effort on lasers concentrates on the increase in drive current and its acceleration by temperature, i.e., its activation energy. For reasons to be discussed later, we will present in this paper primarily our results on the degradation of the threshold current. However, our ongoing program is investigating laser operational failures caused by the excitation of high-order transverse modes, spectral splitting,[6] modal partition noise, change in extinction ratio, occurrence of self-induced pulsations, or mechanical instability in the package (see also Appendix A). Except for the degradation of threshold, we presently do not know whether the rest of the potential failure mechanisms are thermally activated. This important question is being addressed as part of the continuing reliability program.

2.2 Testing methodology

There are many different approaches to the prediction of lifetime at operating temperature. These have been described to some extent in the literature.[7] Each method has its own advantages and disadvantages as far as our 1.3-μm laser program is concerned. For our purpose, we have chosen a combination of isothermal and step-stressing methods, which seem to be best suited for our program.

2.2.1 Step-stress testing

Step-stress testing is an efficient method of measuring a fundamental mechanism by holding most parameters constant, but varying one, e.g., temperature or optical power. Since the test is done on each device, it can, in principle, accurately estimate the dependence of the degradation on the parameter(s) that is varied. However, when this methodology is applied to temperature and power-step-stressed lasers, several issues must be resolved in order to achieve a high confidence level in the results. These step-stressed issues are the following: (1) the amount of bias (if any) in the derived quantity, such as activation energy, due to the particular choice of a temperature sequence (temperature reciprocity); (2) the applicability of the observed power-step-stress results obtained at some temperature, to other temperatures; and (3) whether the results obtained at each step-stress test are representative of the dominant long-term degradation process.

At the present stage of the program, we have preliminary answers to the above-mentioned questions. However, the issues raised have, to a large extent, helped in designing the step-stress experiments so that by the conclusion of the qualification program, the majority, if not all of these questions, will be answered.

A total of 117 lasers (see Appendix B) were stepped in temperature and power. The temperature was stepped in the range between 50 and

80°C, and the power was stepped between 1 and 5 mW/facet. By comparing the degradation rates of threshold current at different temperatures, but at constant power, a measure of activation energy for this parameter was obtained. Similarly, by holding the temperature constant, but changing the optical power, we obtained the dependence of the degradation rate on the level of operation. Our step-stressing methodology subscribed to the following guidelines:

1. The degradation rate of threshold at constant power level was determined at three different temperatures, T_1, T_2, and T_3.

2. The degradation rate was determined at constant temperature, but at several power levels, i.e., 1, 3, and 5 mW/facet.

3. At each test condition, the average increase in drive current was kept within the range of a minimum of ≈5 percent and a maximum of ≈25 percent, at the end of the first test interval (15-percent maximum for subsequent intervals).

4. In each group of lasers being tested, a minimum of five lasers from a particular slice was used, if possible, as indicated in Appendix B.

5. A laser was considered (arbitrarily) to have reached the limit of its light bulb life when the threshold current increased by 100 percent.

The first guideline provided information needed to derive the activation energy(s). At face value, it might seem simple and straightforward. However, to make a valid extrapolation from (T_1, T_2, T_3) to the operating temperature at 10°C, certain requirements must be satisfied. These are (1) a single activation energy for threshold current degradation applies for all four temperatures: T_1, T_2, T_3, and 10°C; (2) the dominant aging mechanism is an extrapolatable function of time at all temperatures (predominantly linear behavior is observed); (3) the results are independent of the chosen sequence of temperature (temperature reciprocity); and (4) the same degradation mechanisms exist at all temperatures but at different rates. All these requirements are currently under investigation, and isothermal testing should provide useful and corroborative information.

The second guideline was motivated by the fact that since some of these BH lasers, for a number of reasons, could not deliver 5 mW at 80°C, we had to lower the power level to 3 mW. The difference in aging rates at constant temperature, but at different power levels, was determined in order to evaluate the impact of the level of operation on the degradation rate, as well as to identify the primary driving mechanism that controls degradation.

In the third guideline, the maximum allowed change was imposed to retain enough life in a laser to undergo subsequent testing at different temperatures or powers. The minimum change of 5 percent was needed to produce changes in the laser that were physically

significant, amenable to analysis, and to ascertain, to the extent possible, that a linear degradation rate was in fact applicable.

The fourth guideline was required for statistical significance. The wide variations observed in the values of degradation rates among various slices prior to purging preclude accurate analysis when performed on randomly selected lasers, at least within the sample sizes of 20 or 40 per test. Instead, the experiments were designed to distribute lasers from the same slice over the various test conditions to better account for slice-to-slice variability. Finally, the last guideline was somewhat arbitrary. It is conceivable that lasers may increase in threshold current by a factor greater than 100 percent before functionally failing. A more comprehensive determination of the end-of-life condition is being made in the context of operational failures by measuring the error rates of aged lasers operating in transmitters.[6]

Figure 1 shows schematically the temperature and power step stressing that two groups of devices, 20 lasers each, underwent. Group 1, detailed in Appendix B, was initially power step stressed by operating at 5 and 3 mW, consecutively. This was followed by the temperature step stressing indicated in Fig. 1. Group 2 (see also Appendix B) was
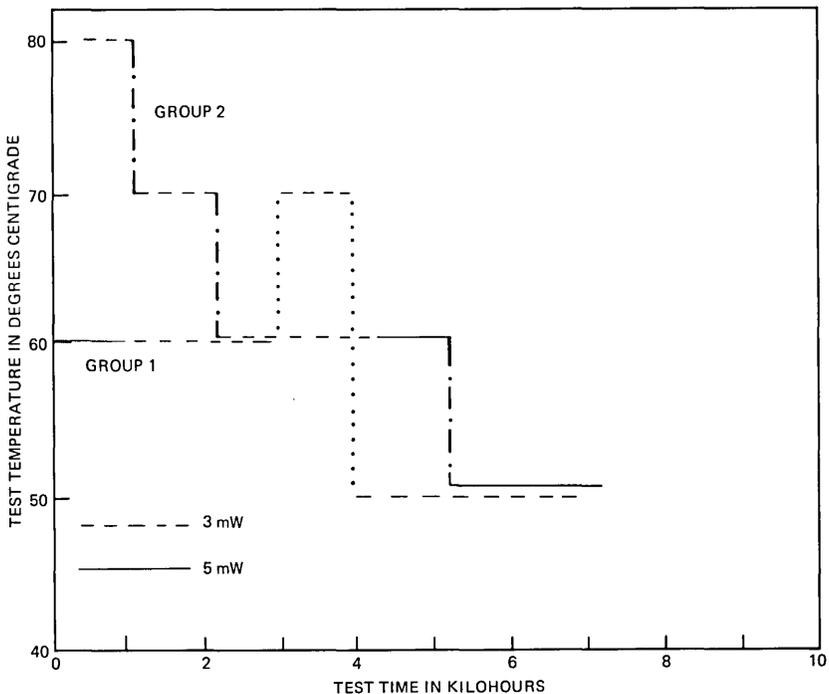
Fig. 1—Sequence of step stressing Groups 1 and 2 in temperature and power. The actual testing is continued beyond 8000 hours (not shown).
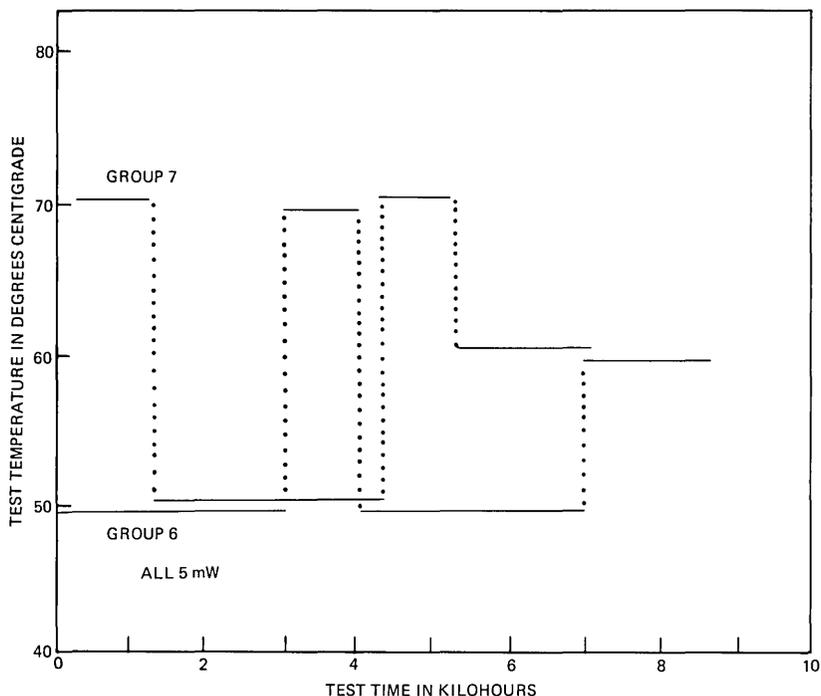
Fig. 2—Sequence of step-temperature testing Groups 6 and 7 at 5 mW per facet.

temperature step stressed by lowering the temperature from 80°C, in 10°C increments, keeping the optical power constant at 3 mW/facet. At 60°C, Group 2 was power step stressed by increasing the power from 3 to 5 mW. The rest of the sequence in Fig. 1 is self-explanatory.

The test sequence of Groups 1 and 2 provided information about activation energy and degradation rate dependence on optical power, in a relatively short period of time. However, the experiment did not provide information about temperature reciprocity, other memory effects, bias in data, etc. For this purpose, two additional groups of devices, 6 and 7, were step stressed in temperature, as shown schematically in Fig. 2. These two groups were selected and tested with the following considerations in mind: (1) The two populations were "alike" in the sense that the lasers used came from the same slices of material, to the extent possible (see Appendix B). This resulted in 37 and 40 lasers used in Groups 6 and 7, respectively. (2) The temperature sequences were mirror images of each other, in order to check temperature reciprocity. (3) Three temperatures were used for aging, in an attempt to confirm an Arrhenius relation for the degradation rates. (4) The power was held constant at 5 mW (system operating level) throughout the test, in order to reduce the variance in the results

associated with the dependence of the degradation rate of threshold current on the optical power level.

2.2.2 Isothermal testing

Isothermal testing of two or more groups operating at different temperatures is the more conventional approach to reliability study.[7] Normally, comparing the time-to-failure distribution curves at different temperatures gives the activation energy for the degradation process.[8] To get an accurate estimate of the activation energy by this method, it is necessary to satisfy two requirements: (1) the device population is homogeneous, i.e., device-to-device variability is smaller than the acceleration associated with temperature; and (2) there is adequate testing time available at the respective temperatures to establish the true nature of the failure distribution curves.

Both of the conditions stated above were difficult to fulfill in the case of our 1.3-μm lasers. In the first place, for any sample size of 40 or more unpurged lasers, obtained from several slices, it was not uncommon to see up to three orders of magnitude variation in rates of degradation. This exceeded the expected difference in temperature acceleration, for practical operating temperatures in the range of 40 to 70°C. Secondly, for practical aging temperatures, the failure rate was predicted to be low enough to require an unrealistic length of testing time to establish a failure time distribution curve. This second drawback was circumvented by using isothermal distribution curves of degradation rates instead of failure rates.[3,9,10] This, in turn, raised another problem, namely, the relationship between degradation rate and end-of-life for a device.

Given the respective advantages and disadvantages of step-stress and isothermal testing, we concluded that both methodologies are essential for a well-balanced program. Therefore, we dedicated a total of 145 lasers for isothermal testing at 10, 30, 40, and 60°C, in groups of 40, 25, 40, and 40 lasers, respectively, as described in Appendix B. To the extent possible, we attempted to design the test groups to be alike. That is, we selected slices that provide many lasers and then distributed the lasers from each slice among the various groups to be tested at different temperatures. This minimized the effects of slice-to-slice variability in degradation rate and enhanced our ability to derive an activation energy from the isothermal data. Hence, at a minimum, we expect this isothermal testing program to provide information about the aging law, correlation between degradation rate and time-to-failure, and the evolution of mechanisms that may lead to functional failures.

2.3 Aging law

The aging law governs the time evolution of device current at various

temperatures. When the lasers are operated at system power level (e.g., 5 mW), it is necessary to distinguish between changes in threshold current and those associated with the modulation current (the latter being the difference between threshold and the current at 5 mW per facet). This is done because there is no a priori reason to expect that the same degradation mechanisms apply equally to both current components. In fact, test results suggest that the aging mechanism for threshold may in some cases be different from that for the modulation current. Hence, as a first step in this analysis, the aging law will be obtained for the threshold current.

Our study of 1.3-μm BH lasers has shown that the threshold current shows the following behavior: (1) at constant temperature (isothermal) the current increases with time; (2) at a specific time (isochronal) the threshold increases with temperature; and (3) at a given temperature and time, a temporary disturbance in the operating condition (a momentary change in operating temperature or lasing condition) results in a transient change in the current-time response. These characteristics differ in their strengths and severity from laser to laser. However, their existence requires that both the experiments and the interpretation of the results account for these phenomena, to the extent possible.

An example of the time dependence of the threshold current of a particular laser is shown in Fig. 3a. The device was operated at 60°C and 5 mW/per facet. The threshold current in Fig. 3a was normalized with respect to its initial value. There are several noteworthy features in the data of Fig. 3a. For instance, the current initially increased at a rapid rate during the first 700 hours. This rapid increase ceased beyond 1000 hours; thereafter the current assumed a more moderate rate of increase. The rapid decrease in aging rate at 1000 hours forms what we call the "knee." At 1800 hours, the testing at 60°C was interrupted to measure the laser characteristics at 10°C. After testing was resumed at 60°C, the current beyond 1800 hours showed a small transient. In this particular device the transient was quite moderate in size. Similarly, at 5500 hours the testing was interrupted but the device stayed at 60°C. This again resulted in a small (almost insignificant) transient. Further test data up to 13,100 hours, of the laser shown in Fig. 3, indicate a continuation of the linear "postknee" linear aging behavior.

Other electro-optical measurements shed some light on the effects associated with the knee, also shown in Fig. 3b. In the first place, the magnitude of the saturable current at the knee seemed to be slice dependent. In the second place, the modulation current, as a function of time, often exhibited a knee similar to that associated with the threshold current. Finally, the junction voltage at threshold also
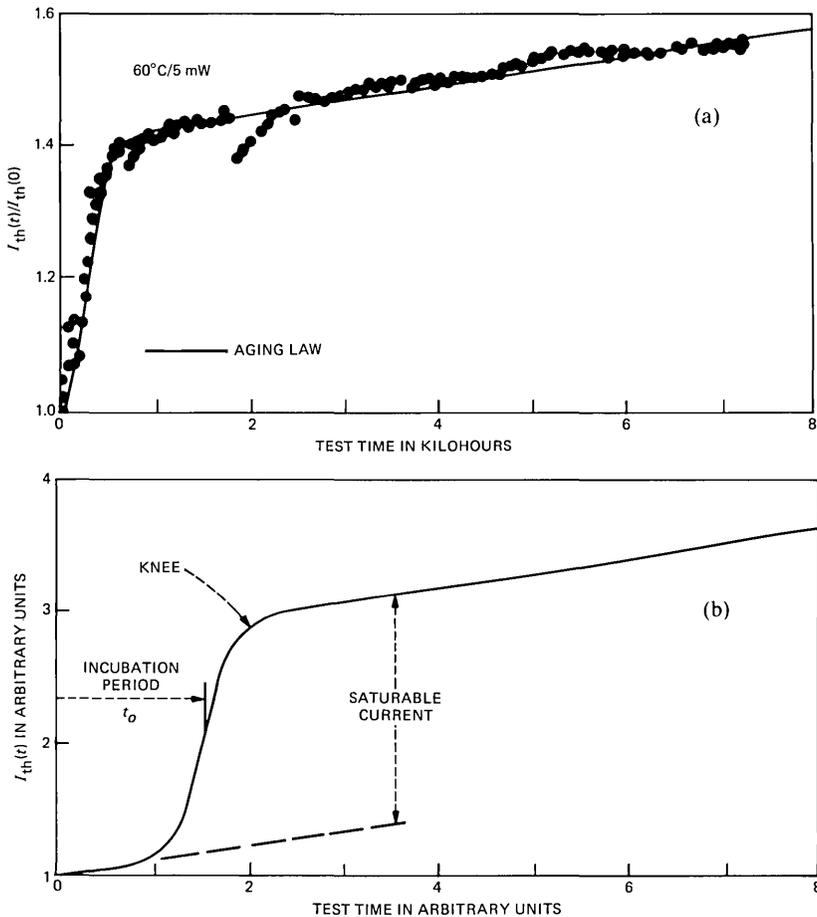
Fig. 3—(a) An example of the change in threshold with time taking place in a laser stressed at 60°C and 5 mW. (b) Schematic representation of the aging behavior of threshold current.

showed a small initial increase (a few millivolts), which saturated beyond the knee. (The junction voltage at threshold is obtained by subtracting from the operating voltage the measured product of threshold current and device resistance.) Hence, the change in threshold current prior to the knee was not an isolated effect.

It is clear that the degradation mechanism(s) in the laser is in fact very complicated. We are not, at the present time, in a position to make unequivocal statements about the fundamental phenomena that are influencing the laser behavior. It is possible, however, to glean from the data information that is suggestive of the basic mechanisms that are active. For this purpose, the aging law for current as a function

of time, an example of which is shown in Fig. 3, is a proper starting place.

The aging law for threshold current can be written in the general form[11]

$$I(t, T') = \left[C + \int_0^t G(t', t, T)dt' \right] F(T'), \qquad (1)$$

where C is a constant, and the integral evaluates the increase in current due to degradation, when the device is stressed at some temperature T for a time t. $F(T')$ is an isochronal function that may be the $\exp(T'/T_0)$ expression typically used for lasers, where T_0 is the characteristic temperature.[12-17] Finally, the application of eq. (1) to actual cases requires care. In the first place, the extent of degradation (due to any combination of current, temperature, optical power, etc.) over any time t is evaluated by the integral of eq. (1). In the special case of isothermal aging, the temperature T is held constant over the aging time t. In the second place, when the aging process is halted at any time t, the instantaneous temperature sensitivity of current is accounted for entirely by the isochronal function $F(T')$.

Equation (1) assumes that as the laser degrades, the isochronal dependence of threshold current on temperature, at any time, remains unchanged. Our data indicate that this is usually a valid assumption. To confirm this, we show in Fig. 4 a plot of the measured relative change in threshold current due to aging at T, i.e., $\Delta I(t, T)/I(0, T)$ and the resulting relative change measured at 10°C, $\Delta I(t, 10)/I(0, 10)$. In spite of the obvious scatter in the data points, it is possible to obtain a linear regression analysis that gives

$$\frac{\Delta I}{I} (10°) = (1.027 \mp 0.047) \frac{\Delta I}{I} (T) + 0.064. \qquad (2)$$

It is seen from eq. (2) that there is a one-to-one correspondence in the *relative* change in threshold current at various temperatures. The cause of the small residual intercept is uncertain. It could be due to measurement error, transients associated with temperature changes, or a small but finite change at 10°C that deviates from a linear relationship. In any case, it is concluded that for significant changes, aging lasers at one temperature produces the same relative changes in threshold at other temperatures. In other words, the characteristic temperature T_0 of the laser remains nearly the same.

The formalism of the aging law given in eq. (1), in principle, can account for the history of the device through the function G. In other words, at any time and temperature the degradation rate depends on the extent of degradation that took place previously. Although concep-
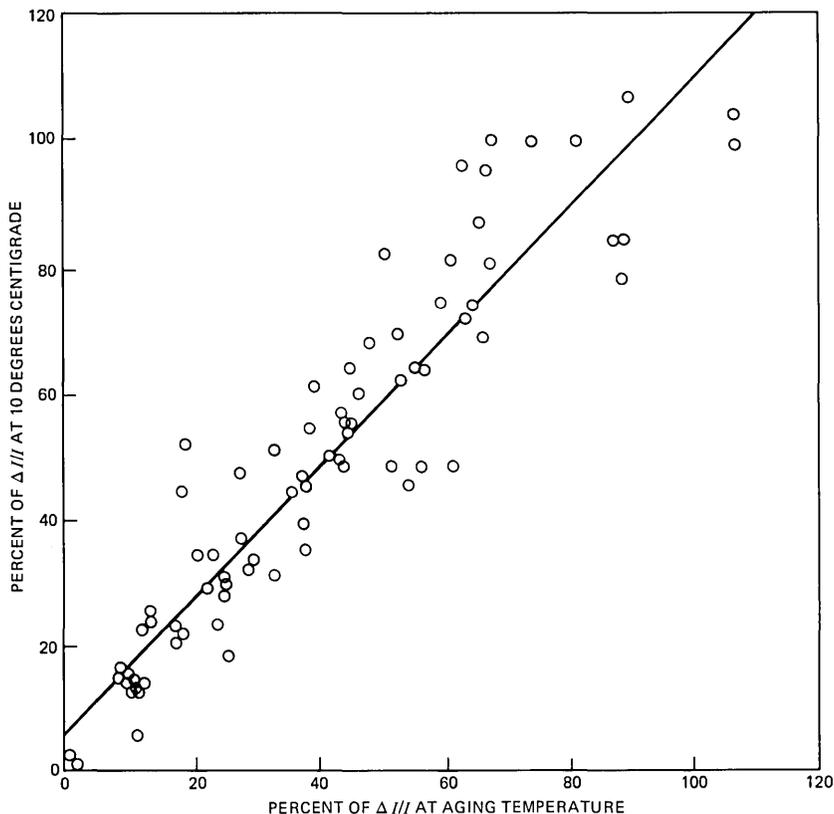
Fig. 4—Correlation between relative changes in threshold incurred at high temperature aging and the corresponding changes measured at 10°C.

tually useful, the integral form will be supplanted by empirical closed-form expressions that can be applied to actual cases. Thus, the isothermal aging law for threshold current can also be written in the form

$$I(t, T') = I_0[I^s_T(t) + I'_T(t)]F(T'), \qquad (3)$$

$$I(0, T') = I_0 F(T'), \qquad (4)$$

where I_0 is a dimensional constant, $I(0, T')$ is the initial value of current at T', $I^s_T(t)$ is a function that accounts for the component of degradation that saturates with time, and $I'_T(t)$ is the nonsaturable long-term degradation function. Finally, in eq. (3) the normalized sum of the functions $[I^s_T(t) + I'_T(t)]$ depends at any time on the thermal aging history of the laser but is not an explicit function of the instantaneous temperature.

A physically reasonable model for the observed saturable current of

Fig. 3 involves diffusion of a nonradiative or optically absorbing center.[9,18] However, a detailed analytical model for such a complicated process is not yet available. We are thus forced to resort to an empirical formulation for the saturable current. After considering a number of analytical expressions, including powers of time, we found the following expression to fit the data best:

$$I_T^s(t) \propto s\left[1 + \exp\left(\frac{t_0 - t}{\tau}\right)\right]^{-1}, \tag{5}$$

where s is the relative strength of the saturable current, t_0 is the incubation period of the saturable current, shown schematically in Fig. 3b, and τ affects the rate of change of the saturable current at t_0. (In many devices, the saturable current does not appear immediately, but rather takes several hundred hours—depending on the stress temperature—before it occurs, i.e., t_0 varies from laser to laser.) Both time contants t_0 and τ are expected to vary with temperature, as will be discussed in later sections.

Finally, the long-term degradation component $I_T^l(t)$ can be described in many ways. The simplest is a linear dependence. Our data on unpurged lasers are consistent with this form, and, in addition, it provides a more conservative approach than the sublinear model. Therefore, we will assume that

$$I_T^l(t) \propto 1 + Rt. \tag{6}$$

The empirical current aging law is obtained by combining eqs. (5) and (6). For the isothermal case and for $T = T'$, we obtain

$$\frac{I(t, T)}{I(0, T)} = \left[1 + Rt + \frac{s}{1 + e^{(t_0-t)/\tau}}\right]\bigg/ N, \tag{7}$$

where

$$N = 1 + s/(1 + e^{t_0/\tau}). \tag{8}$$

N is the normalization constant such that at $t = 0$, $I(t, T)/I(0, T) = 1$, and T is both the aging and measurement temperature.

Figure 3a shows a fit to the data in which $R = 2.2 \times 10^{-5}$ hours^{-1}, $s = 0.46$, $t_0 = 260$ hours, and $\tau = 110$ hours. It is evident that the aging law fits the data. Furthermore, the absence of a clear superlinearity in the long-term aging behavior, in a large number of our unpurged devices, argues against a strong current-driven aging mechanism. Therefore, in the absence of information to the contrary that may be revealed by future testing, the linear model will give us a useful measure of degradation at low temperatures and long times, and will be used in our projections.

2.4 Reliability results

2.4.1 Activation energy

The long-term degradation rate R of threshold current is assumed to obey the Arrhenius relation

$$R = R_0 e^{-E_a/kT}, \tag{9}$$

where R_0 is a constant, E_a is the activation energy, k is Boltzmann's constant, and T is temperature. The activation energy is evaluated in the postknee regime by measuring the function $I_T'(t)$ at two temperatures, T_1 and T_2. Thus, eqs. (6) and (9) give

$$E_a = \frac{kT_1 T_2}{T_2 - T_1} \ln \left(\frac{dI_{T_2}'/dt}{dI_{T_1}'/dt} \right). \tag{10}$$

Conversely, the activation energy can be obtained from eq. (7) in the postknee regime, i.e., $t \gg t_0$, τ by substituting into eq. (10) the derivatives of the total normalized currents. (This presupposes that t_0 and τ for the saturable current have the same activation energy. However, this activation energy can be different from that for the long-term degradation mode, as will become evident from Section 2.4.2.)

Most of our data presently available on activation energy were obtained from the step-stress experiments of Figs. 1 and 2. Although the step-stress test sequence of Fig. 2 is not complete, a fairly substantial amount of data has already been taken. These data, although preliminary in nature, are considered to be sufficiently accurate to make valid projections to the system operating conditions. The isothermal data, which will be treated statistically separately,[4] will be used as a general corroborative test of the validity of step-stress results.

Presently, the potential number of data points for determining activation energy is 177. This is the result of step-stress testing 117 lasers, as in Figs. 1 and 2. (The amount of data will increase further as the test program proceeds.) Given the myriad factors associated with the current aging behavior that can reduce the accuracy of deriving an activation energy, the useful number of data points was less than the total. The data for activation energy, obtained from step-stress experiments, were screened according to the following criteria:

1. The data were taken in the postknee regime of Fig. 3. This was to ensure that the activation energy was being derived only for that mechanism that governs the long-term wear-out of the device, as would apply for purged devices.[5]

2. The degradation rate at high temperature must be sufficiently low to establish a (nearly) linear degradation rate over a significant (≈ 500-hour) period of time.

3. The degradation rate at low temperatures must be sufficient (>0.3 percent per kh) to allow an accurate determination of its value (see Appendix A) over a reasonable test interval (3 kh).

4. Devices should not exhibit severe sensitivity to temperature changes—i.e., transients in the aging behavior—during step-temperature testing.

Out of 177 potential data points, only 26 measurements met the above criteria. The activation energies from these measurements are shown in the distribution plot of Fig. 5. In spite of the precautions listed above, there is still a significant spread in the data. This spread is due to a combination of experimental errors and device variations. It is suspected that the former contributed significantly (≈ 0.1 eV) to the spread of the distribution curve shown in Fig. 5. From the data of Fig. 5, as well as data to be discussed in subsequent sections, a median value of 1 eV for activation energy is reasonable. Figure 5 also indicates that the central 96 percent of the values of activation energy are between 0.76 and 1.28 eV.
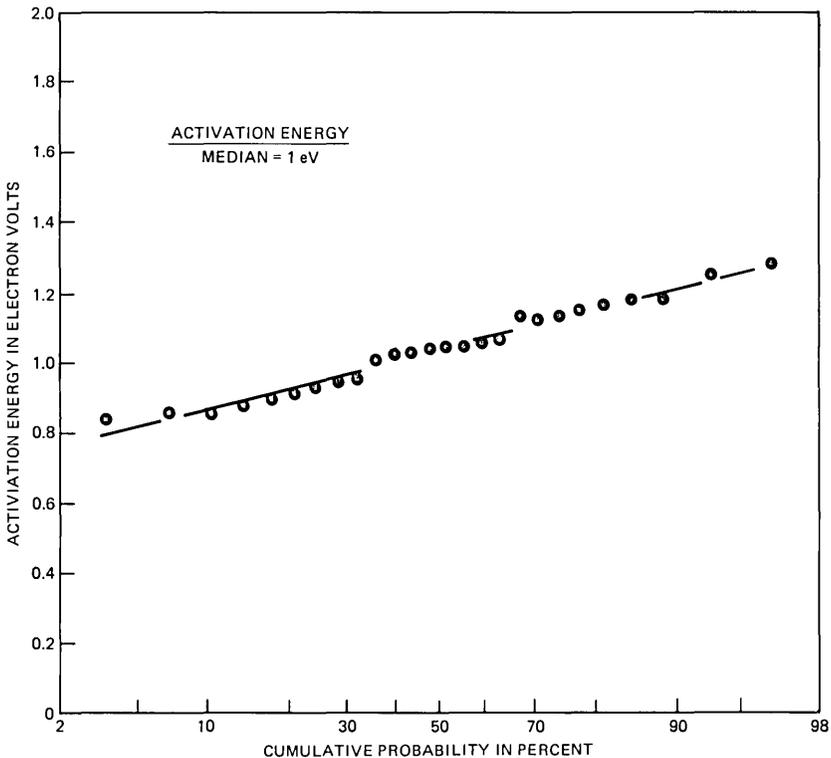
Fig. 5—Distribution curve of activation energy obtained from step-temperature stressing.

Table I—Activation energy and step-stress sequence

Activation Energy (eV)	Temperature Step (°C)					
	80 → 70	70 → 60	60 → 70	70 → 50	50 → 70	40 → 50
Mean	1.09	1.07	0.92	1.07	1.01	0.88
Median	1.02	1.17	0.90	1.10	1.03	0.82
Standard Deviation	0.31	0.33	0.08	0.15	0.31	0.30

Although only 26 measurements of activation energy are presented in Fig. 5, the rest of the data need not be entirely discarded. It can be used in a semiquantitative way to study such effects as temperature reciprocity, the applicability of the Arrhenius relationship, and memory effects. For this purpose, all the data from step-temperature stressing were used, and the results from various test sequences are given in Table I. (Table I also includes the results of step-temperature stressing another group of devices between 40 and 50°C. This is a separate group from those indicated in Figs. 1 and 2.) It is clear from the data that without the application of the guidelines stated above, the standard deviation of the activation energy increased by almost a factor of two. However, the mean and median values were still in the 1-eV range. In addition, two important observations could be made from the results of Table I. First, the mean activation energy, as obtained from the Arrhenius relation, was relatively invariant with temperature over the range of 40 to 80°C. Second, when step-temperature measurements were made in the postknee regime of the current aging characteristic, the derived activation energy showed no significant (∓ 0.1 eV) dependence on the sequence of testing.

The lasers undergoing isothermal testing at 60 and 40°C provided activation energies that corroborated the step-temperature results. However, exact agreement was not possible because the isothermal tests suffered from the following difficulties: (1) With the limited testing time available (≈ 7000 hours), the 60°C group had device currents that developed well beyond the knee; the 40°C group, with 13,700 hours of testing, still exhibited increased degradation rates associated with the residual saturable current. (2) Prior to screening (to be discussed below), the laser-to-laser variation within a slice and the slice-to-slice variation in the degradation rate exceeded (by approximately two orders of magnitude) the expected difference in degradation rate associated with temperature. (3) The number of devices being tested at each temperature (40 lasers) was probably too small to reduce the expected uncertainty in the activation energy below that derived from step-temperature testing.

To overcome the problems of the second difficulty above, data on five slices (H, I, L, M, and N of Appendix B) that were common to

the two test groups were subjected to a linear regression analysis. In addition, to avoid bias in the results, the following measures were taken: (a) in each slice, if a certain number of lasers (0 to 20 percent) degraded too rapidly at high temperature to provide a meaningful measurement, an equal number, from that slice, of the most rapidly degrading lasers at low temperatures was eliminated; (b) in each slice, if a certain number of lasers (0 to 15 percent) at low temperature were not degrading sufficiently to be accurately measured, then an equal number of the slowest degrading lasers at high temperature were also eliminated from the population. This, in essence, maintained the integrity of the heart of the distributions at both temperatures. The resulting activation energy for all the remaining lasers obtained from the five different slices was 0.79 eV, with a standard deviation of 0.19 eV. This was lower than the median value of 1 eV obtained from step-temperature stress. The reason was probably due to some residual effects of the saturable current in the 40°C data. This was supported by the observation that two of the slices (I and M of Appendix B) had low values of saturable currents. The isothermal data on these two slices, when considered separately, gave an activation energy of 1.0 eV and a standard deviation of 0.25 eV. Hence, it is expected that longer testing at 40°C will give lower degradation rates, as the influence of the saturable current component decreased. This will increase the values of activation energy obtained from isothermal tests and provide a more accurate lower bound on the activation energies. Finally, this analysis of the isothermal data is an alternative to the method of deriving the activation energy used by Amster.[4]

2.4.2 Saturable current

One of the factors that complicated the interpretation of both the step-stress and the isothermal data was the presence of the saturable current. This current manifested itself experimentally as shown in Fig. 3 and was represented analytically by the empirical relation given in eq. (5). Although eq. (5) is strictly empirical, it is a useful tool that provides a quantitative measure of the strength of the saturable current and its rate of formation. Thus, from eq. (5) a critical time t_c is defined by

$$t_c = t_0 + \tau, \tag{11}$$

which is the time required for the saturable current to achieve ≈ 73 percent of its final value. To obtain an activation energy for the saturable current, the data of the isothermal test groups, operating at 40 and 60°C, were utilized. In addition, the data obtained during the first phase of testing of the step-stress groups at 50 and 70°C were also analyzed to estimate the values of t_0, τ, and t_c.

These data are plotted as a function of reciprocal temperature in Fig. 6. For each temperature, the median value of critical time (the squares) and the standard deviation (the bars) are indicated. The large variance in t_c at each temperature is obvious. This may be due to large slice-to-slice variation in the data obtained from 16 different slices. Nevertheless, the trend for t_c to decrease at high temperature is quite clear. A linear regression analysis of the data on 73 lasers gave
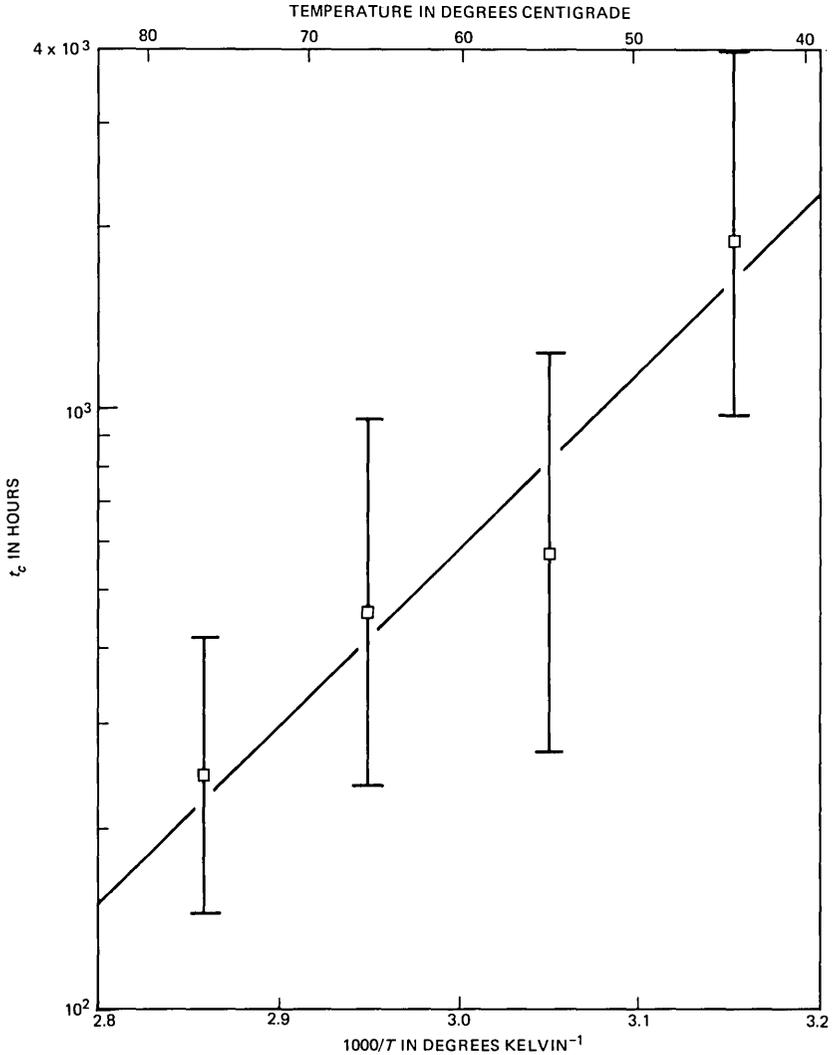
Fig. 6—Data for critical time versus reciprocal temperature, for the saturable current component. The linear regression uses data on 73 lasers.

$$t_c = 7.6 \times 10^{-7} \, e^{E/kT} \text{ hours} \tag{12a}$$

and

$$E = 0.59 \mp 0.06 \text{ eV}, \tag{12b}$$

where the 0.06 eV is the standard deviation. This implies that the time constant for the saturable current has an activation energy that is almost half the value of the main wear-out mechanism. Furthermore, the fit of the linear regression curve to the data is adequate, considering the fact that the curve passes close to the median values at all four temperatures.

Additional observations indicate the following. First, a comparison of the degradation rate in the postknee regime with the strength of the saturable current gave practically no correlation. In other words, the strength of the saturable current was not related in a direct way to the long-term wear-out mechanism. Second, the onset of the saturable current was accompanied by a reduction in the external differential quantum efficiency of the laser. And, finally, the junction voltage at threshold also increased with time in a manner similar to that of the saturable current. Most of these observations are consistent with a model in which the saturable current is associated with an increase in leakage current.[5]

Further clues about the saturable current can be obtained by noting that its activation energy is 0.59 ∓ 0.06 eV. For impurity diffusion in InP, the closest activation energy is that of gold, whose diffusivity is given by the relation[19]

$$D = 1.32 \times 10^{-5} \exp[(-0.48 \mp 0.01)/kT] \text{ cm}^2/\text{s}. \tag{13}$$

Although the activation energies are close, it is not certain whether gold plays a role in the saturable mode of degradation of lasers. The details of such an interaction are even less well understood. We can only conjecture that the complicated reactions involve thermal, electrical, and possibly optical effects.[14,19-28] It is also possible to invoke more complicated models to derive a different value for the activation energy. For instance, formation of dark-spot defects has been described analytically by an exponential function of temperature and a current-squared term.[29] In such a case, the data of Fig. 6 yield an activation energy of 0.24 eV. However, we do not at present have data to justify the use of models that are any more complicated than the simple Arrhenius relation of eq. (12a).

2.4.3 Rate dependence on optical power

A determination of how the degradation rate depends on optical power is necessary for many reasons. From a fundamental point of

view, it is desirable to know whether the long-term wear-out mechanism is being driven predominantly by current (voltage), nonradiative recombination, or the stimulated power. Each of these factors suggests different solutions. From a practical point of view, it is necessary to know how the results of aging lasers under constant power output would compare to the real situation of biasing the lasers at threshold but pulsing them ($\approx$50-percent duty factor) up to the required power output.

As in our methodology for activation energy, we determined the dependence of degradation rate on optical power in the postknee regime by step-power stressing. As shown in Fig. 1, the test was done at constant temperature, i.e., 60°C, and by changing the power between 3 and 5 mW. Furthermore, to safeguard against memory effects, the test was done twice by reversing the power sequence ($3 \rightarrow 5$ and $5 \rightarrow 3$). Finally, in a sequence not shown in Fig. 1, the test is continuing at different power levels (1 and 7 mW), at 60°C.

The dependence of the degradation rate on power $R(P)$ was derived from the step-power experiments by performing a linear regression analysis of the results obtained between any two power levels. This correlation analysis gave the average change in R with power as well as the standard deviation in the estimate of the median value. The results are plotted in Fig. 7. The degradation rate at any power was normalized with respect to its value at 5 mW/facet, i.e., $R(5)$. It is clear from Fig. 7 that as the optical power was reduced, the degradation rate decreased. These results suggest that the dependence of degradation rate on optical power can be represented empirically by the relation

$$R(P) = r_0 + r_1 P, \tag{14}$$

where r_0 is a residual degradation rate which is independent of optical power. For these 1.3-μm lasers, and at 60°C, the value of r_0 was roughly 15 percent of the total degradation rate at 5 mW, as can be seen from Fig. 7.

Equation (14) and the results of Fig. 7 are qualitatively consistent with the results obtained previously.[10,30–32] Except for differences in the functional dependence on power, the results indicate that the degradation rate is comprised of a component r_0 (which may be current driven)[5] and another component $r_1 P$, which is power driven. That this second component is in fact driven by the optical power, and not by current, can be appreciated as follows. At 60°C, the ratio of the drive current at 5 mW to that at 3 mW, $I(5)/I(3)$ is 1.16. Suppose that the observed corresponding change in the ratio of degradation rates of 0.67 is due to current. Therefore, if $R \alpha I^m$, a value of $m = 2.6$ is needed to explain the results. Now this strong dependence of degradation rate
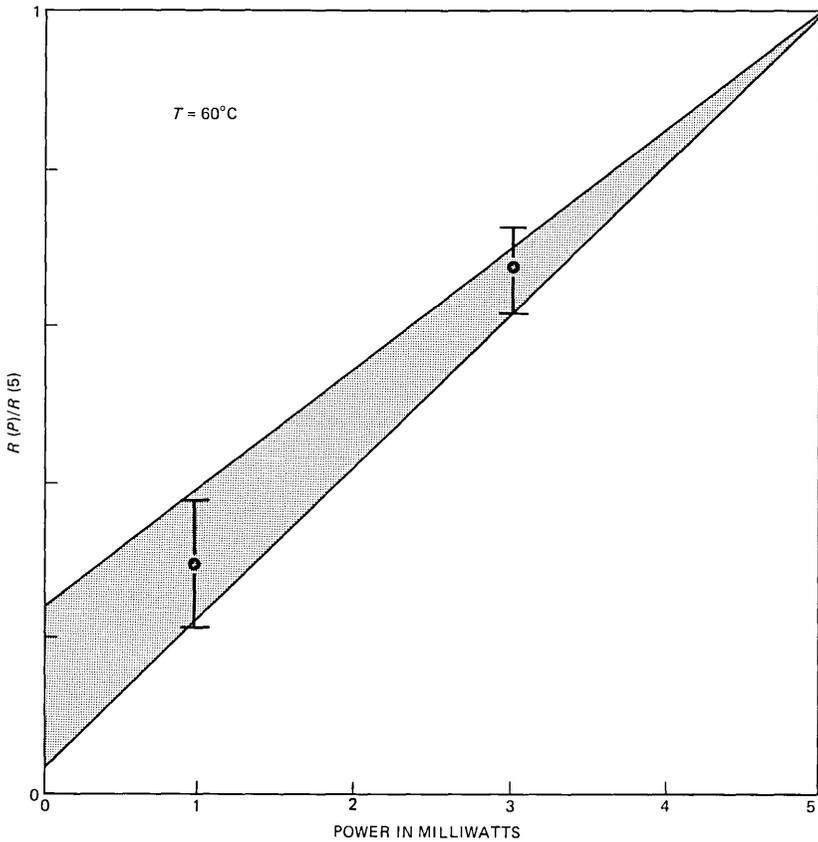
Fig. 7—Dependence of the degradation rate of threshold on optical power measured at 60°C.

on current is inconsistent with our observations on the aging law. A value of $m = 2.6$ leads to a strong superlinearity in the aging characteristics of the lasers. Our data show otherwise. The aging behavior is linear at best and may even have a slight sublinearity in some cases. Hence, a strong current-driven degradation mechanism for our lasers at 5 mW cannot be supported by the observed aging law. Finally, the mode-switching behavior that we sometimes observe, and which will be discussed in subsequent sections, also suggests that the optical fields play an important role in the long-term degradation process.

In summary, the degradation rate at 5 mW per facet and 60°C is made up of two components. One component accounts for ≈15 percent of the total value and may be current driven.[5] The remaining component is driven by optical power. That the dominant degradation process is driven by optical power has at least three justifications. First, for current to explain the degradation results, a strong depend-

ence of degradation rate on current is required. This would imply a superlinearity in the aging law, which is inconsistent with observation. Second, in the step-power experiments at constant temperature, the threshold current is nearly constant. Hence, for significant quantum efficiencies, the residual degradation associated with threshold should remain constant. Finally, the excitation of high-order transverse modes in some degraded lasers suggests a spatially nonuniform degradation, as would result from the optical fields.

2.4.4 Projected operation at 10°C

Out of a total number of 262 lasers tested in this program, more than half (145 lasers) are under isothermal stress test. These groups do not yield activation energy values for individual lasers. The remaining step-temperature-stressed lasers do provide individual activation energies, but only the small fraction shown in Fig. 5 are considered to be sufficiently accurate. The mechanics of projecting the high-temperature data to 10°C becomes problematic. It is possible to treat all the isothermal groups separately, as in the statistical analysis of Amster,[4] or some assumptions can be made that allow all the high-temperature data to be utilized fully.

When a laser is step-temperature stressed between T_1 and T_2, its operation at T_3 can be individually projected. However, there are two sources of error. First, there is the error of estimating its individual activation energy from the measurements at T_1 and T_2. For example, if T_1 and T_2 are 60 and 70°C, respectively, then a 10-percent error in estimating the ratio of degradation rates leads to an error of 0.094 eV in the activation energy. This may seem rather acceptable. However, at 10°C the resulting error in the projected degradation rate is 98 percent, i.e., an order to magnitude larger than the starting measurement error. The reason for this magnification is easy to appreciate in terms of the large "lever arm" involved in such projection.

To minimize the error in projection, the primary source of error associated with activation energy should be minimized. One way to accomplish this is to use the best estimate of activation energy obtained on a fraction of the devices, and use this energy as a "population" parameter to all the lasers. This way, even the lasers being isothermally tested at 40 and 60°C can be projected to 10°C by applying a population activation energy to their measured degradation rates. Our best estimate of activation energy is derived from the results of Fig. 5, in which the median value is 1 eV. Furthermore, 96 percent of the values of activation energy obtained are larger than 0.76 eV and smaller than 1.28 eV.

We have attempted to project the data to 10°C in a manner that will be as representative of the final manufactured product as possible,

which is a difficult task. In the first place, the majority of devices being tested in this program have not been subjected to the rigorous 60°C burn-in screen that the final product will experience. (As described previously, the majority of these devices were subjected to a 50°C/5 mW/100h screen.) This burn-in screen eliminates most lasers with large initial degradation rates (> 50 percent/kh) and large saturable currents. In addition, the present lasers have not been subjected to the purge program, whereas the final product will be.[5] Although the effect of the purge on our projections cannot presently be estimated, it is possible to account for the 60°C burn-in screen. This was done by removing from the statistics all lasers that would fail this burn-in test, had they been subjected to it. This information was available from the degradation of individual devices during the first one hundred hours of stress aging.

When these lasers were subjected to this simulated burn-in screen, and the individual degradation rates projected to 10°C by using a population activation energy, the results are as shown in Fig. 8. For the best estimate of 1 eV for the activation energy, the median degradation rate at 10°C is 4.1×10^{-3} percent/kh. For almost 80 percent of the population, the distribution is fairly well lognormally distributed with a sigma value of 1.8. However, for the remaining 20 percent of the population, which is more rapidly degrading, the curve exhibits truncation. This truncation is due to the 60°C burn-in screen that eliminates rapidly degrading lasers. Further truncation is expected in the final product to be used in the system. The additional truncation results from the added "certification" test to which every laser will be subjected before actually being used.

It is possible to use the projected values of degradation rates at 10°C to estimate the light bulb life of the lasers. Thus, the light bulb Time-To-Failure (TTF) is defined, somewhat arbitrarily, to be

$$TTF \triangleq 10^5/R(\text{percent/kh}) \text{ hours.} \tag{15}$$

For a linear aging law, eq. (15) implies that the light bulb failure occurs when the threshold current increases by 100 percent. Our transmitter testing on a few lasers, to be discussed elsewhere,[6] shows that (15) is a conservative estimate of light bulb failure. This conservative light bulb TTF definition provides a safety margin to protect against unforeseen functional failures.

Figure 8 shows that for an activation energy of 1 eV, the median TTF at 10°C is slightly greater than 2×10^7 hours. The lower 98-percent confidence limit on activation energy yields a median TTF of 5.7×10^6 hours at 10°C. Also evident in Fig. 8 is the truncation of the distribution curve, which removes devices with relatively short projected lives at 10°C. It has the effect of reducing the rate of early
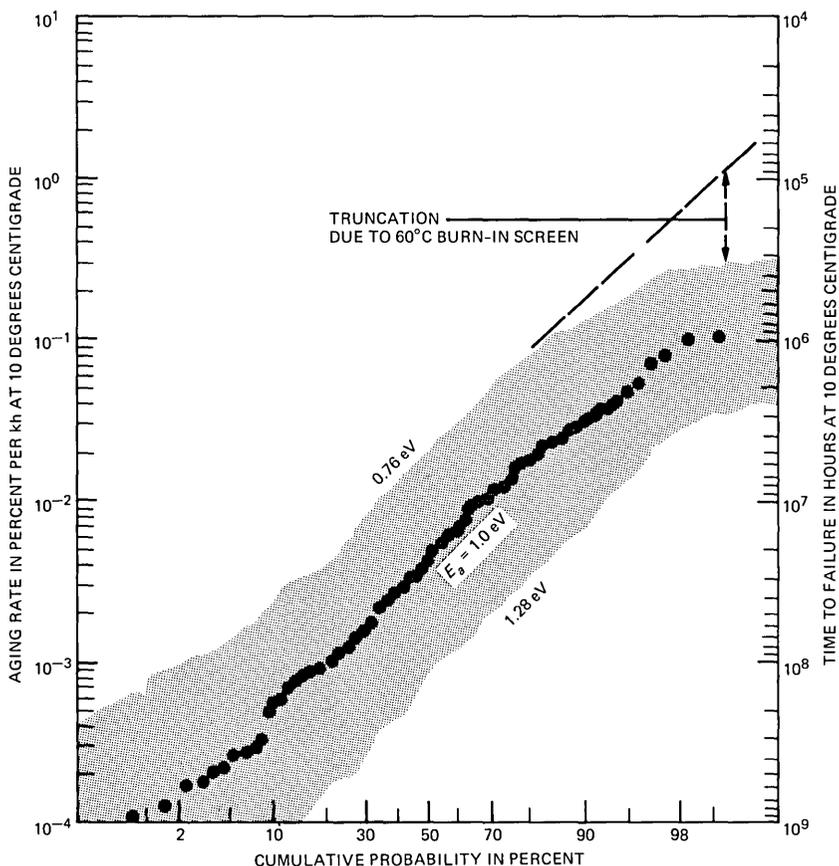
Fig. 8—Degradation rates of individual lasers projected to 10°C using various estimates of activation energy.

failures in the system, below the value expected from the usual log-normal distribution.

2.5 Potential functional failures

Given the relatively long light bulb lifetimes predicted for the 1.3-μm laser at 10°C, it seems prudent to surmise that if and when failures occur, they will be due to operational malfunctions. These potential malfunctions were mentioned in Section 2.1 and include, among other things, change in extinction ratio, excitation of high-order transverse modes, spectral splitting, self-induced pulsations, and fiber-to-laser coupling instabilities in the laser package. Given the incomplete state of our knowledge of these mechanisms, the ability to predict their occurrence is limited. However, some preliminary information about their rate of occurrence and the impact on transmitter performance is beginning to become available.[6]

The occurrence of functional failures should be viewed initially in the context of the overall number of device hours of testing. Thus, the total number of device hours presently available is $\approx 2.4 \times 10^6$ hours. This covers the whole range of temperatures from 10 to 80°C. It is not known, at present, the extent to which functional failures are accelerated with temperature. However, in terms of light bulb life, with an activation energy of 1 eV, the various testing at different temperatures is equivalent to 9.3×10^8 device hours at 10°C. Hence, for a configuration that uses ≈ 1000 active lasers, the equivalent time is ≈ 106 years of system operation at 10°C.

With all the caveats stated above in mind, we can say that there is no available evidence to indicate that functional failures will have a severe impact on system reliability. However, we have identified (in three rapidly degrading lasers that normally would not pass the 60°C burn-in screen) certain anomalies that affect, for instance, the coupling efficiency into the single-mode fiber. Such an example is shown in Fig. 9. The laser had been stressed at 40°C to the extent that its threshold increased by ≈ 50 percent. After this amount of degradation, the laser power output at 10°C, as monitored by the back-face detector, showed an abrupt change in quantum efficiency at a current 9 mA above threshold, as shown in Fig. 9a. The output from the single-mode fiber is even more remarkable. For currents between the threshold value (43 mA) and 52 mA, the fiber coupling efficiency was very low. However, the coupling efficiency above 52 mA increased, as evident from Fig. 9a. The optical spectrum, when measured through the single-mode fiber, is shown in Fig. 9b. For wavelengths between 12,850 and 12,875Å, there was evidence of another family of longitudinal modes that are closely spaced to the main optical mode. It was therefore concluded that this optical spectrum was comprised of a fundamental and a high-order transverse mode whose coupling efficiency to the single-mode fiber was rather low.

There was also direct evidence that the initially fundamental TE_0 (Transverse Electric) mode could change, in some rapidly degrading lasers, to a high-order TE_n mode. An example of an extreme case is shown in Fig. 10. Initially, the far-field emission pattern indicated a well-behaved fundamental transverse mode. After accelerated aging, which produced a 70-percent increase in threshold, the far-field pattern became as shown in Fig. 10. (The final far-field measurements were made by R. T. Ku, who used a numerical aperture of 0.2. The initial far-field measurement used a numerical aperture of 1, which may have lead to some smoothening of the fine structure in the radiation pattern.) The dominant mode in the degraded laser became a TE_1 transverse mode. However, optical spectral measurements of the same laser indicated the presence of four distinct families of
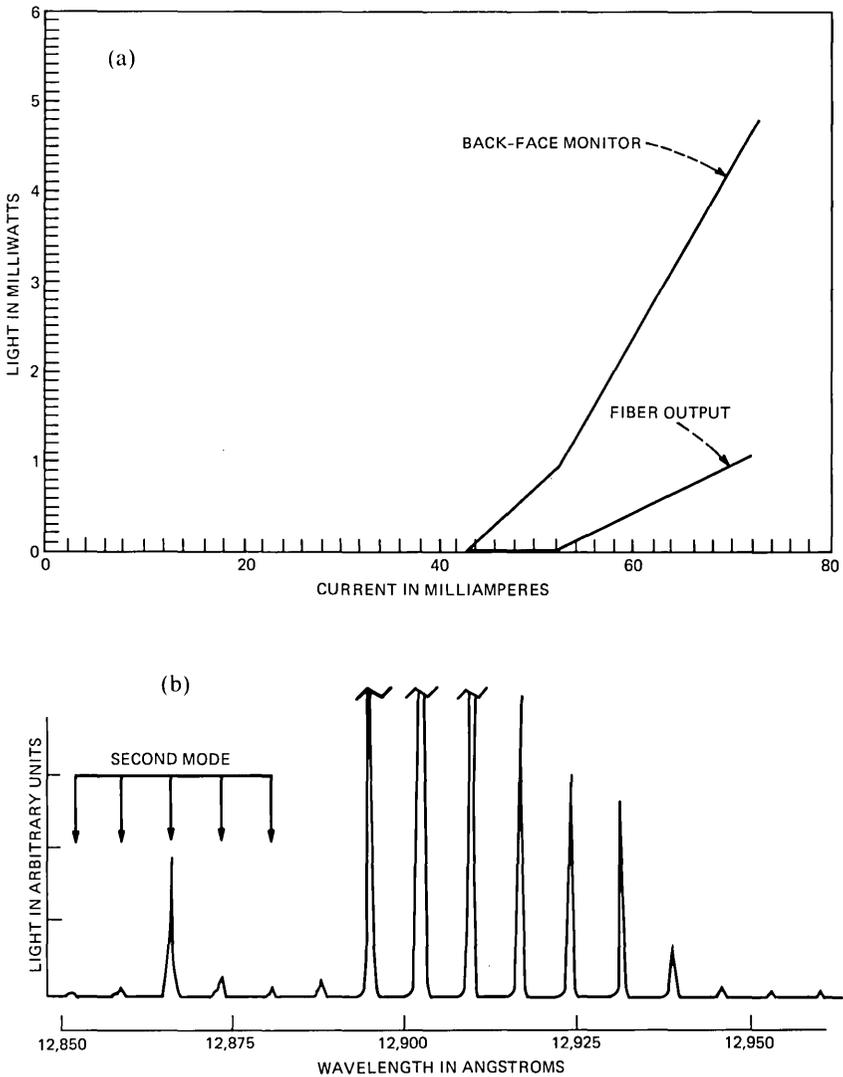
Fig. 9—An example of potential malfunction of a degraded device measured at 10°C. (a) Light-current characteristic and (b) spectrum measured through the optical fiber at 5 mW.

longitudinal modes. Hence, in addition to the TE_0 and TE_1 modes, there were two TE modes that were not very distinct in Fig. 10. In general, the high-order modes encountered after degradation were odd-numbered and TE polarized. Inspection of the mirror facets indicated no obvious erosion. Therefore, the degradation process seemed to be in the bulk of the active regions, and may be similar to the "spottiness" observed in degraded GaAs lasers.[33]
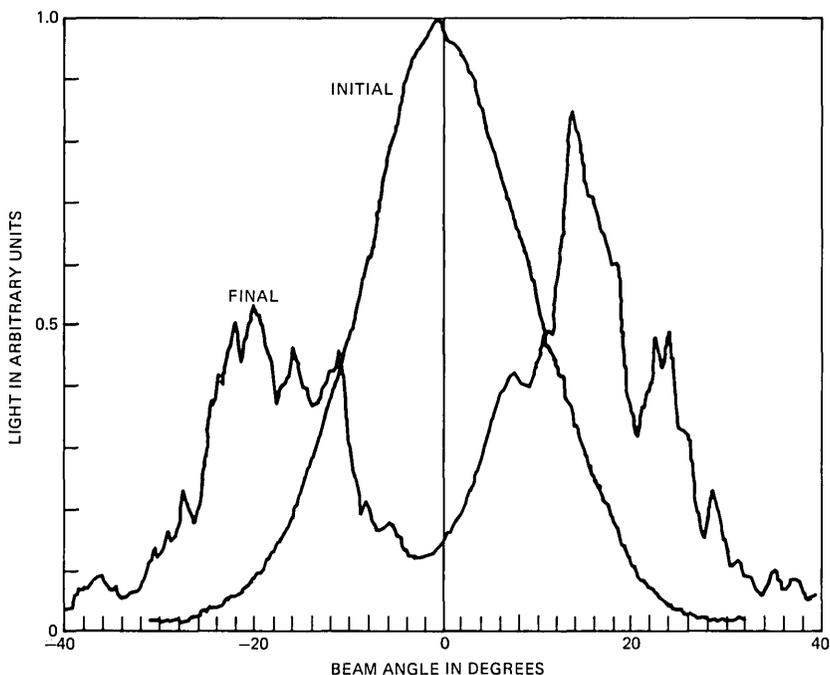
Fig. 10—An example of extreme distortion in the far field of a degraded laser obtained from slice C of Table III.

Further diagnostic tests were performed on selected lasers that exhibit light-current nonlinearities, additional spectral modes, and reduced fiber coupling. In all such cases, the width of the active region was found to exceed the 1- to 1.7-μm target value. For instance, the laser whose characteristic is shown in Fig. 9 was obtained from slice N of Table III in Appendix B. Lasers from this slice had an unusually high rate of incidence of a secondary family in the optical spectra, indicating high-order transverse modes. Several lasers from this slice had their mirrors etched in a potassium-ferrocyanide solution. Subsequent measurements of the etched mirrors in a Scanning Electron Microscope (SEM) indicated that the width of the active region for slice N was ≈ 2.5 μm, i.e., about 1 μm larger than the desired value. The laser whose far-field pattern was shown in Fig. 10 deviated even further from the nominal value. Its active region was ≈ 4 μm wide. In all cases of clear evidence of high-order modes, the width of the active region exceeded 2 μm. The intriguing aspect of this result was that all the lasers initially had a fundamental transverse mode. The high-order modes developed in degraded lasers if the active region exceeded ≈ 2 μm.[32] Therefore, better production or screening control should eliminate these potentially malfunctioning lasers. Such screening tests

may involve freedom from "kinks" in the light-current curves up to 10 mW, noninvasive stripe width measurements, far-field radiation patterns, and spectral measurements. In addition, it is possible to obtain a direct measure of stripe width by performing etching and SEM measurements on the nearest-neighbor laser to that intended for system use.[34]

It is clear that the above examples are but a few of the many ways in which lasers could fail functionally. So far, we have encountered only two degraded (out of 262) lasers that developed weak (−90 dBm) self-induced pulsations occurring at 2.4 and 3.1 GHz, respectively. This is at variance with the high rate of incidence of self-induced pulsations in proton-bombarded GaAs lasers.[35] The reason may be due to the strong index guiding properties of buried heterostructure lasers.[36] However, this by no means exhausts all the possibilities of functional failures, which are actively being investigated.[6]

III. CONCLUSION

In this phase of prequalification reliability program we have defined both the reliability potential as well as the testing methodology of 1.3-μm lasers. There are stringent reliability conditions imposed on all components used in submarine cable systems. This requires accurate knowledge of the fundamental reliability parameters that influence long-term projections. Our experiments on unpurged lasers revealed short-term saturable effects as well as long-term degradation. By carefully choosing experiments in the laser program, we have been able to establish a population value for the activation energy of 1 eV for the long-term degradation process. Furthermore, 96 percent of the measured values of the activation energy are between 0.76 and 1.28 eV. These activation energies allow projections to 10°C that imply a most probable median time to failure in excess of 2×10^7 hours, for light bulb life. By the same token, with 98-percent confidence level, the median time to failure at 10°C is greater than 5.7×10^6 hours. These projections are based on devices that have been subjected to a short (100-hour) screen at 60°C, which eliminates most potential early failures. An additional long-term screen at 60°C is planned for the laser certification program. Both screening operations lead to a severe truncation of the distribution curve for short-lived devices. This reduces the early failure rate below the value expected from the usual lognormal distribution.

The initial phase of this program studied in detail the light bulb life of 1.3-μm lasers and determined that it is adequate for submarine cable application. It is obvious that there is another class of failures that can have an adverse impact on laser system performance. For instance, nonthermally activated failure mechanisms or early random

failures may occur in some devices. These potential failures we expect to reduce to a minimum by subjecting the lasers to a severe purge program.[5] Another class of failures is related to operational malfunctions. This class is currently being studied actively and the information is quite preliminary. However, there are no indications, so far, to lead us to believe that operational failures should significantly alter our conclusions about the utility of the 1.3-μm laser for submarine cable use.

IV. ACKNOWLEDGMENTS

REFERENCES

1. M. Hirao et al., "Fabrication and Characterization of Narrow Stripe InGaAsP/InP Buried Heterostructure Lasers," J. Appl. Phys., 51 (August 1980), pp. 4539–40.
2. K. Mizuishi et al., "Accelerated Aging Characteristics of InGaAsP/InP Buried Heterostructure Lasers Emitting at 1.3 μm," Jap. J. Appl. Phys., 19 (July 1980), pp. 429–32.
3. K. Mizuishi et al., "Reliability of InGaAsP/InP Buried Heterostructure Lasers," Jap. J. Appl. Phys., 21, Supplement 21-1 (1982), pp. 359–64.
4. S. J. Amster, unpublished work.
5. F. R. Nash et al. "Implementation of the Proposed Reliability Assurance Strategy for an InGaAsP/InP, Planar Mesa, Buried Heterostructure Laser Operating at 1.3 μm for Use in a Submarine Cable," AT&T Tech. J., this issue.
6. C. E. Barnes and W. Smith, unpublished work.
7. D. S. Peck and C. H. Zierdt, Jr., "The Reliability of Semiconductor Devices in the Bell System," Proc. IEEE, 62 (February 1974), pp. 185–211.
8. G. A. Dobson and B. T. Howard, "High Stress Aging to Failure of Semiconductor Devices," Proc. Seventh Nat. Symp. Rel. Qual. Contr. Electron. (1961), pp. 262–72.
9. Y. Horikoshi, T. Kobayashi, and Y. Furukawa, "Lifetime of InGaAsP-InP and AlGaAs-GaAs DH Lasers Estimated by the Point Defect Generation Model," Jap. J. Appl. Phys., 18 (December 1979), pp. 2237–44.
10. H. Imai et al., "Accelerated Aging Test of InGaAsP/InP Double-Heterostructure Laser Diodes With Single Transverse Mode," Appl. Phys. Lett., 38 (January 1981), pp. 16–7.
11. R. G. Smith, private communication.
12. G. H. B. Thompson, "Temperature Dependence of Threshold Current in (GaIn) (AsP) DH Lasers at 1.3 and 1.5 μm Wavelengths," IEE Proc. I, Solid State Electron Dev., 128 (January 1981), pp. 37–43.
13. G. H. B. Thompson and G. D. Henshall, "Nonradiative Carrier Loss and Temperature Sensitivity of Threshold in 1.27 μm (GaIn)(AsP)/InP D. H. Lasers," Electron. Let., 16 (January 1980), pp. 42–4.
14. N. K. Dutta and R. J. Nelson, "Gain Measurements in 1.3 μm InGaAsP-InP Double Heterostructure Lasers," IEEE J. Quantum Electron., QE-18 (January 1982), pp. 44–9.
15. M. Yano, H. Imai, and M. Takusagawa, "Analysis of Electrical, Threshold, and Temperature Characteristics of InGaAsp/InP Double-Heterojunction Lasers," IEEE J. Quantum Electron., QE-17 (September 1981), pp. 1954–63.
16. A. Sugimura, "Band-to-Band Auger Recombination Effect on InGaAsP Laser Threshold," IEEE J. Quantum Electron., QE-17 (May 1981), pp. 627–34.
17. Y. Horikoshi and Y. Furakawa, "Temperature Sensitive Threshold Current of

InGaAsP-InP Double Heterostructure Lasers," Jap. J. Appl. Phys., *18* (April 1979), pp. 809–15.

18. W. C. Ballamy and L. C. Kimerling, "Premature Failure in Pt-GaAs IMPATT's—Recombination-Assisted Diffusion as a Failure Mechanism," IEEE Trans. Electron. Dev., *ED-25* (June 1978), pp. 746–52.

19. S. I. Rembeza, "Diffusion of Au in InP Single Crystals," Sov. Phys. Semicond., *3* (October 1969), pp. 519–20.

20. A. K. Chin et al., "Cathodoluminescence Evaluation of Dark Spot Defects in InP/InGaAsP Light Emitting Diodes," Appl. Phys. Lett., *41* (September 1982), pp. 555–7.

21. I. Camlibel et al., "Metallurgical Behavior of Gold-Based Ohmic Contacts to the InP/InGaAsP Material System," J. Electrochem. Soc., *129* (November 1982), pp. 2585–90.

22. O. Ueda et al., "Mechanism of Catastrophic Degradation in InGaAsP/InP Double-heterostructure Light Emitting Diodes and GaAlAs Doubleheterostructure Light Emitting Diodes Applied With Pulsed Large Current," J. Appl. Phys., *53* (December 1982), pp. 9170–9.

23. O. Ueda et al., "Transmission Electron Microscope Observation of Dark-Spot Defects in InGaAsP/InP Double-Heterostructure Light-Emitting-Diodes Aged at High Temperature," Appl. Phys. Lett., *36* (March 1980), pp. 300–1.

24. S. Yamakoshi et al., "Reliability of High Radiance InGaAsP/InP LED's Operating in the 1.2-1.3 μm Wavelength," IEEE J. Quantum Electron, *QE-17* (February 1981), pp. 167–73.

25. S. D. Hersee, "Long Lived High Radiance LEDS for Fiber Optic Communications Systems," Digest Seventh Int. Electron Dev. Meeting, 1977, pp. 567–70.

26. K. Wakita et al., "Transmission Electron Microscope Observation of Dark Defects Appearing in InGaAsP/InP Double Heterostructure Lasers Aged at Accelerated Operation," Appl. Phys. Lett., *40* (March 1982), pp. 525–7.

27. K. Ishida, T. Kamejima, and Y. Matsumoto, "Lattice Defect Structure of Degraded InGaAsP/InP Double-Heterostructure Lasers," Appl. Phys. Lett., *40* (January 1982), pp. 16–7.

28. S. Komiya et al., "Optically Induced Glide Motion of Misfit Dislocations in InP/In$_{1-x}$Ga$_x$As$_{1-y}$P$_y$ Double Heterostructure," J. Appl. Phys., *54* (February 1983), pp. 1058–61.

29. S. Yamakoshi et al., "Degradation of High Radiance InGaAsP/InP LEDs at 1.2-1.3 μm Wavelength," IEDM Tech. Digest, Washington, D.C., 1979, pp. 122–5.

30. K. Hori, H. Imai, and M. Takusagawa, "Long-Lived GaAlAs-GaAs DH Laser Diodes for Optical Fiber Communications," Fujitsu Sci. Tech. J., *15* (December 1979), pp. 95–109.

31. H. Ishikawa et al., "Accelerated Aging Test of Ga$_{1-x}$Al$_x$As DH Lasers," J. Appl. Phys., *50* (April 1979), pp. 2518–22.

32. K. Mizuishi et al., "Reliability of InGaAs/P/InP Buried-Heterostructure 1.3 μm Lasers," IEEE J. Quantum Electron., *QE-19* (August 1983), pp. 1294–301.

33. H. Kressel and J. K. Butler, *Semiconductor Lasers and Heterojunction LEDs*, New York, NY: Academic Press, 1977.

34. C. B. Swan, private communication.

35. T. L. Paoli, "Changes in the Optical Properties of CW (AlGa)As Junction Lasers During Accelerated Aging," IEEE J. Quantum Electron., *QE-13* (May 1977), pp. 351–9.

36. B. W. Hakki, "Instabilities in Output of Injection Lasers," J. Appl. Phys., *50* (September 1979), pp. 5630–7.

APPENDIX A

Aging and Testing Facilities

The variety of testing conditions required that a large quantity of lasers operating at various temperatures must be accommodated at the same time. The following criteria governed the design of the 1400 Continuous Wave (CW) and 120 pulsed aging sites for packaged 1.3-μm lasers. We will demonstrate in this Appendix that we have con-

structed a test facility which meets the criteria listed below, particularly the third item.

1. In-place testing and aging—For efficient aging of a large number of devices, in-place testing becomes mandatory.

2. Automated optical and electronic measurements—The large amount of data taken requires that all of the optical, as well as the electronic, measurements are made in situ and under computer control.

3. Precision and stability—The precision and stability of the measurements must allow the laser characteristics to be projected accurately to 200,000 hours, based on less than 10,000 hours of data on each laser. With proper balance of all design factors, we can make the most efficient use of an expensive facility.

A.1 Description of facilities

The SL laser CW aging facility consists of aging cabinets; an uninterruptible (battery reserve) power system; a master computer; and multiplexing for the electrical, optical, and radio frequency spectrum measuring equipment. Twenty different temperatures (70 lasers each) can be controlled simultaneously. All testing is performed in place and is initiated by a master computer. The operating currents of these lasers have temperature sensitivities in the range of 0.5 to 2 mA/°C, depending on the laser and the temperature. Approximately 0.1-mA stability is needed to take data in less than 10,000 hours and project it to 200,000 hours with greater than 98-percent confidence that the current will not double. We therefore made a preliminary estimate that ±0.05°C temperature control was necessary for testing InGaAsP lasers. The exact details depend on the laser and the conditions at which it is aged.

During aging, the computer-controlled temperatures of the devices are held at ±1°C within the range of 10 to 90°C. The lasers are aged at a current level that is continuously adjusted, using analog circuitry, to provide constant power output. This power output is typically 5 mW per facet, which is the system operating power level. Four times a day, the Light-Current-Voltage-Temperature (L-I-V-T) is measured to ensure the integrity of the automatic feedback circuitry and temperature control.

Every 50 hours, the computer initiates L-I-V in-place tests on each laser. During the test, the relative temperature is held to ±0.05°C of the nominal aging temperature. The absolute temperature is known to within 0.5°C. The current on each device is decreased in increments of ≈0.4 mA from its operating current down to zero. At each current level, the light from the fiber (L_F), light on the rear-face detector (L_R), current, and voltage are measured. From this set of measurements are derived the IdV/dI, I^2d^2V/dI^2, dL/dI, d^2L/dI^2 curves for each device

at a particular temperature and time. The derivatives for light are computed only for the rear-face light. The data are analyzed and stored by the computer to give information about threshold, modulation current, nonradiative or leakage current, kinks in the LI curve that may indicate high-order transverse modes, diode resistance, and spontaneous emission at threshold that can affect the extinction ratio. The fiber-to-rear-face power ratio gives the fiber coupling efficiency (η_c).

The test facility can also provide tests relevant to possible operational failures of the laser. For instance, the optical spectrum of the light launched into the single-mode fiber is measured. The data are digitized and stored in the computer. Spectral splitting, mode broadening, or excitation of an additional family of longitudinal modes, though rare, have been detected from these data.

The existence of self-induced pulsations or excessive noise is monitored by analyzing the microwave spectrum of the laser output from 100 kHz to 3.5 GHz. Depending on the frequency and intensity of the microwave noise, the error rate in the optical link could increase above the maximum allowed limit of 10^{-9}/s.

An auxiliary package pulse aging facility consists of 120 positions capable of modulating the laser using pseudorandom pulses at data rates to 300 Mb/s. In addition, all the tests described for the CW facility are made for the pulse system as well. The controller is a separate desk top computer separate from the one used for the CW facility.

The main purpose of pulse aging is to compare the pulsed aging rates with the CW aging rates. In the qualification program about 10 percent of the lasers to be aged will be aged under pulsed conditions to assure that no degradation mechanism unique to pulsed operation exists.

A.2 Performance

Because of the high temperature coefficient of laser current at a fixed light output, temperature variations are a major contributor to variations in laser current. In our aging racks, we have achieved a day-to-day control of the relative temperature of ±0.05°C during measurements of laser characteristics. We have measured the long-term standard deviation of the laser current (I_5) at a fixed light output of 5 mW and determined it to be 0.10 mA. This value, (I_5), is included with the short-term and long-term measurement accuracies of important laser parameters in Table II. Threshold current (I_{th}), which is derived digitally from the maximum in d^2L/dI^2, has a long-term standard deviation of 0.20 mA. The mean value for the short-term standard deviation for I_5 was 5 microamperes, a factor of 20 less than the long-term value. This result is consistent with our original estimate that

Table II—Measurement accuracy

Parameter	Units	Short-Term Standard Deviation	Long-Term Standard Deviation
I_{th}	mA	0.1	0.20
$I_{5(mW)}$	mA	0.005	0.10
dI_{th}/dt	mA/kh	—	0.04
dI_5/dt	mA/kh	—	0.02
$\eta_c = L_F/L_R{}^*$	dB	0.01	—
T	°C	0.008	0.036

* L_F is the light from the fiber and L_R is the light from the rear face of the laser.

temperature control would dominate the long-term stability measurements of threshold and drive current on these lasers.

The aging rates of threshold and 5-mW operating currents are two of the most important parameters that we study. If the linear law of eq. (6) in the main text is assumed, the Gaussian statistics for deriving the aging rate from estimates of measurement variance can be modeled. The results depend on the frequency and distribution of the measurements over a given time interval. For the 0.1-mA, long-term standard deviation for I_5 given in Table II and 60 measurements made uniformly over 3000 hours, the uncertainty in the aging rate would be 0.02 mA/kh. We also get this value for the standard deviation of the aging rate, R for I_5 in Table II, which is the average of many least squares fits to $I(t)/I_o = 1 + Rt$. The data used were from devices with 2000 to 4000 hours of stable aging characteristics at 12 and 50°C and no obvious knee characteristic.

We will be able, assuming that the linear model continues to provide an upper bound to the extrapolated operating current, to extrapolate with a few milliamperes of uncertainty the operating current to 200,000 hours at any temperature with 3000 hours of data. In particular, we can certify each laser package for current stability by operating it at 10°C for 3000 hours. In addition, each laser chip will have seen a high-temperature aging during a purge that is equivalent to tens of thousands of hours of system operation at 10°C. After packaging, the laser will again see the temperature accelerated equivalent of at least one million system hours during a 60°C part of the package certification.

Our situation in fiber coupling measurements and measurements of the stability of the test apparatus is not yet firmly established. It is desirable that we have an engineering safety factor on the fiber coupling stability verifiable on individual lasers by temperature acceleration. The uncertainty in documenting this stability is related to the early stage of this phase of the program. The main concern in packaged lasers is related to the potential for mechanical relaxation phenomena causing changes in the fiber coupling. Our observation to date has been that the rate of change of fiber coupling decreases with time; i.e.,

the rate is sublinear. Therefore, our initial measurements will provide an upper bound on the linear decoupling rate.

Our short-term measurement error for fiber coupling is given in Table II. The short-term value is small enough to make 3,000-hour coupling projections to 200,000 hours with only a few percent uncertainty in the coupling. As in the case of current stability, the assumption is made that a linear fit is an upper bound on the aging rate. This assumption will be checked on individual lasers.

We do not yet have data on the long-term standard deviation for fiber coupling. The temperature coefficient is 0.02 dB/°C, three to four times smaller than the laser current temperature coefficient. Therefore, we expect the long-term measurement error to be dominated by the repeatability of the alignment of our computer-controlled optical multiplexer. The short-term measurement error, 0.01 dB, is a measure of this repeatability.

In conclusion, at AT&T Bell Laboratories we have built and tested aging facilities that meet the criteria outlined at the beginning of this section. Assuming a linear model, we can project, with a few milliamperes uncertainty, laser 5-mW operating current stability to 200,000 hours or 24 years with 3000 hours of data. This can be done at any of our operating temperatures, including the 10°C system operating temperature. Our short-term fiber coupling efficiency measurements indicate similar accuracies and projection test times for fiber coupling efficiency measurements. We expect our long-term accuracy to be compatible with the required 200,000-hour extrapolations.

APPENDIX B

Design of the Reliability Experiments

All lasers are aged in packaged form. Two types of packages are used. The first is the so-called multimode package, which uses a Ge detector to monitor the power output from the back face of the laser. This package also uses a multimode fiber to couple the power output from the front face of the laser. The second type of package is the so-called single-mode package, in which the output from the front face is coupled into a single-mode fiber. It uses an InGaAs PIN detector to monitor the power output from the back face of the laser. The multimode package is easier to make and couples more effectively into high-order transverse modes. Therefore, the multimode package is a useful vehicle for some diagnostic purposes. On the other hand, the single-mode package is ultimately the final design configuration that provides information about the stability of the coupling of the fiber to the laser. Out of a total of 262 lasers tested in this reliability program, 129 are in single-mode packages and 133 are in multimode packages.

Table III—1.3-μm laser and slice allocation

Slice Number	Group 1 (Step)		Group 2 (Step)		Group 3 (10°C)		Group 4 (60°C)		Group 5 (40°C)		Group 6 (Step)		Group 7 (Step)		Group 8 (30°C)	
	M	S	M	S	M	S	M	S	M	S	M	S	M	S	M	S
A	1															
B	8		1													
C	5		1													
D	6		6												2	3
E			10			5						4		1		
F						2	5			1						2
G					4	2					3	1				
H					7		6			9						4
I							10		5	3						
J											2	2	6			
K											4	1	5			
L							5	2	2	3						4
M							6		6							
N							6			5						4
O				2					6							
P												5		3	6	
Q												8		7		
R											4	3		5		
S														6		
T					5									4		
U					4	1										
V						3								3		
W						3										
X					3	1										

* M—Multimode. S—Single mode.

The 262 lasers being tested are divided into eight groups listed in Table III. Four of the groups (1, 2, 6, and 7) are step stressed in temperature and power. The remaining four groups (3, 4, 5, and 8) are isothermally tested at 5 mW and the temperatures indicated in Table III. Groups 1 and 2 are stepped in temperature and power according to the sequence shown schematically in Fig. 1. Groups 6 and 7 are step-temperature stressed at 5 mW in the sequence shown in Fig. 2. In addition, to the extent possible, Groups 6 and 7 contain lasers obtained from split-lot slices, as indicated in Table III. This is done to provide a degree of homogeneity in the results between the two groups. By the same token, Groups 4 and 5, isothermally tested at 60 and 40°C, respectively, also involve split-lot slices. This attempt to homogenize the isothermal groups is mandatory for a proper interpretation of the results. The reasons for this are obvious. Within each slice, the laser-to-laser variability in degradation rate is about an order of magnitude. In addition, slice-to-slice variability can be two orders of magnitude.

Hence, to measure the temperature acceleration of degradation between 40 and 60°C, which is expected to be one order of magnitude, is very difficult if each individual population has a spread of three orders of magnitude. Hence, to compensate for this large variance within each test group, an attempt to homogenize comparable groups is made by selecting devices from the same slices. This provides by no means a perfect solution, but is an attempt to ameliorate an otherwise intolerably large variation.

Groups 3 and 8 of Table III are also isothermally tested at 10 and 30°C, respectively. Group 8 attempts to match devices from Groups 4 and 5. The choice of 30°C testing provides a check of the Arrhenius relation as obtained from isothermal testing.[4] The testing time at 30°C, required to establish valid long-term trends, is unfortunately very long, $\geq 10^4$ hours. Hence, at the present time, the 30°C testing data are too preliminary to provide accurate quantitative information. On the other hand, the 10°C isothermal group contains samples from various slices that are being tested at high temperatures, either step stressed or isothermal. The utility of the 10°C test group is to provide, in the long term, a correlation between high-temperature test data and the system operating temperature. In addition, it is a necessary experiment that would uncover any laser performance malfunctions that may be unique to 10°C operation. So far, no fundamental problems have been encountered that are peculiar to 10°C operation.

Finally, it is seen that Table III lists lasers obtained from 24 slices. This represents a fairly wide cross section of material made from early development (e.g., slices A, B, C, and D) to slices made under strict production control environment (e.g., slices T, U, V, W, and X). The basic laser design is almost unchanged throughout. The fact that a good performance record can be demonstrated for lasers, going from development to production, attests to the basic viability of this device. It should also be emphasized that the number of lasers associated with each slice in Table III does not represent the total number from each slice. The lasers used in this reliability program represent only a portion of the total devices. Other lasers are used in diagnostic evaluation, laser-transmitter evaluation, purge development, other experiments, etc.

AUTHORS

Thomas F. Eltringham, A.A.S. (Electronics Engineering Technology), 1967, DeVry Institute of Technology; AT&T Bell Laboratories, 1967—. After joining AT&T Bell Laboratories, Mr. Eltringham was involved in the design of test equipment for phase-array radars. He then worked on the design, testing, and reliability studies of light-emitting diodes and lasers. He is currently involved in laser reliability studies for undersea applications.

Phillip E. Fraley, B.S. and M.S. (Engineering Physics), 1962, Ph.D. (Physics), 1966, The Ohio State University; Aerospace Research Laboratory (USAF), 1966–1969; AT&T Bell Laboratories, 1969—. At AT&T Bell Laboratories, Mr. Fraley has worked on the design and characterization of more than 60 Light-Emitting Diodes (LEDs). He then worked on the reliability and characterization of GaAlAs lasers. Mr. Fraley's recent work has been the use of specialized computer hardware to obtain precise measurements of the optoelectronic parameters needed to quantify InGaAsP laser stability.

Basil W. Hakki, B.S.E.E., 1957, M.S., 1958, Ph.D., 1960 (Electrical Engineering), University of Illinois; AT&T Bell Laboratories, 1963—. After joining AT&T Laboratories, Mr. Hakki was involved in the design and analysis of GaAs two-valley microwave oscillators. He then worked with III–V LEDs and lasers. His work on GaAs double heterostructure injection lasers covered many aspects of the laser, including reliability, device design for CW and high-power operation, and device physics. He is currently involved in the study of quaternary lasers.

Implementation of the Proposed Reliability Assurance Strategy for an InGaAsP/InP, Planar Mesa, Buried Heterostructure Laser Operating at 1.3 μm for Use in a Submarine Cable

By F. R. NASH, W. J. SUNDBURG, R. L. HARTMAN,
J. R. PAWLIK, D. A. ACKERMAN, N. K. DUTTA, and
R. W. DIXON*

(Manuscript received March 5, 1984)

We discuss the implementation of a strategy designed to provide laser-light-emitting reliability assurance for 1.3-μm InGaAsP/InP lasers of the planar mesa, buried heterostructure type for use in a submarine cable application. The testing regimes include initial characterization (cosmetic and light-current curve inspection), passive aging (elevated temperatures [85 to 175°C] without bias, with and without humidity [≤85-percent relative humidity]), overstress active aging (high temperatures [150°C], high currents [250 mAdc]), and long-term rate-monitoring active aging (elevated temperature [60°C] burn-in [3 mW/facet]). Overstress testing is designed to compel a timely (~10^2-hour) identification of premature failures, due to modes of degradation other than the long-term ultimately controlling wear-out mode, and to stabilize transient modes. To identify premature failures of the wear-out type, survivors of overstressing are subjected to rate monitoring in which wear-out degradation rates, established in a reasonable time (~10^3 hours), may be sorted. The principal results of the important overstress aging were the detection of an initially occurring saturable degradation mode, present to some extent in most lasers, and a regimen to force its rapid stabilization, so that it would not obscure determination of the activation energy of the wear-out mode. With a credibly determined value for the latter, it was deterministically inferred from

---

* Authors are employees of AT&T Bell Laboratories.

rate-monitoring results that the light-emitting reliability of the screened lasers at ocean bottom temperatures (10°C) is more than adequate to meet the system design lifetime of 25 years.

I. INTRODUCTION

One promising candidate for the lightwave submarine cable transmitter source is the InGaAsP/InP etched planar mesa, buried heterostructure laser operating at 1.3 μm.[1-4] (A light-emitting-reliability advantage in operating at 1.3 μm had previously been shown[5] by the finding that the average room temperature lifetime of Light-Emitting Diodes [LEDs] fabricated from III-V ternary alloys of the InGaAsP family increases exponentially with decreasing band gap. An LED operating at 1.3 μm would thus outlast one at 0.80 μm by a factor $\approx 10^5$. This was confirmed by the demonstration of room temperature lifetimes $\approx 10^7$ hours for AlGaAs LEDs[6] and similar lifetimes $\approx 10^{12}$ hours for InGaAsP LEDs.[7]) In this paper we discuss the *initial* (cosmetic inspection and light-current characterization), *passive aging* (elevated temperatures [85 to 175°C] without bias, with and without humidity [≤85-percent relative humidity]), overstress *active aging* (high temperatures [150°C], high currents [250 mAdc]), and long-term rate-monitoring active aging (elevated temperature [60°C] burn-in [3 mW/facet]) screening tests that have been applied to lasers of this type to assure light-emitting reliability. Together, these tests constitute elements* of a reliability strategy that has been discussed.[8,9] The objectives are (1) the detection of flaws that have reliability implications and that can be corrected by timely changes in laser design or fabrication; (2) the stabilization of temporally saturable annealing and degradation mechanisms ("drift") that could prevent the reliable determination of both the thermal activation energy and the gradual degradation rates associated with the long-term ("wear-out") mode of failure of the lasers; and (3) the identification of those lasers that are likely to suffer premature failure, i.e., failure prior to the 25-year design lifetime of the subcable system operating at 10°C. Emphasis is placed on two active aging (i.e., forward bias operation) regimes. The *overstress* regime employs high temperatures (60 and 150°C) and high currents (250 mAdc) for relatively short periods of time ($\sim 10^2$ hours). The object is to compel a timely identification of those lasers likely to be premature failures due to modes of degradation other than the

---

* It should be understood that this paper represents the *first* attempt at establishing a reliability strategy for one particular laser structure. The elements discussed show promise. Additional work to optimize the strategy may result in the addition, replacement, or refinement of elements.

fundamental, long-term, tolerably slow-acting wear-out mode and to stabilize transient modes. To identify those premature failures that are early wear-out failures, the survivors of the overstress tests are subjected to the *rate-monitoring* active aging regime in which are sorted the wear-out degradation rates, established in an extended duration burn-in at a lasing temperature sufficiently high so that unambiguous degradation occurs in a reasonable time ($\sim 10^3$ hours).

The principal results of the overstress testing were the detection of an initially occurring saturable degradation mode, present to some extent in most lasers, and a scheme to force its rapid stabilization. With this accomplished, credible values of the activation energy and degradation rates for the slower-acting wear-out mode can be determined. Thus, from the rate-monitoring results, it was deterministically inferred that the *light-emitting reliability* of the screened lasers is more than adequate for subcable applications. (Later studies will address the transmitter or *electro-optic reliability* [e.g., absence, with aging, of beam wander, spatial mode switching, light-current abnormalities, extinction ratio deterioration, and relative motion of the laser and optical fiber] of the transmitter.)

A word now about nomenclature. The reliability assurance results[8,9] that we propose to produce are called STP (Stabilization, Truncation, Purge). The *purge* consists of any scheme (e.g., optical inspection, light-current characterizations, overstress aging using high temperatures, currents, optical power, etc.) by which actual or potential premature non-wear-out failures are eliminated from the population. The overstress regime of the purge will also produce *stabilization* of initial saturable transient modes of degradation that appear to exist in most lasers. Rate monitoring of long-term degradation permits *truncation*, i.e., an elimination of early wear-out failures, those which would occur prior to the desired system lifetime.

II. INITIAL SCREENS (PURGE)

Initial screening involves tests that do not generally produce degradation. They may be passive (e.g., optical or Scanning Electron Microscope [SEM] examination) or active (e.g., light-current characterization).

2.1 Optical inspection

Microscopic inspection to check for missing, broken, or unattached wires; cracked, misaligned, facet-contaminated, or misbonded chips; or handling damage has obvious value. Our primary discovery was the existence of torn fragments of Au (from and still attached to the Cr-Au p-metallization and produced during facet cleaving) on the mirror

~1 μm Cr–Au p-METALLIZATION

CLEAVED
MIRROR FACET

4 μm
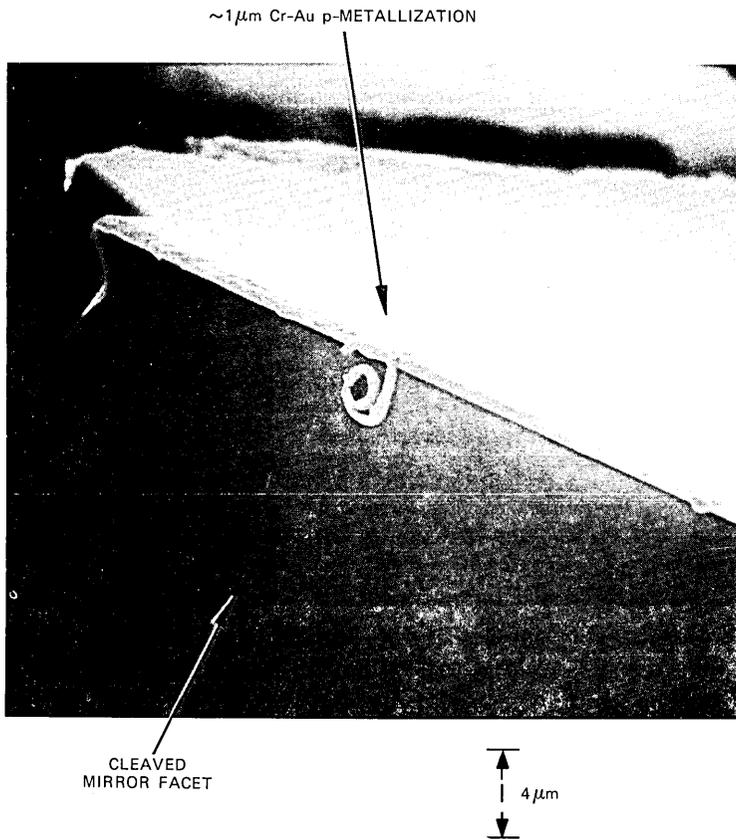
Fig. 1—SEM photograph of a fragment of Au, torn from the p-contact to which it is still attached, which lies on a cleaved facet.

surfaces, across one or more of the epilayers, and on occasion within the region of the mesa. Figure 1 is an SEM photograph of one such case.

There are three ways in which such Au fragments could be inimical to laser reliability: (1) Light blockage—in the course of system installation, long-term mechanical relaxation, or mechanical vibration tests, fragments within the mesa might move to partially obstruct the optical output. (2) Electrical shorting—the recovery of lasing operation after sudden termination, effected by mechanical removal of Au fragments from the mirror of a laser that exhibited no preceding or subsequent signs of degradation, suggested that the torn metal could externally short the p-n junction. (3) Au in-diffusion—Dark-Spot Defect (DSD) formation is a significant mechanism for degradation in InGaAsP/InP LEDs[7,10-15] and lasers,[16-18] and appears to be caused in some cases by the in-migration of Au from the p-contact.[12,13,15]

The overhanging Au problem is an example of the purge used to detect a defect that could cause premature failure and that is avoidable through a relatively simple design change (e.g., thinner or patterned Au).

2.2 L-I-V characterization at 30, 50, and 70°C, pulsed and dc

Initial pulsed (in pulsed operation the junction of the laser remains at the stud or ambient temperature) and dc electrical and optical characterization, performed at ambient temperatures of 30, 50, and 70°C, consists of plots of the following quantities as functions of current (I): light (L), dL/dI,[19] voltage (V), IdV/dI and I^2d^2V/dI^2. For purposes of clarity, we exhibit only the first two of these quantities in Fig. 2 for an acceptable laser. The sublinearity in the Continuous Wave (CW) L-I curves commences at threshold and is more pronounced than in the pulsed curves. (The implications of sublinearity will be considered in more detail in Section 4.1.) Figure 3 represents a laser with unacceptable CW L-I curves. Temperature-dependent anomalies (kinks) in the L-I curves are more apparent in the dL/dI curves. The kinks may be the result of beam wander[20] or a switch to a higher-order mode, which often results in a split mode spectrum.[21-23] Both effects can be caused by an inadvertently too large lateral active region width. Both effects can significantly reduce the power coupled into a single-mode fiber.

III. PASSIVE AGING SCREENS (PURGE)

Passive aging involves tests designed to produce degradation without current injection.

3.1 Temperature humidity

To test the integrity of the stud, standoffs, and solder, we used an 85°C 85-percent Relative Humidity (RH) ambient (the partial pressure of water vapor was 383 Torr, the volume density of water molecules was 10^{19} cm^{-3}, and the fractional concentration of water molecules was 500,000 ppm). Since the facets of the lasers were not dielectrically coated, we chose not to bias the lasers in order to avoid the risk of a lasing-induced facet degradation,[24-26] which might have masked other degradation mechanisms whose presence it would be important to detect.* It was not expected that the Pb/Sn solder used in bonding the laser (p-side up) would degrade since it appears that eutectic alloy

---

* In Section 7.2 we discuss the role of facet erosion under forward bias and conclude, in accordance with Refs. 25 and 26, that it is inconsequential for InGaAsP/InP lasers operated in the environment of interest.
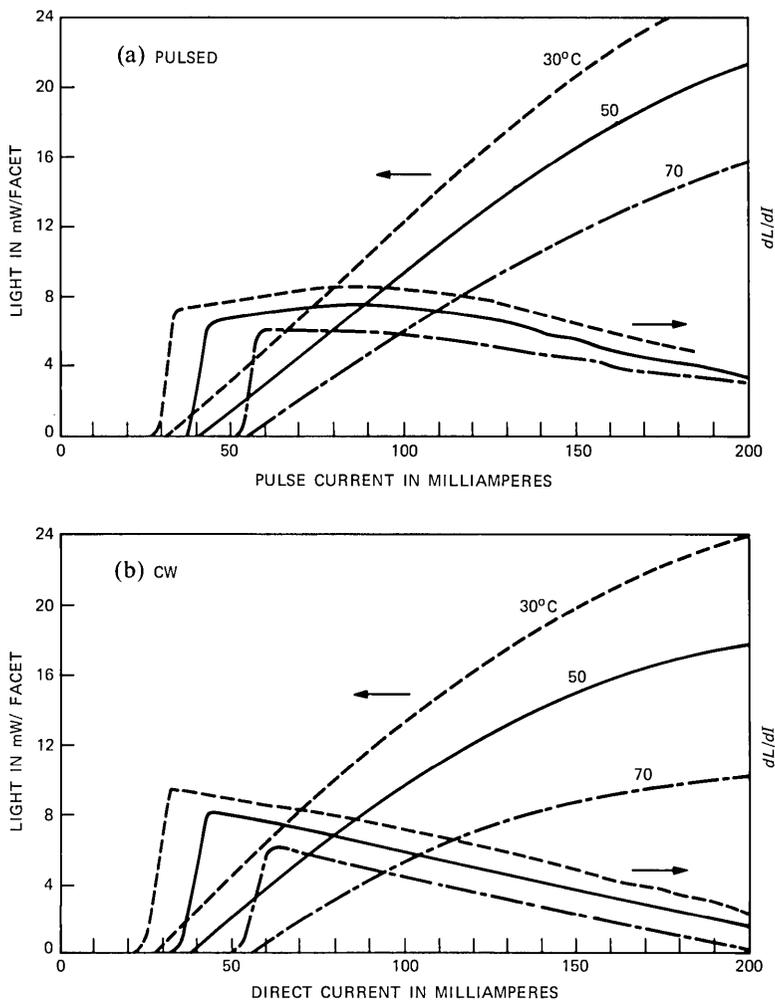
Fig. 2—L and dL/dI versus I as a function of temperature for a typical laser, (a) pulsed and (b) CW. The dL/dI-I curves in CW operation exhibit a negative slope, which reflects the rollover tendency of the corresponding L-I curves.

solders (e.g., Au/Sn* and Au/Ge) remain stable[28-30] for laser operation at 70°C for 10^3 to 10^4 hours, in the apparent absence of any attempt to minimize environmental humidity.

Four lasers were optically examined several times in the course of ≈ 2300 hours of aging without bias. Although two of the lasers operated

---

* A recent study (Ref. 27), however, has revealed Sn-whisker growth at the surface of Au/Sn bonding solder.
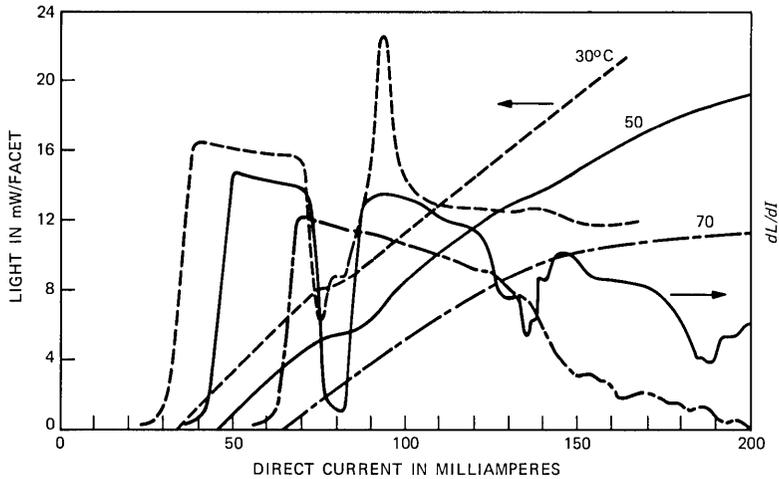
Fig. 3—*L-I* and *dL/dI-I* curves for a laser that exhibits temperature-dependent anomalies (kinks).

with only slightly altered *L-I* characteristics at the termination of aging, three corrosion products of major proportions were present on all lasers and studs. One corrosion product was present in the form of a thermally induced, nonuniform, multicolored (optically thick, ~1000-angstroms) oxide covering the laser facets; at the start of the aging, the facets had been spotlessly clean. A second corrosion product was a grayish powdery deposit on the top and sides of the standoffs, lasers, subcarriers, and studs. Using the energy dispersive X-ray attachment of an SEM,[31] it was shown that a lead oxide was predominant. A third corrosion product was associated with the presence of a steel weld ring (to which a hermetic can is eventually affixed) on the top of the bolt head part of the stud. Although the entire stud is Au-plated, the plating at the weld ring was breached. After ≈2300 hours, a red-brown iron oxide[31] (rust) was observed on top of the bolt head and on nearby surfaces of the stud remote from the ring.

The device-threatening formations of oxides of InP, Pb, and Fe can be substantially suppressed if the water vapor content within the hermetic can is made low. The water vapor content of the ambient outside the can should also be kept low to prevent a breach of the hermetic seal, from without, caused by an inadvertent exposure of some part of the steel ring during the welding. Operation of the lasers at ocean bottom temperatures (~10°C) will also assist in retarding oxide formation.

3.2 High temperature

This regime emphasized storage aging at 175°C, near the melting
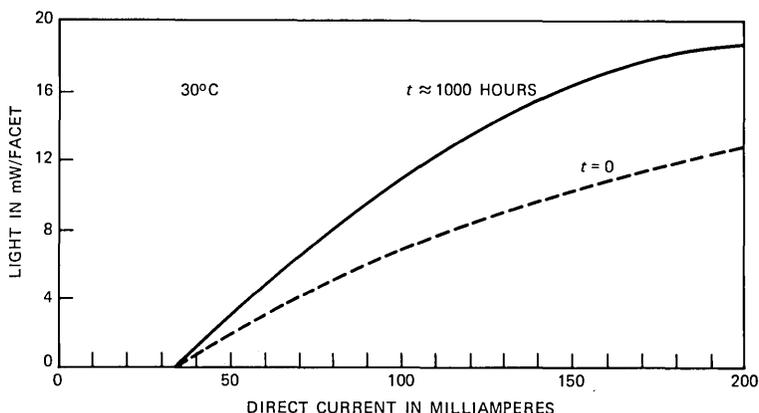
Fig. 4—Change in the 30°C CW L-I curve produced by unbiased storage aging at 175°C for 1 kh.

temperature $(183°C)^{32}$ of the Pb/Sn solder; the test chamber had an opening to the laboratory ambient.* It was not expected that significant deterioration of the Pb/Sn would occur because similar storage-aging experiments had apparently confirmed the stability of eutectic alloy solders for LEDs (Au/Ge, 230°C, 2000 hours)[7] and lasers (Au/Sn, 100°C, 2000 hours).[33]

No oxide growth was observed on any laser or stud that had been aged at 175°C for ≈4000 hours and periodically examined.† The lasing thresholds at room temperature remained unaltered. Several lasers exhibited increases in slope efficiency, however, as typified by Fig. 4. No change after 1000 hours was observed; it is not known when the improvement occurred prior to the first recharacterization at 1000 hours. Annealing effects similar to that shown in Fig. 4, but of smaller magnitude, were noted in the course of active aging; they occurred whenever forward bias was removed from a device at some elevated temperature (e.g., 60°C).

IV. ACTIVE AGING SCREENS (STABILIZATION, PURGE, TRUNCATION)

4.1 High-current L-I characterization—the role of shunt currents

Prior to forward bias aging at high currents, it is necessary to characterize the L-I curves at high currents. At elevated temperatures

---

* The partial pressure of water vapor was 15.5 Torr (the same within and without the chamber); within, the relative humidity was ≈0.2 percent, the volume density of water molecules was 3.3×10^{17} cm^{-3}, and the fractional concentration of water molecules was ≈20,000 ppm. Relative to the 85°C 85-percent RH ambient, there were ≈25 times fewer water molecules.

† In Section 7.3 we shall discuss the effect of storage aging on the possible in-migration of Au from the p-contact.
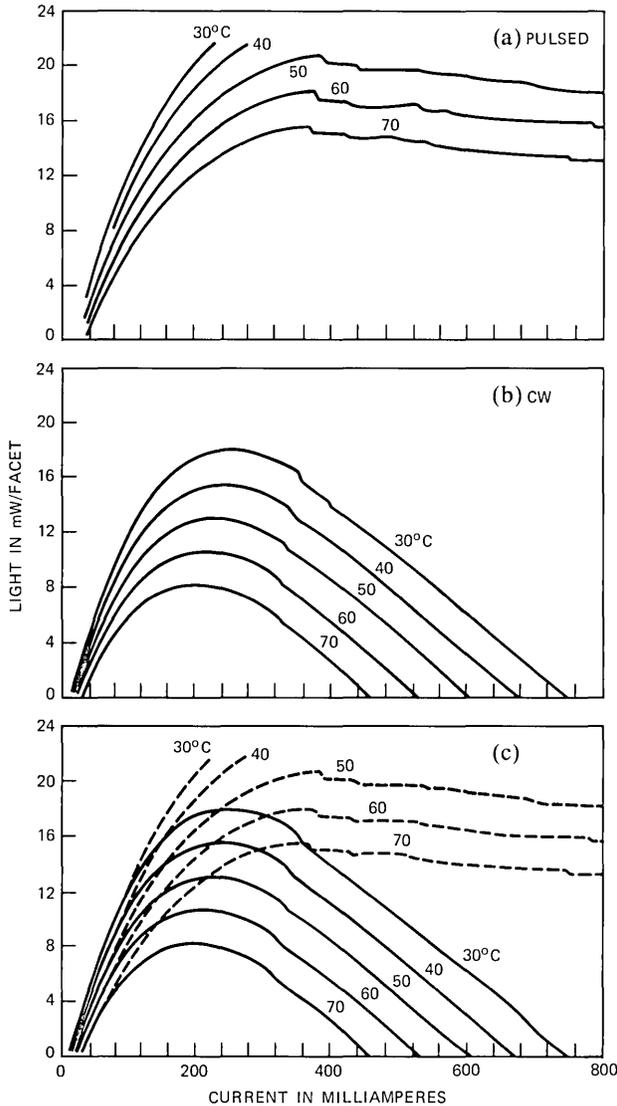
Fig. 5—(a) Pulsed, (b) CW, and (c) composite L-I curves as a function of temperature.

and/or at high currents, a major fraction of the injection current is shunted around the active region. This peculiarity of those lasers that employ reverse-biased junctions to confine current must be taken into account in the design of high stress (current, temperature, optical power) active aging screens.

A typical set of high-current L-I curves is shown in Fig. 5 for pulsed

and dc operation at several different ambient temperatures. The pulsed *L-I* curves exhibit "saturation" behavior, i.e., beyond a critical current (~400 mA for the particular laser), the light output remains essentially constant. The dc *L-I* curves exhibit "rollover" behavior and an upper, as well as a lower, threshold.

The sublinearities present in the pulsed *L-I* curves are due to shunt (leakage) paths, which divert an increasingly larger fraction of the terminal current away from the lasing region, as either temperature or current is increased. In pulsed operation (no heating caused by the current), there appears to be a critical current, independent of ambient temperature, beyond which no additional current can be injected into the lasing region (see Fig. 5a). (In an ideal laser, threshold represents that current above which all additional current is converted into stimulated emission in the lasing modes. The critical current in Fig. 5a is also a threshold current, not for lasing, but for current diversion.)

To understand the rollover in the dc *L-I* curves (see Fig. 5b), it is helpful to superimpose a family of pulsed *L-I* curves (see Fig. 5c), which can be used as thermometers. The 70°C pulsed *L-I* curve crosses the 30°C dc *L-I* curve at ≈360 mA. At the crossover point, the pulsed and dc curves correspond to identical light outputs and currents, so that the active region in dc operation must be 40°C above the ambient or stud temperature. (The validity of this technique requires the assumption that the distribution of currents between the active and shunt paths in the nonheating pulsed regimes is not frequency dependent.) We may use a standard expression for the thermal impedance (R_{th}),

$$R_{th} = \Delta T[°C]/VI[W],\qquad(1)$$

where ΔT is the temperature rise of the active region above the stud temperature, and the voltage-current product is the electrical input power dissipation, which is responsible for ΔT. For the specific example, $\Delta T = 40°C$, $V = 2.0$ volts, and $I = 0.36$ ampere, it is computed that $R_{th} = 56°C/W$. This value is typical[2] for lasers of this type bonded p-up.

The 30°C, dc *L-I* curve is initially coincident with the 30°C pulsed *L-I* curve (see Fig. 5c), but as the current is increased, the active region in dc operation heats up, and the dc *L-I* curve can be seen to progressively cross the 40, 50, 60, and 70°C *L-I* curves of pulsed operation. Rollover in dc operation, caused by internal (ohmic or nonradiative) heating, is then seen as a consequence of having the operating point in the dc *L-I* space pass from one pulsed *L-I* curve to another as the current is increased. Since the maximum lasing power output in pulsed operation decreases with increasing temperature, the dc operating point will roll over and proceed toward the upper lasing threshold. For

the 30°C dc $L\text{-}I$ curve of Fig. 5, the upper lasing threshold occurs at $I = 0.74$ ampere, so by using (1), $R_{th} = 56°C/W$, and $V = 2.0$ volts, it is computed that $\Delta T = 83°C$. Thus, the ambient temperature at which this device could no longer be pulse-operated as a laser is $83 + 30 = 113°C$, which is in good agreement with high-temperature pulsed measurements on similar lasers. Although there is some current flowing into the active region above the upper threshold that produces $\sim 100\ \mu W$ output, the 113°C temperature is viewed as that temperature for which lasing threshold can never be attained because of the dominance of the shunt path. (A thermal runaway of threshold might be expected to occur because of the well-known empirical relationship, $I_{th} \propto \exp(T/T_0)$, where for the present lasers, $T_0 = 70K$. For three reasons, this effect is not believed to be controlling the rollover behavior in Fig. 5b: (1) According to calculations, the output power would fall precipitously beyond the point of rollover because of the exponential function, and this is not in agreement with the shape of Fig. 5b. (2) If shunting were a minor effect, there should have been significantly more spontaneous output (~ 3 mW) at the upper threshold (740 mA) of the 30°C dc $L\text{-}I$ curve of Fig. 5b, since LED studies[14] have shown that InGaAsP does not become nonradiative at $\sim 100°C$; no light output in excess of ≈ 0.1 mW is apparent beyond the upper threshold of Fig. 5b. (3) Using a convenient relationship between the dc and pulsed thresholds, $I_{th}^{dc} = I_{th}^{pul} \exp(\Delta T/T_0)$, where ΔT is due to current heating, $I_{th}^{pul} = 35$ mA, and $\Delta T = 83°C$ in dc operation at $I = 740$ mA in a 30°C ambient, it is calculated that I_{th}^{dc} (113°C) = 114 mA; the device should have lased at a junction temperature equal to 113°C, in the absence of shunting, but it did not. Thus thermal runaway is not dominant.)

It is known[1,3,34] that leakage or shunting currents are significant in these devices. Figure 6 is a schematic of the etched mesa, buried heterostructure laser showing two possible shunting paths for the terminal current, in addition to the current through the active region. The electrical characteristics of such a laser with a diode shunt path have been calculated.[35] The $L\text{-}I$ signatures for pulsed operation have been calculated for the pnpn shunt alone,[36,37] and for the pnpn and diode paths simultaneously present.[38] Using credible values of the relevant parameters, the latter study[38] permits the conclusion that the dominant shunt current is I_{pnpn}. Calculated representative pulsed $L\text{-}I$ curves for several ambient temperatures are shown in Fig. 7. A "knee" is apparent, suggesting a light output saturation such as that shown in Fig. 5a. This model, however, does not show leakage at threshold, in disagreement with experiments.

In addition to the *gradual* increases in leakage current corresponding to the increasing sublinearity in the $L\text{-}I$ curves of Fig. 5, there also
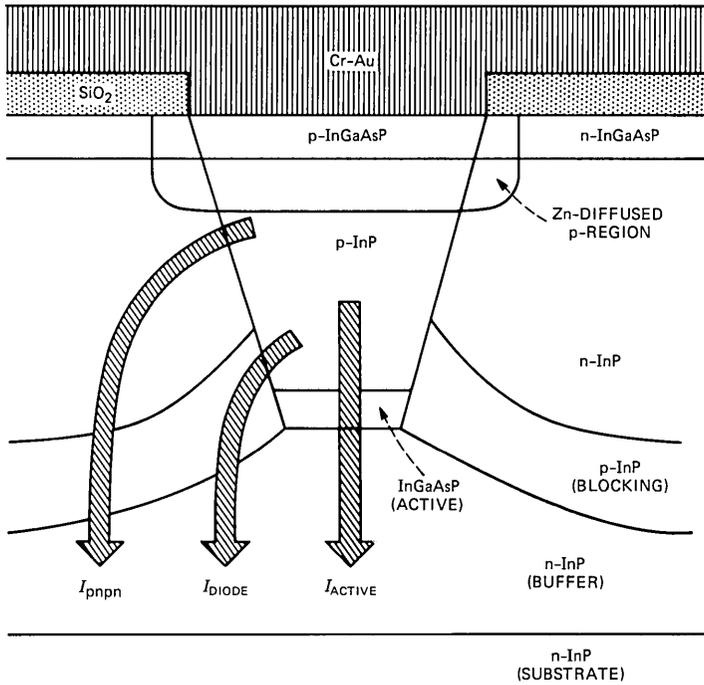
Fig. 6—Schematic cross section of the planar etched mesa, buried heterostructure laser with several current paths.
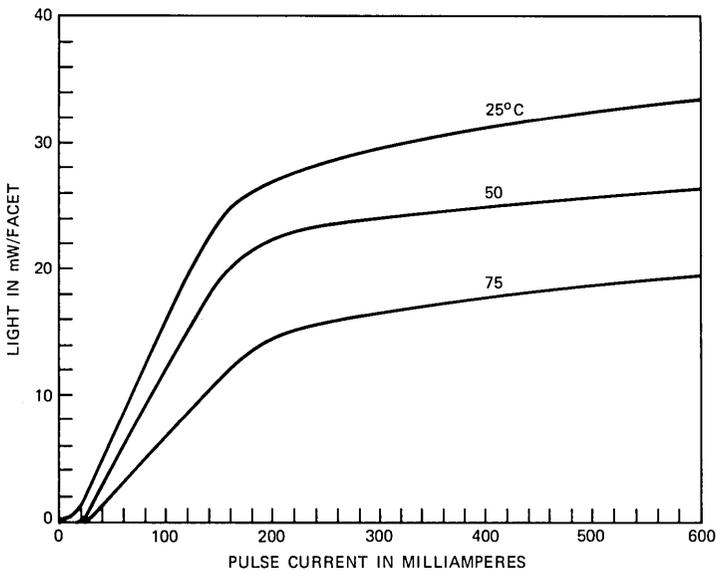
Fig. 7—Calculated pulsed L-I curves assuming significant shunt current flow in the pnpn path of Fig. 6.

appear *abrupt* decreases in output for small positive increments in current in both pulsed and dc operation (see Fig. 5). These sudden decreases on occasion may be large ($\Delta L \sim -10$ mW/facet); they invariably and very frequently occur beyond the critical current for output power saturation, they have been attributed[1,34] to sudden increases in I_{pnpn}, and the current at which they occur often does not decrease with accelerated aging degradation, so their existence is not automatically grounds for laser rejection.

The existence of shunt currents has been confirmed by the direct observations of spontaneous emission from lateral regions outside the active region, either from an n-side "window"[4] or facet viewing.[39] The dL/dI curves characterizing dc operation indicate that sublinearity in the *L-I* curves is present at lasing threshold and above (see Fig. 2b). This is true for all lasers of this type examined in the temperature range 0 to 70°C. An advantage[1] in the existence of dc-rollover (i.e., current shunting) is that catastrophic optical damage of the laser facets[40] may not occur even though the lasers are operated repeatedly up to 30 times threshold. Since unexpected current surges are a likely cause of random failures,[9] the built-in surge protection is important in a high-reliability application. *Short-duration, high-current* pulses might, however, remain a problem (e.g., compare Fig. 5a and b).

4.2 Initial elevated temperature burn-in

All of the lasers used in our tests were the survivors of two elevated temperature burn-in tests conducted prior to our receipt of the devices: (1) 60°C, 5 mW/facet, 24 hours, $\Delta I_{op}/I_{op} \leq 1.5$ percent; and (2) 60°C, 5 mW/facet, 100 hours, $\Delta I_{op}/I_{op} \leq 5$ percent, where ΔI_{op} and I_{op} are the increase in operating current and initial value of operating current to produce a given output. Burn-in tests (operation at normal outputs and elevated temperatures for a short duration, $\sim 10^2$ hours) are useful for eliminating some so-called infant failures. Potential low thermal activation energy modes of failure may not, however, be identified in this kind of screening test.[8,9] We shall show that due to the existence of an initially occurring saturable mode of degradation, which exists in some degree in virtually all lasers, and which typically can escape detection in burn-in tests because it has an incubation period, the conventional elevated temperature burn-in was found to be inadequate to assure reliability of its survivors. The purge (understood, hereinafter, to mean the overstress regimes) that was imposed on the burn-in survivors was essential to the success of the undertaking.

4.3 Degradation accelerants and driving mechanisms

Discussions of screening strategies[8,9] have emphasized the concepts of the *harshness*, *selectivity*, and *tailoring* of the accelerants used to

detect the presence of premature non-wear-out failure and transient modes. Initial device characterization and an understanding of how the device operates is crucial in giving specificity to these concepts. The laser is a forward bias device. Potential accelerants for the light-emitting degradation of lasers under forward bias are temperature, injection current, optical power, and humidity.

In the absence of forward bias, *humidity* is both a driving mechanism for, and an accelerant of, degradation; it is not, however, selective, because it treats all lasers alike (see Section 3.1). For a laser operated in forward bias without protective facet coatings, humidity in the testing ambient will also nonselectively promote and accelerate degradation due to the formation of an oxide on the emissive portion of the mirrors.[24,25,41] The degradation may take many forms (e.g., reduce the output at a fixed current, redirect the output beam, promote electro-optic deficiencies such as pulsations and light jumps),[41] but it should have approximately the same effect on all lasers, i.e., it makes no distinction between lasers that are inherently reliable and those that are not. The lasers used in our tests had uncoated facets. Our initial test results showed that no light-emitting degradation that we detected was due to facet oxidation (see Section 7.2). This is in accord with recent studies[25,26,31] that have concluded that InGaAsP/InP lasers of the etched mesa, buried heterostructure type can operate at elevated temperatures ($\sim70°C$), in laboratory ambients (~30-percent RH), with outputs ≈5 mW/facet, without any appreciable degradation due to lasing-induced facet erosion for times that are more than an order of magnitude larger than the design lifetime (2×10^5 hours) of the subcable lightwave system. Thus, uncoated lasers can be used to find which of the other possible accelerants might be made selective in detecting premature failures, without having our results significantly affected by facet effects due to the presence of humidity in the testing ambient.

Temperature alone is not a driving mechanism for degradation in our lasers, and hence acting alone it is not an accelerant. Elevated temperature shelf lives are very long. In the presence of forward bias, under which optical power and/or injection current are driving mechanisms, temperature is an accelerant. The melting temperature ($183°C$) of the Pb/Sn solder is the approximate upper limit on the harshness of the ambient temperature. For purposes of stressing the active region in CW lasing operation, a more important upper limit on the ambient temperature is the temperature ($\sim100°C$) at which most of the injected current flows into a shunt path.

To the extent to which optical power, as distinct from current, is both a driving mechanism for, and an accelerant of, degradation, it is typically limited to a maximum of ~20 mW/facet at room temperature

(see Fig. 5). It is inconvenient to lower the temperature in order to produce larger maximum CW outputs, and it is somewhat unwise, as well, because the deceleration of degradation due to a lowered temperature could offset the acceleration of degradation due to the increased optical power.

We shall show that direct current, as distinct from optical power, is a driving mechanism for the most significant degradation that we observe. If the direct current is too large, it will flow predominantly into a shunt path, resulting in nonlasing and hence no stressing of the active region (see Section 4.1 and Fig. 5). An upper limit on the current stressing of the shunt path will be related to the melting temperature of the solder. Thus, it is clear that temperature, optical power, and current either cannot or should not be elevated indiscriminately if a proper attempt at reliability assurance screening is to be made.

4.4 Optical power-current (P/I) overstress regime

To stress the active region, current must pass through that region. A lasing output is the best evidence of this. Since substantial currents are required to get substantial output powers at room temperature, the tailored stressing of the active region requires the simultaneous use of optical power and current as accelerants. If P_p is the peak available output power, i.e., the power at which rollover (see Fig. 5) occurs in CW operation, and I_p is the corresponding direct current, then, e.g., in a 30°C ambient, P_p may be as large as 25 mW/facet, and I_p is typically 180 to 250 mA, which is 5 to 15 times the threshold current (I_{th}) at 30°C. Since many of the lasers had P_p (30°C) ≈ 25 mW/facet, it was possible to select a group for constant optical power aging at ≈ 20 mW/facet; this level was chosen to be below P_p to allow for some decrease in P_p with aging. The operating current to maintain an output ≈ 20 mW/facet was continually monitored. This P/I regime was characterized by the following constant stress conditions: $P \approx 20$ mW/facet, $T_{amb} = 30°C$, uncoated facets, and laboratory ambient (≈ 50-percent RH).

Plots of operating current (I_{op}) versus time are given in Fig. 8. The bimodality in degradation behavior is only apparent; a variety of degradation rates are actually present. The initial values of I_{op}, which are ≈ 5 times I_{th}, are, on average, lower in the group showing marked degradation than in the group showing little or no degradation; $I_{op}(0)$ is not indicative of degradation potential. Judging from the sublinearity present in the dc L-I curves, i.e., the substantial departure of P_p from the power output that would have been expected from a straight line extrapolation of the initial value of the L-I slope up to a current I_p, the P/I regime is unavoidably stressing both the active region and
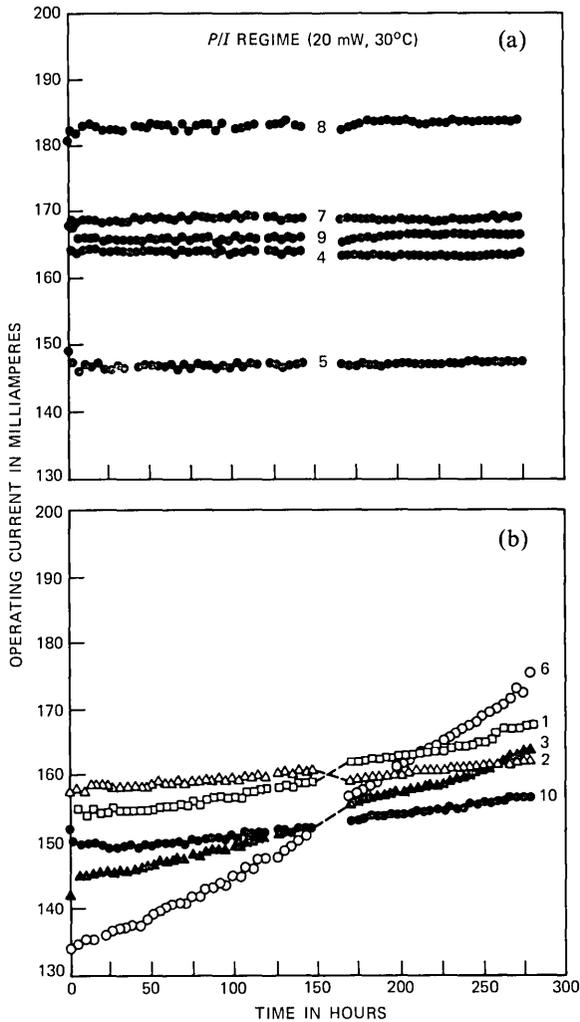
Fig. 8—Plots of operating current (to maintain 20 mW/facet in a 30°C ambient) versus time for ten lasers that had previously only experienced a burn-in (5 mW/facet) at 60°C for 100 hours, and had shown $\Delta I_{op}/I_{op}(0) \leq 5$ percent.

shunt current paths. Although there is some sign of an initial incubation period present in the degradation behavior of some lasers, no stabilization occurs in 275 hours. Even if power, as distinct from current, is an important accelerant, its usefulness is limited not only by the fact that P_p rarely exceeds 25 mW/facet, but as well by the substantial downward spread in actual values of P_p in a given population (see Fig. 9). Henceforth, in attempts to stress the active region, current will be emphasized as the practically important accelerant.
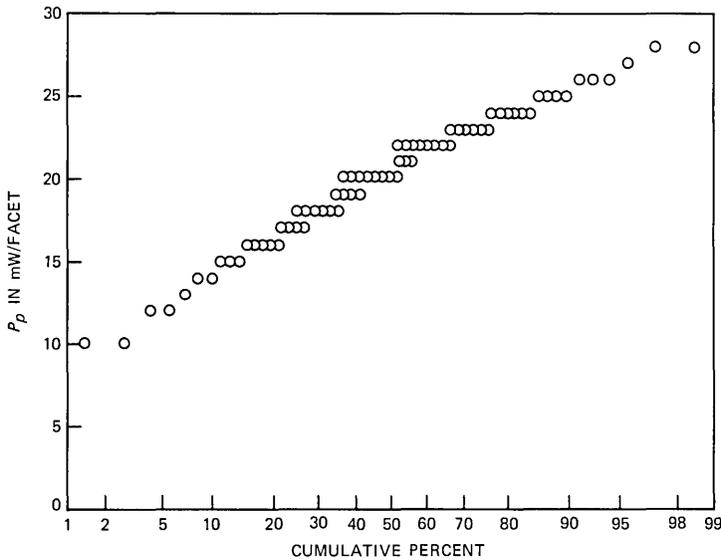
Fig. 9—Normal plot of the peak or rollover CW power (P_p) at 30°C for 70 lasers that had been subjected to only a burn-in (see caption of Fig. 8).

4.5 Temperature-current (T/I) overstress regime

High temperatures and currents will exclusively stress the shunt path. The sublinearity present in the dc *L-I* curves at threshold indicates that some current is always flowing into the shunt path, regardless of the temperature and output power at which the laser is operated. To stress the shunt path alone, the ten lasers aged previously in the *P/I* regime (see Fig. 8) were subjected to the *T/I* regime characterized by the following constant stress conditions: $I(\approx I_p) = 250$ mA; $T_{amb} = 100$°C; uncoated facets; and laboratory ambient. Periodically, the ambient temperature was lowered to 60°C, and the operating current to yield 3 mW/facet was recorded (see Fig. 10).

The most important discovery in the degradation patterns of these lasers is the stabilization effect that occurs within $\lesssim 20$ hours. The lasers (numbers 3 and 6) that exhibited the largest change in Fig. 8 failed to emit 3 mW/facet at 60°C and consequently are not represented in Fig. 10. Lasers (numbers 1, 2, and 10) that exhibited intermediate changes in Fig. 8 exhibited in Fig. 10 increases that were similar to one another. The lasers (numbers 4, 5, 7, 8, and 9) that exhibited the smallest changes in Fig. 8 continued to exhibit relatively small or no changes in Fig. 10. (We shall subsequently present data that show that virtually all lasers exhibit an initial transient mode of degradation, which can be compelled to stabilize rapidly [~a few hours] in a high temperature-high current stress regime; after stabilization,
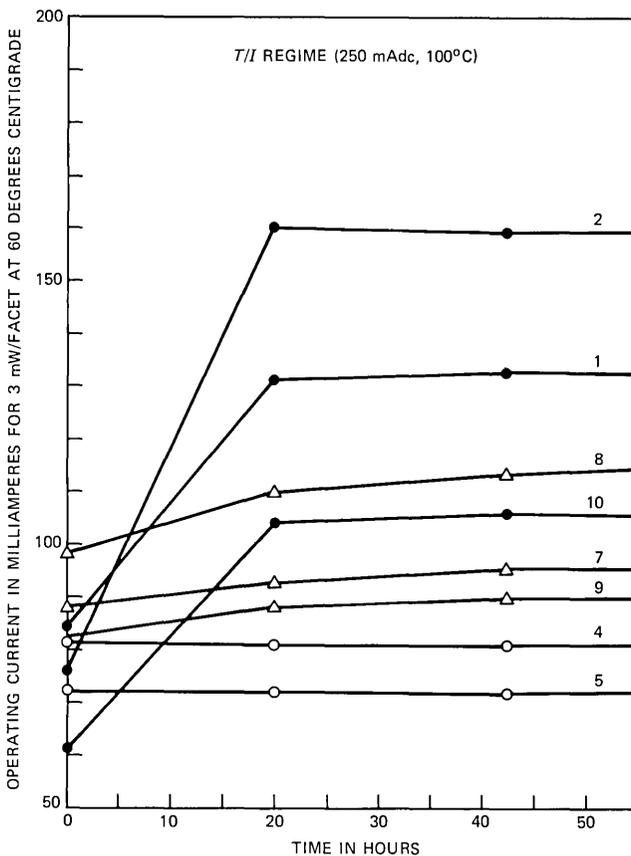
Fig. 10—Plots of operating current (to maintain 3 mW/facet in a 60°C ambient) versus the time that the lasers were subjected to overstressing in a nonlasing regime (100°C, 250 mAdc); the lasers used had previously been subjected to a power-current overstressing (see Fig. 8).

the degradation rates of all surviving lasers are substantially lower and very similar to one another, despite the larger rate differences in the prestabilization time period.) Since the P/I regime stresses the active region and shunt current paths simultaneously, and since the T/I regime almost exclusively stresses the shunt path, the correlation between Figs. 8 and 10 suggests that the prestabilization increases in operating current occur because of some mechanism related to the shunt path. This is important because the shunt path is always operative, even at threshold.

4.6 Effect of T/I and P/I overstress regimes on the initial saturable degradation

To confirm the conclusion that the prestabilization degradation is

largely a shunt path phenomenon, the order in which the P/I and T/I stress regimes had previously been imposed was reversed, and more comprehensive monitoring was done by periodically generating L-I curves. For a second set of lasers the T/I regime was characterized by $T_{amb}^{\dagger} = 150°C$, $I = 250$ mAdc, time = 10 hours, uncoated facets, and laboratory ambient. An L-I curve was generated at a test ambient temperature of 60°C, at the start and finish of this regime. Immediately thereafter, we imposed a P/I regime whose characteristics were $T_{amb}^{\dagger} = 60°C$, $I^{\ddagger} = 200$ mAdc; time ≈ 200 to 300 hours. L-I curves were generated periodically at 60°C. The results of the two sequential regimes are shown for one laser in Fig. 11. The important change occurred in the T/I regime, in which the terminal current was diverted almost entirely into the shunt path. Much less significant changes were observed during the P/I regime, in which the active region path and shunt path were comparably and simultaneously stressed. This suggests that stabilization had been completed for this laser within 10 hours, since further stressing of the shunt path produced no comparably dramatic effect. Another example of degradation during the T/I and P/I regimes is shown in Fig. 12. Note that after the T/I regime, it was not possible to get 3 mW/facet at 60°C at any value of current. A similar large decrease in P_p is the likely explanation for why devices numbers 3 and 6 of Fig. 8, which had exhibited the largest degradation, were unable to emit 3 mW/facet at 60°C and hence were not represented in Fig. 10. The laser of Fig. 12 continued to exhibit marked degradation in the P/I regime; this may have been due to degradation of the active region path as distinct from the shunt path, or it may reflect an additional nonsaturable, or not yet saturated, degradation of the shunt path.

To demonstrate the time evolution of the stabilization of the initial transient mode, another group of lasers were T/I stressed at $T_{amb} = 135°C$ and $I = 250$ mAdc. The rollover power, P_p, at 60°C was recorded at various times throughout the stress duration of 200 hours (see Fig. 13a). This experiment was repeated with another set of lasers stressed

---

† The temperature was chosen so that with $I = 250$ mA, the junction temperature would remain below the melting point (183°C) of the Pb/Sn solder.

‡ This value of current was chosen because $I_p \approx 180$ to 250 mAdc, i.e., the current (I_p) at which the peak output occurs (independent of ambient temperature) is, on average, ≈ 200 mAdc. The emphasis for this version of the P/I regime was not on optical power output as an accelerant, even though $P_p \approx 10$ mW/facet at 60°C. Rather, it was, in part, upon maximizing the current through the active region, whose stressing we desired. More of the injected current (200 mA) would have flown through the active region at a lower temperature (e.g., 30°C), but we also wanted to maintain some thermal acceleration, as well as current acceleration, of the active region path, so 60°C was chosen as the ambient temperature. Temperatures below 60°C are difficult to obtain, even if desirable, because of the current heating of the laser bar holder.
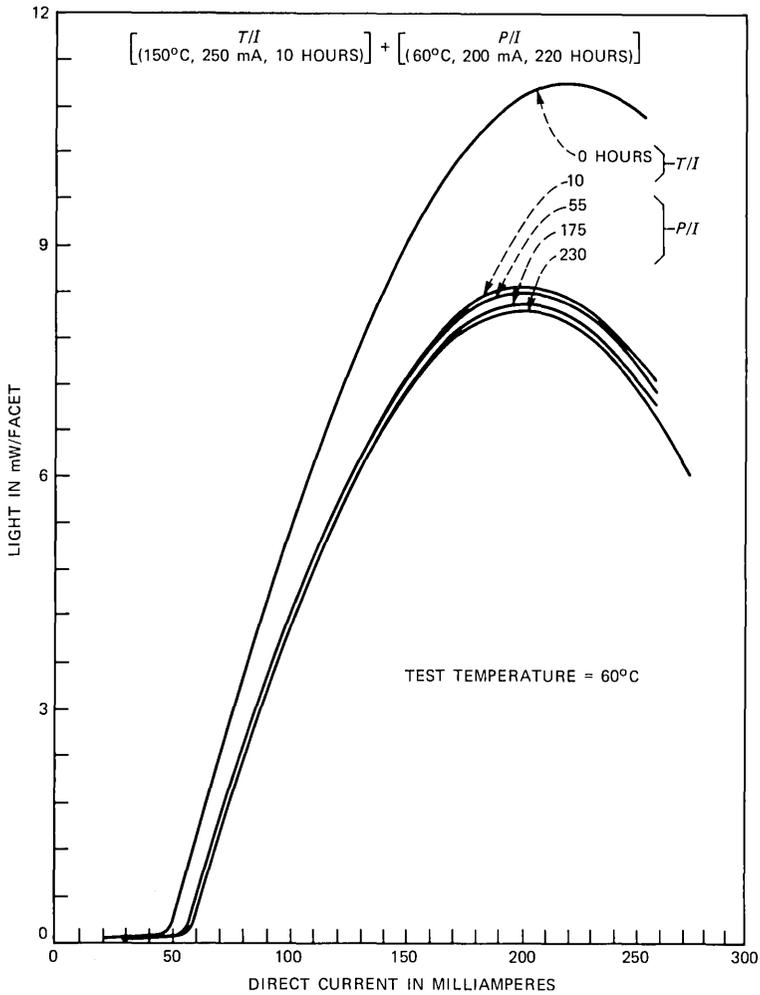
Fig. 11—The effects on the 60°C CW *L-I* curve of a purge, i.e., a sequential temperature-current and power-current overstressing for the time durations shown.

at $T_{amb} = 150°C$ and $I = 250$ mAdc (see Fig. 13b). Stabilization, in the contexts of Fig. 13a and b, means that an initially large rate of degradation of P_p decreases with time until it becomes constant. The current was constant (250 mAdc) in both T/I regimes, and in both cases it flowed exclusively in a shunt path. The optical output was also constant in each case, i.e., very small, because no current went into the active region.

4.7 Activation energy of the saturable initially occurring mode of degradation

Many degradation mechanisms involve the diffusion of atoms and
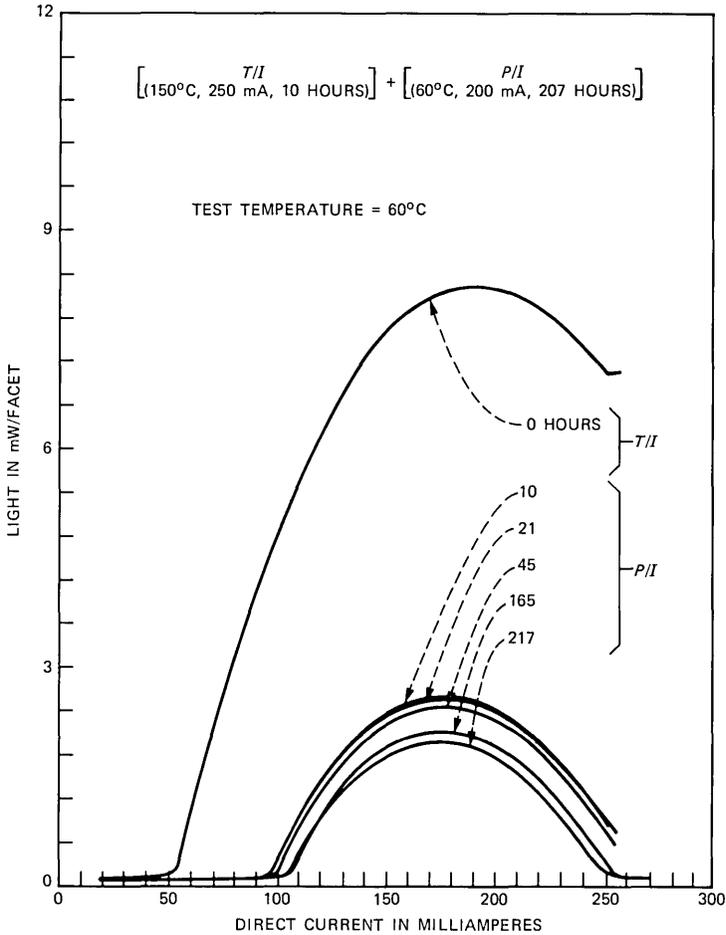
Fig. 12—The effects of sequential overstressing (purging) on another laser; regimes are similar to those of Fig. 11.

hence possess an Arrhenius time-temperature dependence, i.e., a reaction rate given by

$$R \propto \exp\left[-\frac{E_a}{kT}\right], \tag{2}$$

where k is Boltzmann's constant, T is the absolute temperature, and E_a is the activation energy. More complicated temperature-dependent reaction rates might be used, e.g., the Eyring equation.[42] Since there is no physical basis for making a rational choice, the simplest, (2), will be used. We shall follow the approach of Joyce et al.[43,44] R is understood to be the rate of degradation of some parameter that occurs at an
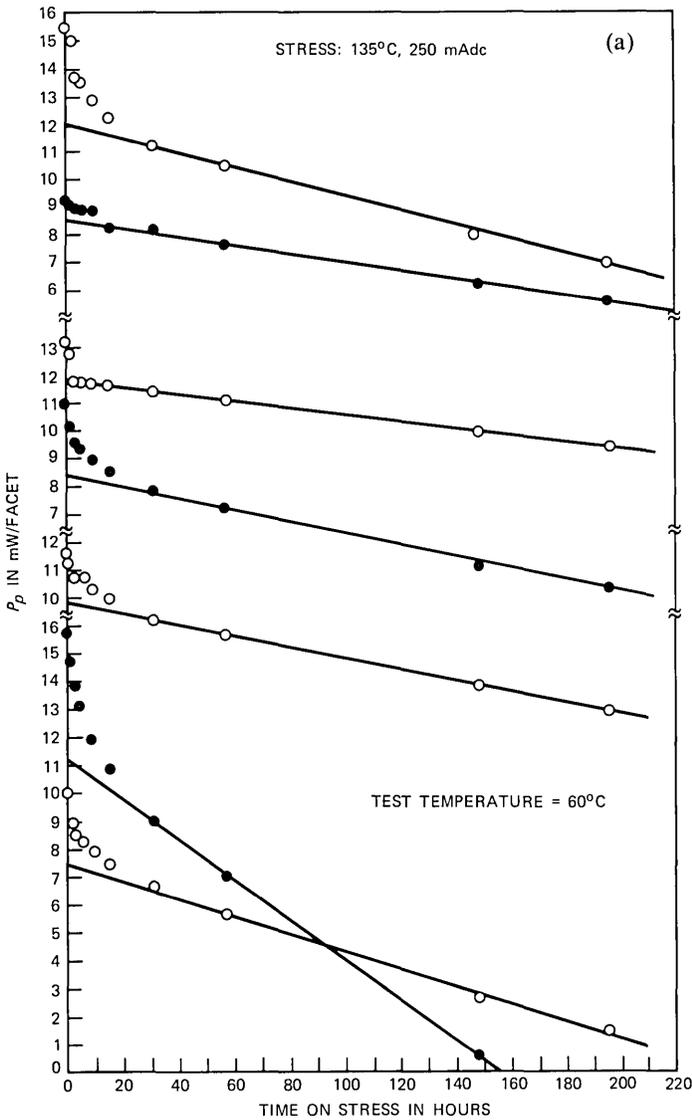
Fig. 13—Peak CW power (P_p) versus time for seven lasers; (a) the stress condition was 135°C and 250 mAdc and the L-I curve monitoring of the effects was done at 60°C. (Cont.)

elevated accelerated aging temperature, T, as measured at some lower reference temperature. This is the activation energy approach.[44] Since no rates of degradation are monitored at the accelerating tempera-ture(s), this approach is suitable for the case of Fig. 13, where the accelerating temperatures exceed that at which lasing occurs. With T_1
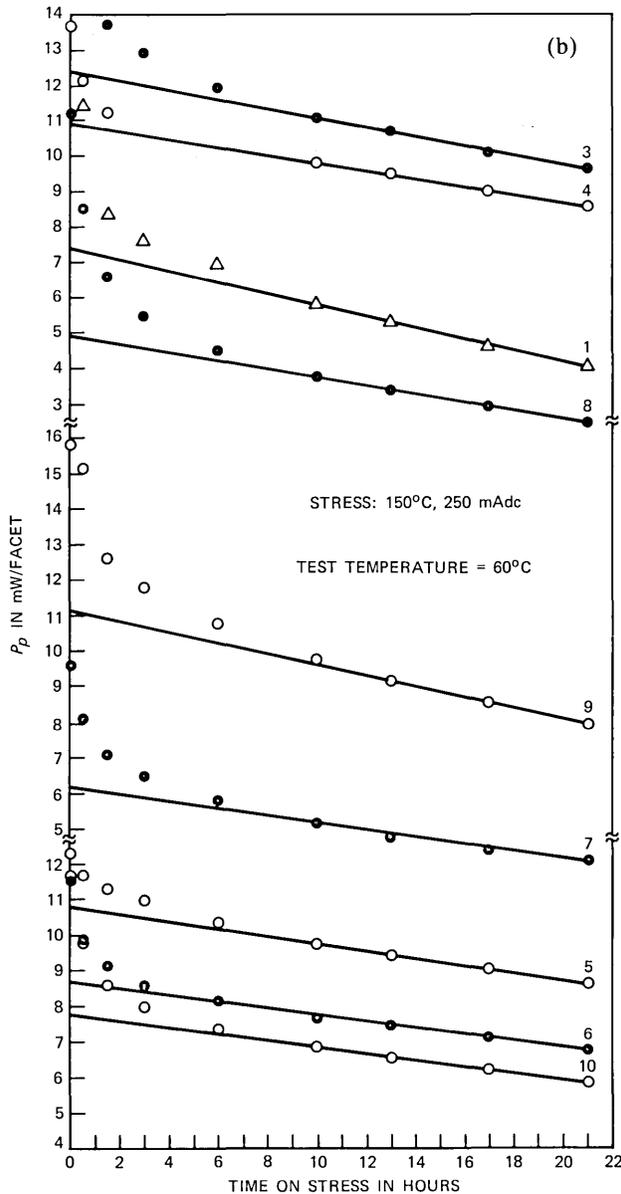
Fig. 13—(b) Same as Fig. 13a except that the overstress temperature is 150°C.

and T_2 as the two accelerating temperatures, E_a may be calculated from

$$E_a = k \left[\frac{1}{T_1} - \frac{1}{T_2} \right]^{-1} \ln \left[\frac{R_2}{R_1} \right], \qquad (3)$$

RELIABILITY ASSURANCE IMPLEMENTATION   831

an expression that is independent of the reference temperature. No explicit current density dependence of the rate of degradation was included in (2) because the current was the same in the thermal regimes (see Fig. 13) to be compared and such dependence would be expected to cancel out in forming the ratio, which is (3).

Adding a current-induced temperature increase (25°C) to the ambient temperatures of 135 and 150°C enables us to compute that $(E_a)_i \approx 1.5$ eV for the initial mode, using the median decreases in rollover power in the first hour, $R(150°C) = 1.9$ mW/h, and $R(135°C) = 0.48$ mW/h, as obtained from Fig. 13. Alternatively, one might suppose that the time duration (τ) required for stabilization also obeys an Arrhenius equation like (2), so that with $R(150)/R(135)$ replaced by $\tau(135)/\tau(150)$, it is calculated that $(E_a)_i \approx 1.2$ eV, using $\tau_{av}(135°C) \approx 24$ hours and $\tau_{av}(150°C) \approx 8$ hours. There is no a priori reason that both calculations should have given equal answers, even apart from uncertainty in the data. For lasing operation in the ambient temperature range 40 to 70°C, it has been found[23] that $(E_a)_i = 0.54$ eV. All estimates of $(E_a)_i$ have been made by comparing degradation properties among groups of lasers that are isothermally aged at different temperatures.

It may always be speculated that the large difference between the 0.54 and 1.5 eV values is due to the presence of two different modes of degradation. A more likely explanation[23] for at least some of the discrepancy is related to the fact that the comparisons leading to estimates of $(E_a)_i$ are being made between *statistically inequivalent* populations that are isothermally aged; considerable variations in rates of degradation and stabilization times can exist within a population at a fixed temperature. Step temperature stressing on individual lasers is preferred[23] as a method for avoiding the large within-population variations, but this scheme is impractical to employ when the rates of degradation are changing significantly throughout the test period. From a practical standpoint, however, knowledge of $(E_a)_i$ is much less important than the knowledge about how to compel a timely stabilization of the initial mode.

4.8 Conclusions about prestabilization degradation

The important result of the two active aging overstress stages of the purge, which were intended to use temperature and current as accelerants for the degradation of both the active region and shunt paths, is that virtually all lasers exhibit an initially occurring transient mode of degradation. Further on, it will be shown that this mode may require 300 to 1000 hours to stabilize in an elevated temperature (60°C) burn-in (3 mW/facet). The rates of degradation associated with this mode

are considerably larger than the rates associated with the mode that controls the long-term behavior.

The stabilization of the initial saturable or transient mode is associated with considerable degradation,* as might be monitored by an increase in operating current to maintain a constant optical output. The prestabilization degradation produced during the 10 hours of T/I overstressing (150°C, 250 mAdc) will be shown to be substantially in excess of the additional degradation subsequently experienced by the same laser during an entire subcable system lifetime (2.2×10^5 hours) at 10°C.

The prestabilization degradation produced during the purge overstressing is caused by current (not by optical power, which is negligibly small in the T/I regime) flowing exclusively in a shunt path (very little current flows simultaneously in the active region path). Current is the crucial stabilizing accelerant.

To determine the rates of degradation and activation energy for the long-term wear-out mode, the initial mode must be stabilized. It is unsatisfactory to produce stabilization solely by means of an elevated temperature burn-in because, as will be shown, it can require $>10^3$ hours of aging to make certain that stabilization has actually been achieved. Since the times-to-stabilization and the extent of the stabilization (increase in current to maintain a fixed optical output) are variable, any attempt to determine activation energies and degradation rates, which are expected to accurately characterize the bases for long-term lifetime predictions, while stabilization is still in progress, will introduce considerable uncertainty[†] into these predictions. Activation energies and degradation rates that include variable amounts of pre- and poststabilization modes of degradation can lead to either overly optimistic lifetime predictions or to unduly pessimistic predictions requiring expensive redundancy in laser installation. Having recognized the existence of an initially occurring mode that eventually saturates or stabilizes, it is imperative that the stabilization process be temporally accelerated to facilitate a timely access to the longer-term ultimately controlling mode.

The overstress regimes of the purge serve *sorting*, and *filtering*, as well as *stabilization*, purposes. Surviving lasers may be sorted according to the increase in operating current (to maintain a constant output) produced during the T/I regime; all else being equal, the lower the operating current, the better, since current is certainly a candidate for

---

* Speculations about possible specific mechanisms will be reserved until Section VII.
[†] See Section 4.9 for a specific discussion of how the initial transient mode can distort determinations of the degradation rates and the activation energy for the long-term mode.

being an accelerant of wear-out. Some filtering (identification and elimination of infant failures) may have been accomplished during preliminary prior-to-receipt burn-in (see Section 4.2). Additional filtering results from the imposition of the T/I regime. Subsequently conducted long-term rate monitoring (see Section 4.9) requires measurable degradation to occur during an elevated temperature (60°C) burn-in (3 to 5 mW/facet) for a laboratory experiment convenient time period (~1 kh). If the degradation that occurs for a laser during the T/I regime is so potent that it is not possible to then get 3 to 5 mW/facet at 60°C for any value of operating current, then that laser is rejected because it is *practically* untestable, even though it might be capable of quite stable long-term operation at lower burn-in temperatures. Thus, if a laser achieves stabilization and is able to emit 3 to 5 mW/facet at 60°C, with some margin, it is a survivor. On the other hand, if a laser loses the capability of emitting 3 to 5 mW/facet at 60°C, whether or not it has stabilized, then it is rejected.

4.9 Long-term rates of degradation—Elevated temperature burn-in rate-monitoring regime

To this point, we have examined the harsher shorter-term components (T/I and P/I regimes) of the active aging portion of the purge. Yet to be considered is the longer-term elevated temperature burn-in of the truncation active aging regime on devices that have passed through (or survived) the earlier imposed harsher stress regimes of the purge. The general intent of the burn-in testing is to detect the early failure fraction of the wear-out population.[9] Specifically, the elevated temperature burn-in stage may be used to do the following: (1) to establish a "fan" in the distribution of the long-term degradation rates to permit a deterministic selection of the least rapidly changing devices for subcable applications; (2) to determine whether the harsher prior stages of the purge actually performed their intended function, i.e., to stabilize an initially occurring transient mode of degradation;* and (3) to demonstrate, by comparison with the burn-in of lasers that had not been exposed to the T/I and P/I regimes, that these purge regimes do not introduce degradation that would not otherwise have been present.

Forty lasers were divided into two equal groups. Each group had an equal number of representatives from each of 11 wafers. One group, but not the other, was exposed in sequence to a T/I regime ($T_{amb} = 150°C$, $I = 250$ mAdc, laboratory ambient, and time = 10 hours) and

---

* If stabilization has not been achieved, as might be detected in the first ~100 hours of the burn-in, the relevant lasers could be recycled through the prior harsher stages. Those devices which remained unstabilized would be discarded after another elevated temperature burn-in.

then a P/I regime ($T_{amb} = 60°C$, $I = 200$ mAdc,[†] laboratory ambient, and time = 350 hours). Both classes were then subjected to an elevated temperature (60°C) burn-in (3 mW/facet) for 1000 hours. An ambient temperature equal to 60°C is approximately the lowest temperature at which unambiguous degradation is likely to be seen in a convenient time period (~1 kh). The 3 mW/facet level was chosen so as not to eliminate from the burn-in test any laser that might not have been able to produce a higher output (e.g., 5 mW/facet), but that would otherwise have exhibited tolerably slow degradation. Figure 14a displays the smoothed aging behavior for the group that was not exposed to the overstress regimes (T/I and P/I). Figure 14b is the smoothed behavior for the group that had previously been overstressed. Of the 20 overstressed lasers, two were not admitted to the burn-in test because they could not emit 3 mW/facet, at any current, at 60°C; this is the result of L-I degradation similar to that shown in Fig. 12. Three of the overstressed lasers had operating currents >120 mA and were omitted from Fig. 14b; 15 lasers are represented in Fig. 14b. A number of conclusions about Fig. 14 can be made.

The previously overstressed population of Fig. 14b exhibited low rates of degradation in contrast to the previously unstressed group of Fig. 14a, which exhibited a variety of much larger rates of degradation. Despite the variety, all of the degradation patterns of Fig. 14a are caused by an initially occurring saturable (in different times, 300 to 1000 hours) mode of degradation that can be compelled to stabilize in ~10 hours during a T/I overstress regime. Although the overstress regimes caused an increase in operating current (I_{op}) to maintain a constant output, the subsequent degradation rates during burn-in are largely immeasurably low in the 1-kh time period. If the presence of the initially occurring transient mode is not recognized as being controlling for virtually all of the lasers of Fig. 14a, then the aging behavior during a conventional 3-mW/facet, 60°C, 100-hour burn-in, without prior overstressing, can give misleading indications about long-term reliability.

Prior to any attempt to measure E_a for the long-term mode of degradation, it is crucial to stabilize the initially occurring transient mode. To see why, consider the laser represented by the uppermost curve in Fig. 14a, which exhibited stabilization in ~400 hours. Imagine that a step temperature stress procedure[23,45-47] employing measured

---

[†] Since the P/I regime is designed to overstress the active region current path, the stressing current was lowered from the 250 mAdc of the T/I regime to 200 mAdc in order to avoid the sudden pnpn switching effects (e.g., Fig. 5b) that can divert significant additional current into the shunt path. These sudden switching effects invariably occur above $I_p \approx 200$ mAdc.
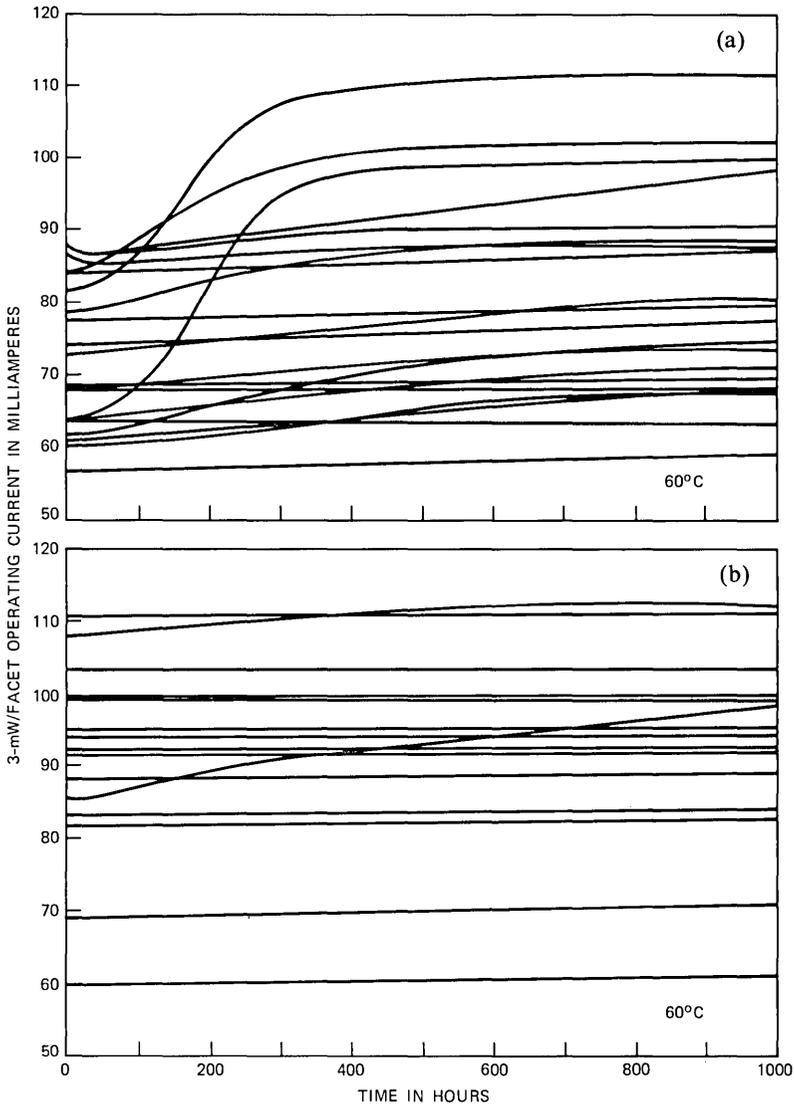
Fig. 14—Comparison plots of aging behavior (operating current to maintain 3 mW/facet) at 60°C for companion groups of lasers which (a) had not, and (b) had been previously overstressed (purged).

rates of degradation has been used to determine the E_a for this laser, which had not passed through the T/I and P/I overstress regimes. Suppose that after testing at 60°C for 200 hours, the ambient temperature had, for example, been increased to 70°C; the degradation rate would naturally be expected to increase. Since, however, the rate at 60°C would have undergone a significant reduction in the 200- to 400-

hour period had aging been continued at 60°C, the rate at 70°C would actually be less than it might have been if the initial (0- to 200-hour) rate at 60°C had persisted. Consequently, this particular step stress direction (an increase in ambient temperature, 60 to 70°C) would yield a value for E_a that was too small. Similar reasoning shows that if the step stress direction had been reversed (a decrease in ambient temperature, e.g., 60 to 50°C), E_a would have appeared too large.

For another example, if sublinear behavior is detected for a given laser, then any temporal extrapolation based upon fitting the degradation data to a sublinear time dependence may give greatly overestimated laser lifetimes, if the time dependence of the degradation were to become linear after the initially occurring mode has stabilized. Alternatively, if the degradation time dependence were to be *incorrectly perceived* as linear during an initial burn-in ($\sim 10^2$ hours), an unduly pessimistic lifetime prediction might be made upon the incorrect assumption that a large degradation rate remained uniform in time well beyond the burn-in period.

From a comparison of Fig. 14a and b, it is seen that the increases in I_{op} caused by the overstress regimes are comparable to those that occur in the absence of any prior overstressing.* The overstress regimes compelled the stabilization to occur quickly and caused the elimination of 5/20 lasers, which were viewed as having incurred too much degradation.† The overstress regimes did not appear to cause any degradation that, but for the overstress, would not otherwise have been present.

Figure 15 exhibits the normalized operating currents for 14/15 of the lasers of Fig. 14b for ≈ 7000 hours of a 60°C, 3-mW/facet burn-in. The largest increase in $I_{op}/I_{op}(0)$ was ≈ 22 percent in ≈ 7000 hours. The purge regimes that preceded the extended duration burn-in at 60°C did not consume intolerable fractions of useful operating lifetimes, nor were they inherently destructive of good lasers. The previously unstressed lasers (Fig. 14a) also exhibit very low rates of degradation at 60°C beyond 1 kh, which are comparable to those of Fig. 15. Thus,

---

* Because of this, it cannot be argued that the initial increase in current during the 150°C T/I purge regime is caused by changes in the Pb/Sn solder, since similar increases occur at 60°C, over much longer periods of time.

† Continuous operation of a laser at any current less than, but near, I_p (the current corresponding to the peak power output, P_p, of the rolled-over dc L-I curve, e.g., see Fig. 12), where the slope efficiency is low, will be likely to lead to a sudden failure. Even a minor increment in degradation, which has the effect of shrinking the L-I curve in the manner of Fig. 12, can cause a substantial increase in current to maintain a constant output. When the operating current equals I_p, any further degradation causes the current to increase rapidly to the limit of the power supply, since at no value of current is it any longer possible to obtain the desired output; P_p is now less than the desired output. The time dependence of this degradation would be superlinear in operating current.
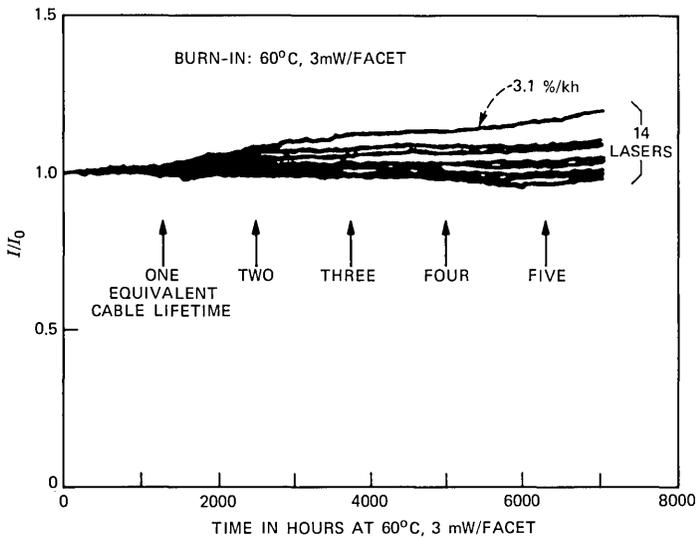
Fig. 15—Elevated temperature (60°C) burn-in (3 mW/facet) for 7 kh of a group of 14 purged lasers.

stabilization of the initially occurring transient mode will eventually take place without the necessity for overstressing; the penalty exacted is ~1 kh of aging at 60°C before it is reasonably certain that a long-term mode of degradation has become controlling. No correlation was found between the purge-induced initially occurring decrease in P_p (Figs. 11 through 13) and the long-term wear-out degradation rate.

To test whether the long-term degradation rates were increased at higher output power levels, another batch of lasers was purged and burned-in at 60°C for 2000 hours, with half aged at 5 mW/facet and half at 3 mW/facet. The Normalized Degradation Rates (NDR) are shown in Fig. 16. It is uncertain from these data that there exists any significant dependence upon level of operation. Other workers,[23,31] however, using a larger range of outputs (1 to 7 mW/facet) and step-stress aging, were able to show that operation at higher outputs produced larger degradation rates. These findings[23,31] are expected since the unbiased shelf life of these devices is long; current and/or optical power are certain to be long-term degradation driving mechanisms and accelerants. Our inability to perceive a significant difference in degradation behavior between two levels of operation (3 and 5 mW/facet), by comparing the widely varying NDR of two separate groups of lasers (see Fig. 16), emphasizes the benefit of the step-stress aging technique.[23,31]

Two points may be made about the dependence[23,31] of the long-term NDR upon the output power level of operation. The first is that, at
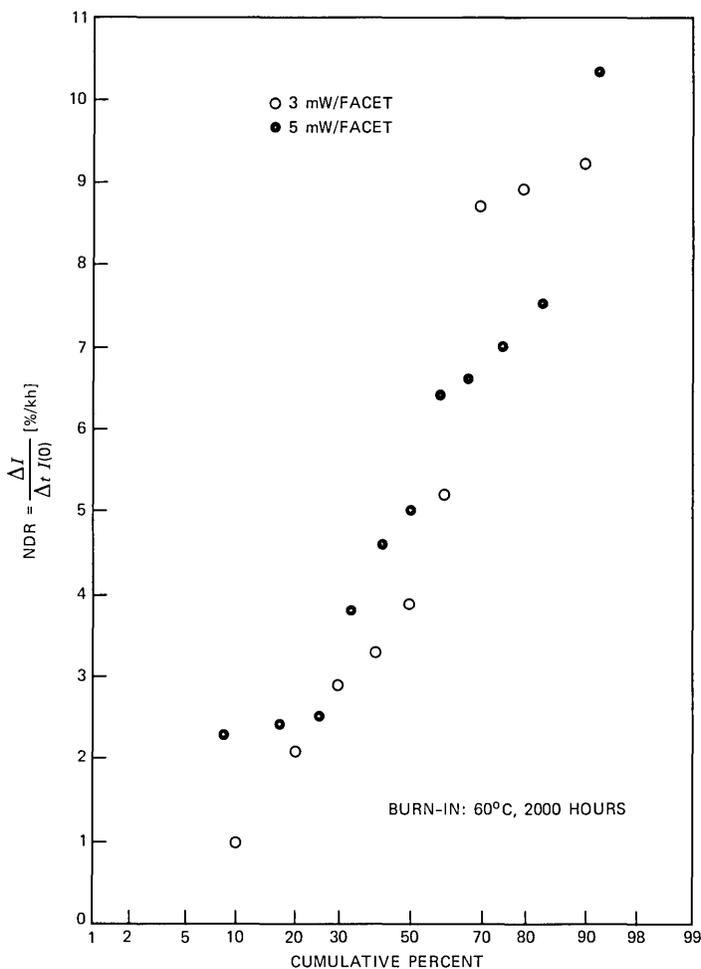
Fig. 16—Normalized degradation rates for two different groups of purged lasers operated at 60°C at 5 and 3 mW/facet for 2000 hours.

least for subcable use, the NDR are so small that demands for larger (than 5 mW/facet) laser transmitter output powers could be satisfied. The second relates to whether the driving mechanism for long-term degradation is optical power or current. To get a higher output power at some temperature, the current must be increased. As the current is increased, the fraction of that current diverted into a shunt path is also increased; this is a nonlinear effect (see Fig. 5b). In our view it has not yet been conclusively established whether, e.g., it is the optical power[23] within the active region or the current in some shunt path (which is activated even at threshold) that is responsible for long-term degradation.
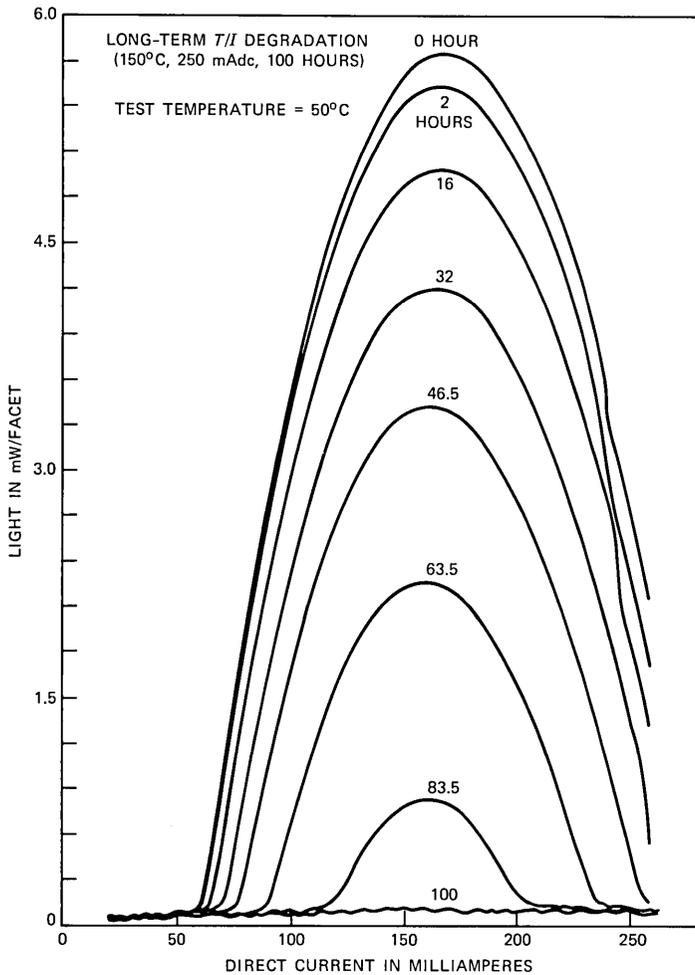
Fig. 17—The effect on the 50°C CW *L-I* curve of a laser as it is aged to complete failure in a 150°C, 250-mAdc, 100-hour regime.

V. ACTIVATION ENERGY OF A LONG-TERM NONLASING SHUNT PATH MODE OF DEGRADATION

Figure 13 shows that beyond the time of stabilization in a T/I purge regime, there is a longer term mode of degradation present. An example of complete degradation due to long-term T/I aging, as monitored by the lasing output at 50°C, is shown in Fig. 17. We wish to determine the activation energy for the long-term T/I mode, because, if too high a purge ambient temperature is chosen, it is possible that a significant fraction of the useful lifetime of a "good" device would be consumed if the duration of the T/I regime were excessive. It is also desirable to

know the activation energy for this shunt path mode, because the shunt path is activated even at the lower temperatures and currents associated with normal laser operation.

The process of using a T/I regime to stabilize the initial transient mode in a timely fashion and then using a T/I regime to produce long-term degradation, with the eventual purpose of determining some upper limit on T, has a distinct bootstrap appearance. We were, however, aided by our initial observation (see Fig. 10) that stress temperatures $\approx 100°C$ were not significantly deleterious, and by a practical upper limit that was the melting temperature (183°C) of the Pb/Sn solder. Using our previous determination (see Section 4.1) that the thermal impedance $R_{th} \approx 50°C/W$ and that $V = 2$ volts when $I = 0.25A$, the use of (1) yields $\Delta T = 25°C$. An upper limit on the ambient temperature is given by subtracting the incremental junction heating (25°C), which occurs during the T/I overstress, from 183°C. For safety we picked $T_{amb} \leq 150°C$. (Extra margin is derived from the p-up bonding. Even if the active region reached 183°C, it is unlikely that this temperature would be felt at the n-side, Pb/Sn solder interface.)

For purposes of transient mode stabilization, a first T/I regime was imposed and characterized by $T_{amb} = 125°C$, $I = 250$ mAdc, and $\Delta t = 150$ hours. Next the two accelerating T/I regimes were imposed: $T_{amb} = 100°C$, $I = 250$ mAdc, $\Delta t \approx 200$ hours; and then $T_{amb} = 150°C$, $I = 250$ mAdc, and $\Delta t = 100$ hours. Optical power was not a significant accelerant in either of these regimes, since the power outputs were low and similar (<0.1 mW). The stressing current, $I = 250$ mAdc, was kept constant.* Periodic monitoring was performed at 50°C where a dc L-I curve was generated each time.

A number of different parameters may be used to record the rate of degradation, e.g., the threshold current (I_{th}), the current required to produce a 3-mW/facet output (I_3), I_5 $\cdots$ or P_p. Since I_{th} is somewhat ambiguous in view of an increasing "softness" in the turn on with increasing time, since a variety of kinks appeared in some of the curves at 3 and 5 mW, and since some curves had no output at 5 mW at 50°C, we restricted our attention to the increases in $I_{1.5}$ and the decreases in P_p. Correcting the ambient temperatures (100 and 150°C) by the current-induced junction heating (+25°C), which is the same for both regimes, the rates of degradation of $I_{1.5}$ and P_p and use of (3) lead to the values of $(E_a)_{sh}$ for the long-term shunt path mode shown in Table I for seven lasers. $I_{1.5}$ as a function of time is shown in Fig.

---

*If the stressing current had been different in the two accelerating regimes, then deriving a value for E_a based upon a ratio of the degradation rates at the two temperatures would impute solely to temperature an effect which may actually have been a function of both temperature and current.
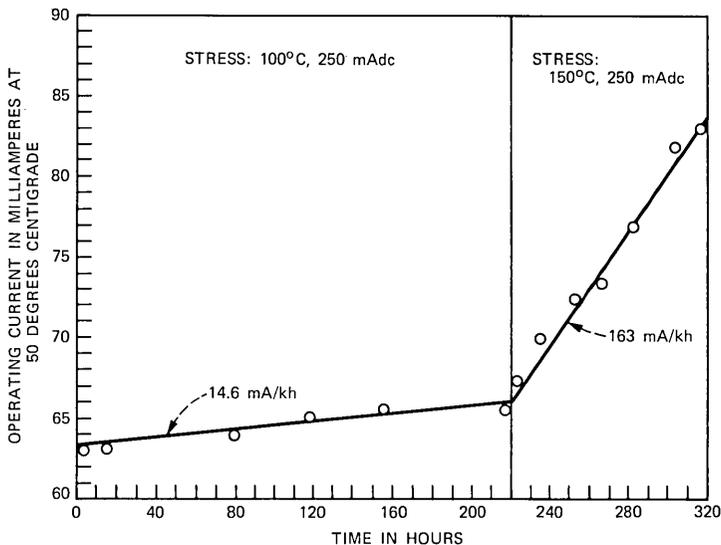
Fig. 18—Plots of the 50°C values of the operating current ($I_{1.5}$) to produce 1.5 mW/facet in the course of stressing at 250 mAdc, first in a 100°C and then in a 150°C ambient. Previously, overstressing was used to stabilize an initial transient mode.

18 for laser number 3 and P_p as a function of time is shown in Fig. 19 for laser number 4. Since step-stress aging was used, a value of $(E_a)_{sh}$ for each laser could be determined.

From the apparent linear degradation* with time after stabilization (Figs. 13, 18, and 19) it is reasonable to assume that there exists a single dominant mode of degradation that is responsible for the changes in both $I_{1.5}$ and P_p. Consequently, the values determined from these two indicia might be expected to be the same. The device-to-device differences in $(E_a)_{sh}$ are not considered meaningful. We shall use $(E_a)_{sh} \approx 0.85 \pm 0.1$ eV as a common population descriptor[23,44] of thermal aging in the nonlasing T/I regime.

It is of interest to compare our findings with the activation energies determined by others for operation in the light-emitting regime. Step stressing was used to determine[23] an activation energy, in long-term lasing operation, equal to 1.05 ± 0.15 eV in the ambient temperature range 40 to 80°C. In the course of lasing operation, both the active region and shunt paths may be simultaneously degrading, since both

---

* For the reason given in Section 4.9, the rate of increase of $I_{1.5}$ will tend to become superlinear as P_p degrades downward toward 1.5 mW; as P_p reaches $\leqslant 1.5$ mW, I_p will tend toward infinity, which in a practical case is the maximum current available. The linear rates observed in Fig. 18 are expected to hold provided that P_p does not approach too closely the desired constant output power.
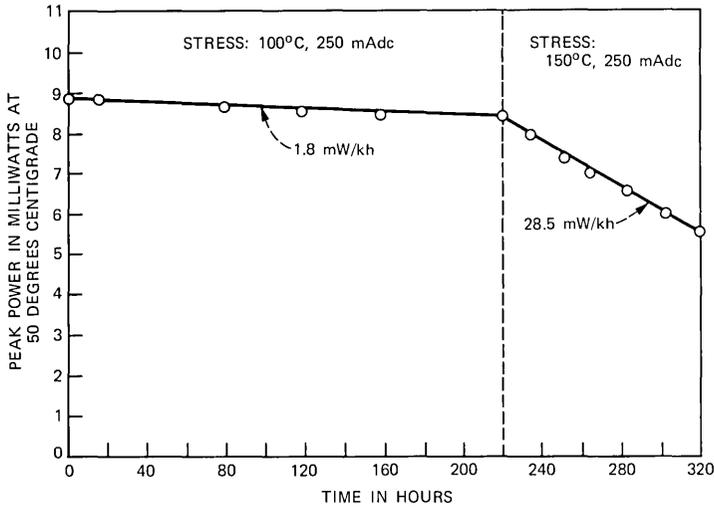
Fig. 19—Same as Fig. 18 except that the monitored parameter is the peak CW power (P_p) for another laser.

Table I—Experimentally determined activation energies using step temperature aging

Laser Number	$I_{1.5}$ (eV)	P_p (eV)	Average E_a (eV)	Measurement Error (eV)
1	0.86	0.80	0.83	±0.03
2	0.85	—	—	—
3	0.74	0.94	0.84	±0.10
4	0.94	0.85	0.90	±0.05
5	0.81	0.77	0.79	±0.02
6	0.86	0.74	0.80	±0.06
7	0.99	0.82	0.91	±0.08
Average	0.86	0.82	0.85	±0.06
Standard Deviation	±0.08	±0.07		

are being stressed. Isothermal aging of lasers, of exactly the same structure and composition, in the ambient temperature range 50 to 70°C, yielded[4] an extrapolation energy[43,44] $E_e \approx 0.9$ eV. For the temperatures of interest it may be computed (see Ref. 44, eq. [33]) that $E_e = E_a + 0.14$ eV, so that $E_a \approx 0.76$ eV. This measurement also presumptively involved degradation of either one or both of the active region and shunt paths. In an LED configuration (no shunt path), the activation energy in the ambient temperature range 170 to 250°C for the thermally induced degradation of the active region path, as measured at room temperature, was found[7] to be $E_a \approx 1.0$ eV.

Considering the experimental uncertainties (step-stress versus isothermal* aging), and the differences in operating conditions (lasing measurements[4,23] were for constant optical power, while the LED measurements[7] were at constant current), the above values of E_a, including ours for the shunt path, are not viewed as significantly different. This does not mean, however, that for our lasers the shunt path mode of degradation is controlling during low temperature lasing operation; we have not shown that the shunt path has an $(E_a)_{sh} = 0.85$ eV at low temperatures. The similarities in the above values of E_a only suggest that if there is an active region mode of degradation that is different from a simultaneously present shunt path mode, then they may have nearly the same values for the long-term E_a.

VI. LONG-TERM LIFETIMES

This paper is concerned exclusively with light-emitting reliability, and not with any lifetime limitation imposed by kinks in the L-I characteristic, beam wander, mode hopping, etc. Figures 8 and 10 through 19 indicate that all degradation of our lasers is characterized by both increased operating current (I_{op}) and decreased rollover or peak power (P_p). The system-imposed lifetime limitation on P_p is $P_p \geq 5$ mW/facet, and the limitation on I_{op} is $I_{op} \leq I_m$, where I_m is the maximum available current from the transmitter compensating circuit. If $I_p (= 175$ to 250 mAdc) is the current at which P_p generally occurs, we shall assume that $I_m < I_p$, so that the lifetime will be current- and not power-limited.

For the purpose of making plausible lifetime estimates, we shall assume that the operating current (I_{op}) to produce a constant optical output is linearly dependent upon time. If I_f is the value of I_{op} at failure $(t = \tau)$, then the lifetime (τ) is given by

$$\tau = \frac{\dfrac{I_f}{I_{op}(0)} - 1}{\text{NDR}}, \tag{4}$$

where the Normalized Degradation Rate (NDR) is given by

---

* Batch isothermal aging to deduce activation energies can be meaningful only if the populations are large and statistically equivalent. Our step-stress aging measurements yielded $(E_a)_{sh} = 0.85$ eV. In contrast, if $(E_a)_{sh}$ had been determined from the isothermal poststabilization data of Fig. 13, it would have been found that $(E_a)_{sh} \approx 1.9$ eV. The discrepancy is believed to be due to the effect produced by large variations in degradation rates among devices at a fixed temperature. Another example of what might be the same problem comes from isothermal measurements made on another structure, the double-channel planar BH laser operating at 1.3 μm. The E_a for the long-term mode appears to fall in the range of a few tenths to 0.4 eV (see Refs. 48 and 49), which is considerably below the step-stress values that we and others[23] have found for our structure.

$$\text{NDR} = \frac{1}{I_{op}(0)} \frac{dI_{op}}{dt}. \qquad (5)$$

An upper limit which always exists on I_f is the value I_m. It may be desirable, however, to allow for the imposition of a stricter requirement upon I_f, so that one may write

$$I_f = \text{lesser of } [\beta I_{op}(0) \text{ or } I_m]. \qquad (6)$$

The rationale for choosing a small value of $(\beta - 1)$ is that if I_{op} is not permitted to change by much within a subcable system lifetime, then very likely no other property will change much either. This can provide added confidence that a variety of electro-optic deficiencies will not occur within the system lifetime. A small value of $(\beta - 1)$ also tends to validate the assumption that I_{op} is increasing uniformly with time.

As an example, it appears possible to choose $\beta = 1.05$, which permits only a 5-percent increase in I_{op} in a system lifetime. Assuming that $I_{op}(0) \le 0.95 \, I_m$, then from (4) and (6) one gets

$$[\tau][\text{NDR}] = \text{constant} = 0.05. \qquad (7)$$

This will be true at all temperatures. With a very large yield, it appears that postpurge aging at 60°C, 5 mW/facet produces NDR (60°C) $\le$ 5 percent/kh. Consequently, from (7) it is computed that $\tau(60°\text{C}) = 1$ kh. Assuming that τ scales with temperature by a factor

$$\gamma = \exp\left[\frac{E_a}{k}\left(\frac{1}{T_1} - \frac{1}{T_2}\right)\right], \qquad (8)$$

then with $T_1 = 10°\text{C} = 283\text{K}$, $T_2 = 60°\text{C} = 333\text{K}$, and $E_a = 1.0 \text{ eV},$[23] it is computed that $\gamma = 469$ so that $\tau(10°\text{C}) = 469$ kh, which is twice the required subcable system lifetime (25 yr $= 220$ kh) at 10°C. Thus, by the use of the best estimate[23] of the value of E_a for long-term degradation and an exceptionally strict definition of lifetime, the lasers appear more than adequate for the needs of the system.

Alternatively, one may ask about the consequences of using a lower, and hence more conservative, value of the activation energy. If, for example, $E_a = 0.85 \text{ eV}$, the low end of the measured[23] range, then $\gamma = 186$, and thus a $\tau(10°\text{C}) = 220$ kh has a 60°C equivalent lifetime of $\tau(60°\text{C}) = 220/186 = 1.2$ kh. In this instance, the 14 lasers of Fig. 15 have lived more than five cable lifetimes, and in that period the largest increase in the operating current has been ≈ 22 percent, an increase easily accommodated by a feedback compensating circuit.

A third alternative view emerges from answering the following question. Assuming that the correct value of E_a is lower than has been measured,[4,23] how low could it be before the subcable system would become unfeasible? Suppose that the transmitter compensating cir-

cuits can accommodate a 50-percent increase in I_{op}, then $\beta = 1.5$ in (6), so that from (4) it is calculated that

$$[\tau][\text{NDR}] = \text{constant} = 0.5. \tag{9}$$

Given that NDR (60°C) = 5 percent/kh, an upper limit on observed rates, it follows that τ(60°C) = 10 kh. Since τ(10°C) = 220 kh, it also follows that $\gamma = 220/10 = 22$, so that from (8) it is calculated that $(E_a)_{min} \approx 0.5$ eV. It should be noted, however, that after several thousand hours of a postpurge 60°C, 5 mW/facet burn-in, it is found[50] with a high yield that NDR (60°C) ≤ 1 percent/kh, so that on this basis, τ(60°C) = 50 kh and $(E_a)_{min} = 0.24$ eV. (The decrease in NDR is due to a sublinear time dependence.)

This leads to a fourth alternative framework for viewing the adequacy of the lasers. If it can be shown that the lasers will live adequately long (220 kh) at 60°C, then regardless of the activation energy, they will be adequate, a fortiori, at 10°C. Substituting τ(60°C) = 220 kh into (9) yields $[\text{NDR (60°C)}]_{max} = 0.23$ percent/kh. There are lasers which satisfy this condition, but more work is required to produce a reliable yield estimate.

All of our concern up to this point has been with the question, will the lasers last as long as the system? We have not made any serious attempt to answer the question, how long can the lasers live? The sublinear behavior[50] in postpurge burn-in tests offers the hope that the lasers have the potential for long life. An upper bound may be the estimated[7] 60°C lifetimes of LEDs that are $\approx 10^9$ hours. This is, however, probably too optimistic; measured[7] lifetimes of LEDs in a 170°C ambient are $\approx 10^5$ hours, while the lifetimes of our lasers in a 150°C ambient are orders of magnitude shorter (see Figs. 13b and 17 through 19). If LED lifetimes are assumed to reveal a "material" limit upon light-emitting lifetimes, then our lasers may have a "structure"- or "processing"-induced limit.

VII. CONSIDERATION OF POSSIBLE ORIGINS OF THE INITIAL SATURABLE MODE OF DEGRADATION

In this section we shall consider three possible specific explanations for the origin of the initial transient mode, which stabilizes after some aging: (1) an increase in temperature sensitivity, e.g., thermal impedance (R_{th}); (2) lasing-induced facet oxidation; and (3) the formation of precipitate-like defects caused by the in-migration of Au from the p-contact. The literature of reliability shows that each has been associated with an initially occurring temporally saturable mode of degradation. Existing evidence can be used to rule out the significance of (1) and (2); existing evidence tends to favor (3) or something similar to (3).

7.1 Possible increase in temperature sensitivity of pulsed and dc thresholds

7.1.1 Increase in pulsed threshold due to a possible decrease in T_0

The empirical expression universally used[51] to relate the threshold current of a laser to the temperature of the active region is given by

$$I_{th}(T) = C \exp\left[\frac{T}{T_0}\right], \tag{10}$$

where C and T_0 are empirical constants, the latter being so only for a restricted temperature range. In low duty cycle pulsed operation, the temperature of the active region is just the ambient or stud temperature. To detect changes in T_0, we compared ratios of pulsed thresholds at different ambient temperatures before and after aging. Using (10), T_0 is calculated to be

$$T_0[K] = \frac{40}{\ln[I_{th}^{pul}(70°C)/I_{th}^{pul}(30°C)]}. \tag{11}$$

Before aging, $[I_{th}^{pul}(70°C)/I_{th}^{pul}(30°C)]_{t=0} = 58.9 \text{ mA}/35.6 \text{ mA} = 1.65$, and hence $T_0 = 79.9$K, for a particular laser. After a T/I stressing at 125°C, 250 mAdc for 20 hours in order to stabilize the initial transient mode of degradation, the thresholds had increased approximately 50 percent, $[I_{th}^{pul}(70°C)/I_{th}^{pul}(30°C)]_{t=20 \text{ hours}} = 86.1 \text{ mA}/51.8 \text{ mA} = 1.66$ and $T_0 = 78.9$K, a value essentially unchanged because of the initial degradation. Alternatively expressed, if a decrease in T_0 were responsible for causing an increase in I_{th}^{pul} (70°C) from 58.9 to 86.1 mA, then it is calculated that $(\Delta T_0)_{cal} = -6.5$K; from above it is seen that $(\Delta T_0)_{exp} = -1.0$K. Thus, the observed initially occurring stabilizable degradation appears unrelated to T_0, which remained unchanged.

Similar comparisons of experimentally determined values of T_0, before and after a much more significant stressing (30°C, 20 mW/facet, 280 hours, followed by 100 to 125°C, 250 mAdc, 250 hours) were made on ten lasers. Although most of the lasers exhibited an apparent decrease in T_0 of ≤6 percent, this is not viewed as being outside experimental error,* which when taken into account permits the conclusion that T_0 does not decrease significantly in the course of degradation. This agrees with other work.[23] At least with regard to pulsed threshold increases, in the course of aging, the degradation must affect the prefactor C in (8), rather than T_0.

---

* The algorithm for determining I_{th} depended on the shape of the pulsed L-I curve in the vicinity of threshold. Prior to aging, the 30°C L-I curve is slightly concave upward, which leads to an overestimate of I_{th}. After aging, the 30°C L-I curve is slightly concave downward, which underestimates I_{th}. These deviations from ideality tend to make T_0 appear lower after aging. Additionally, two lasers that suffered virtually no degradation showed $\Delta T_0 \approx -2$ percent, while another laser that showed substantial degradation had a $\Delta T_0 = +1$ percent.

7.1.2 Increase in dc threshold due to a possible increase in thermal impedance (R_{th})

(GaAl)As lasers bonded p-side down with In solder have a well-documented history of reliability problems associated with temporally occurring metallurgical interactions between In and Au. Thus, it was found that voids could form in the mixed In-Au layer[52,53] during device operation, that voids can also form[52] during 1000 hours of unbiased storage aging at 70 to 100°C, that the bonds of life-tested devices exhibit poor adhesion,[54] that aged chips can often be easily pulled off their studs,[53] and that the thermal impedance increases during device operation[46,53,54] so as to exhibit what appears to be a saturable mode of degradation,[53,54] reminiscent of the knee-like effects shown in Figs. 10, 13a and b, and 14a. We thought it was important to consider a possible instability of the Pb/Sn bond despite the demonstrated stabilities of other eutectic alloy solders.[7,27-29,33]

The dc threshold may be written in terms of the pulsed threshold as

$$I_{th}^{dc} = I_{th}^{pul} \exp\left[\frac{R_{th} V I_{th}^{dc}}{T_0}\right]. \tag{12}$$

The incremental temperature increase of the active region above the ambient temperature, due to injection current heating, equals $R_{th} V I_{th}^{dc}$, where, as in (1), R_{th} is the thermal impedance and $V I_{th}^{dc}$ is the input power dissipation at threshold. Consequently,

$$R_{th} = \frac{T_0}{V I_{th}^{dc}} \ln\left(\frac{I_{th}^{dc}}{I_{th}^{pul}}\right), \tag{13}$$

$$\approx \frac{T_0}{V I_{th}^{dc}}\left[\frac{I_{th}^{dc}}{I_{th}^{pul}} - 1\right], \tag{14}$$

since to a sufficient approximation, the exponential in (10) may be expanded as $e^x = 1 + x$. In general, we do not know the dc and pulsed thresholds with enough accuracy to use (11) or (12) credibly. For example, because of the opposite concavity of the dc and pulsed L-I curves near threshold (see previous footnote), it is often found that $I_{th}^{dc}/I_{th}^{pul} \lesssim 1$, a physically unreasonable result. When it is found that $I_{th}^{dc}/I_{th}^{pul} \gtrsim 1$, calculations using (11) or (12) sometimes show that R_{th} has decreased with degradation, while in other cases apparent increases are computed. The calculational uncertainty occurs because I_{th}^{dc}/I_{th}^{pul} exceeds unity by only a few percent; thus, an error of only a few percent in this ratio can easily alter the calculated value of R_{th} by a factor of two.

To more persuasively address the question of whether R_{th} decreases

and causes threshold degradation, we employed the crossing technique, discussed in Section 4.1, which we view as a substantially more error-free and credible approach for examining the behavior of R_{th} before and after aging. Prior to aging a particular laser, whose $P_p = 18$ mW in dc operation at 30°C, it was found that the 70°C pulsed L-I curve crossed the 30°C dc L-I curve at $I = 0.37$ ampere and $V = 1.9$ volts and crossed the 50°C dc L-I curve at $I = 0.22$ ampere and $V = 1.7$ volts. Using (1), it was computed that at these two crossings $R_{th} = 57$ and 53°C/W, respectively. After thoroughly degrading the laser so that P_p at 30°C was reduced a factor of 300, the above crossings occurred at $I = 0.365$ ampere and $V = 2.0$ volts, and $I = 0.228$ ampere and $V = 1.8$ volts. The after-aging values were determined to be 55 and 49°C/ W, respectively. The reasonable conclusion is that R_{th} was unaffected by the degradation. In any case, the indicated decrease in R_{th}, if real, would have an annealing rather than a degrading effect upon threshold or operating current.

7.2 Lasing-induced facet oxidation (erosion)

A lasing (forward bias)-induced facet erosion (oxidation) exhibits itself on (Al,Ga)As lasers, in its early stages, as an initial saturable mode of degradation.[41,55-58] The oxidation that occurs on the emissive portion of a laser facet can be suppressed if a dielectric (half-wavelength) film is deposited on the facet soon (~1 h) after cleaving.[41,58] If the film deposition is deferred for several days, the facet erosion is only mildly retarded.[41] The proposed explanation[41] for this inadequacy is related to the existence of a natural oxide (thickness ≈ 30 angstroms), which forms with a logarithmic time dependence[59,60] on the entire facet after cleaving. The oxide film, probably spongy, sandwiched between the GaAs and the dielectric coating is a likely conduit for the transportation of ambient oxidants to the emissive portion of a laser facet. If the coating is applied soon after cleaving, the natural oxide will be only ≈ 10 angstroms thick and a less permissive conduit. The rate of formation of the natural oxide on InP is the same as that for GaAs.[61] Therefore, if the lasing-induced oxidation were the same for InP as for GaAs, then quickly deposited coatings would be essential. Our present purge studies have been performed on uncoated lasers, hence our concern about this mechanism. For the following reasons, we believe that the initial saturable mode is not caused by facet erosion.

1. The initial saturable mode in the etched mesa BH lasers is extremely variable[4] in its potency and duration (e.g., see Figs. 8 and 10 through 14). A notable feature of degradation caused by facet erosion in (Al,Ga)As lasers is that for fixed powers and temperatures, the magnitude of the increases in operating current and the period of stabilization were approximately the same for all lasers.[41] The varia-

tions which we and others[4] observed are inconsistent with expectations about the operation of facet erosion based upon commonality of the following conditions for the aged lasers—same wafer, percent relative humidity, temperature, operating current, material composition, etc.

2. The initial saturable mode is also seen[23] to a comparable extent for packaged lasers that are in a hermetic can containing significantly less water vapor than a laboratory ambient.

3. Although preliminary experiments[4] with sputtered coatings indicated an apparent effectiveness in the suppression of the initial saturable mode, recent work[31] has shown that coatings are actually ineffective.

4. A recent study,[33] using an etched mesa laser structure very similar to ours, concluded that for operation at 30 mW/facet in a 25°C, room temperature ambient, an uncoated laser will suffer substantial degradation, $\Delta I_{op}/I_{op}(0) \geq 20$ percent in 250 hours; the cause was alleged to be facet erosion.[33] Since it has been shown[26] that the thickness of the lasing-induced oxidation $d \propto P$ (optical power), we might have expected to see operating currents increase by $\gtrsim 13$ percent in Fig. 8, where similar high power operation was monitored. Instead, none of the lasers in Fig. 8a show any credible increase at all in 250 hours, and only 1/5 lasers in Fig. 8b fulfills this expectation. Our explanation for the reported[33] degradation when $P = 30$ mW/facet, which takes into account the runaway or catastrophic appearance of end-of-life,[33] is that operation was too close to P_p, the rollover power (see Fig. 9), and any slight degradation (e.g., our initial saturable mode) caused P_p to fall toward and then drop below 30 mW/facet, at which point the current supply limit of 250 milliamperes[33] is rapidly attained.

Not only are we convinced that the initial saturable mode is not induced by facet erosion, but we are also persuaded that within the system lifetime (2.2×10^5 hours at 10°C), lasing-induced facet oxidation will very likely never promote degradation that is significant enough to cause premature failure; thus, uncoated lasers may be used even in high-reliability applications. The reasons are the following:

1. Uncoated lasers and Al_2O_3-sputtered coated lasers of a structure identical to ours were aged at 15 mW/facet and 25°C for 4000 hours in a laboratory ambient (<30-percent RH) and a dry nitrogen ambient; there were no significant differences in aging behavior.[26]

2. The thickness of the lasing-induced facet oxidation is proportional to the logarithm of time,[26] unlike the square-root time dependence found for AlGaAs/GaAs, so that in a 25°C, 30-percent RH ambient, operation at 5 mW/facet will produce only ~70 angstroms of oxide in 2×10^5 hours, a thickness insufficient to promote degradation related to optical scattering, reduced facet reflectivity, or surface current leakage.

3. This is supported by the observation[31] that an uncoated laser aged at 50°C and 5 mW/facet for 6000 hours showed only a native oxide layer ($\approx$30 angstroms) at the emissive portion of the facets.

4. In another study,[25] it was determined that the time for $I_{th}^{pul}/I_{th}^{pul}(0)$ to reach 1.5, because of facet erosion, was $\approx 10^7$ hours for an InGaAsP/InP laser operating in a 70°C, 1 to 10-percent RH ambient at a power output $\sim$5 mW/facet. In a hermetic can with reduced water vapor at 10°C, this time would be considerably longer.

7.3 Dark-spot defects and gold in-migration

A predominant mode of degradation in InGaAsP/InP DH LEDs[7,10-15] and lasers[16-18] is related to the formation of Dark-Spot Defects (DSD), although instances of Dark-Line Defect (DLD) formation have been reported.[10,16-18,62] The early manifestation of the degradation associated with the nonradiative DSD and DLD appears as an initial saturable mode in lasers.[16-18,62] The crystalline defects connected with DSD are precipitate-like[7,10,13,63] and recently were found[12,13,15] to contain gold (Au), which had migrated into the vicinity of the active region from the p-metal contact.

A qualitative description of the Au in-diffusion and degradation has been given[13,15] as follows. During the p-contact alloying procedure, nonuniformities in the alloyed region are formed, which can have a spikelike appearance.[64] Under forward bias, the current density is larger in these regions in which the contact metals, especially Au, have preferentially penetrated. Electromigration and thermomigration (promoted by local ohmic heating) cause even deeper local penetration of Au. DSD, formed when the Au reaches the active region, can act as optical and current (nonradiative recombination) losses.[13,15,18] Local heating has been observed at DSD in the active regions of lasers.[17] Outside the active region,* in the nearby shunt paths of an etched mesa BH laser, the presence of Au may also cause local heating, ohmic and nonradiative; electrical shorting of the blocking junctions, which promote increased shunt current flow independent of heating, is also possible. Any increase in local temperature, at a fixed terminal current, whether the defect site is inside or just outside the active region, will promote an increase in shunt current (by virtue of Fig. 5a), i.e., degradation, since the terminal current must be increased to preserve the intended optical output. Consistent with our view that laser-

---

* Within the active region, the precipitate-like defects, caused by local Au deposits, are manifested as DSD in an electroluminescent view of that region; equivalent deposits outside the active region, i.e., in the shunt path that borders the active region, do not have such a visible manifestation and so cannot be called DSD. Our discussions use the term DSD because that describes how the defects are seen when they are within a strongly luminescing region.

degradation-promoting defects exist in the shunt path and can be caused by exclusive shunt path operation are the observations[31,33] that initial degradation in lasers of our structure is not necessarily accompanied by any darkening of the active region.

The existence of random local invasions of Au from the p-contact is consistent with the observations of DSD made by others. After an initial period, the number of DSD in laser operation remained fixed, if they occurred at all, and this number-saturation coincided with the knee-formation or degradation stabilization process.[18] In LEDs, it has been reported that both the number and size of DSD increased with time.[11,13] Evidence of a darkening of a portion of the active region in etched mesa BH lasers has been associated[4] with degradation in laser performance. The time for the first appearance of a DSD is strongly current density dependent.[10,18]

A certain number of observations about the initial transient mode and its stabilization are consistent with the idea of randomly located spikelike regions in which remotely situated nonradiative centers can electromigrate to the vicinity of the active region. The following are some examples: (1) On occasion double knees may occur in the course of T/I stressing, which could correspond to the sequential establishment of two distinct and locally separate diffusion fronts. (2) Some lasers exhibit no evidence of an initial saturable mode, and for those that do, the manifestations are quite varied, i.e., both the increases in operating current required for stabilization and the times-of-stabilization can span a range of values. (3) Figure 14a displays "incubation" periods prior to the occurrence of the initial saturable mode of degradation, which strongly suggests that little or no degradation can take place until some in-migrating degradation-causing centers reach the active region,* or its vicinity, and that once there, the deposited centers promote a locally confined degradation (increase in nonradiative current), which becomes self-limiting, because each center can cause only so much additional current loss. (4) After substantial degradation (induced by extensive T/I and P/I stressing), a laser may operate pulsed but not CW at, e.g., 70°C; since an increase in thermal impedance seems ruled out (see Section 7.1.2), local heating, produced by defects that did not exist prior to aging and that cause substantial current shunting in dc operation, seems a likely explanation. (5) To support this contention of additional heating at the same value of terminal current after aging, the crossing point method (see Section 4.1) was used to show that, e.g., the 70°C pulsed L-I curve always crossed the 50°C dc L-I curve at a lower value of current after degrada-

---

* The existence of the incubation period explains why all of the lasers of Fig. 14a passed the initial burn-in tests (see Section 4.2).

tion had occurred. (6) In several instances a thermal annealing effect, which is present in the absence of forward bias, has been observed (e.g., Section 3.2), and in one case the threshold of a degraded laser was decreased by ~10 mA after storage aging at 175°C for 60 hours; annealing is consistent with a movement or homogenization of a degrading species.

Still another point of comparison is the activation energy for the initial saturable mode. We have determined (see Section 4.7) that $(E_a)_i$ = 1.2–1.5 eV for the initial mode, while others[23] have found $(E_a)_i$ = 0.54 eV. In an LED study,[10] the activation energy associated with the first appearance of a DSD was found to be $(E_a)_{DSD}$ = 1.2 eV, while a laser study[18] found $(E_a)_{DSD}$ = 0.16 eV. The substantial variations, due perhaps to the uncertainties caused by using the isothermal aging comparison technique, prevent any credible conclusion.

Another area of apparent concern is why elevated temperatures alone, without forward bias, do not produce degradation, when diffusion calculations indicate that they should. For the interstitial diffusion of Au in InP, the distance traveled is related to the diffusion constant and time by

$$x = \sqrt{Dt}, \tag{15}$$

where[65]

$$D = 1.32 \times 10^{-5} \text{cm}^2\text{s}^{-1} \exp\left[-\frac{0.48 \text{ eV}}{kT}\right]. \tag{16}$$

Since the p-contact annealing (350°C for 10 minutes) is known to cause Au to penetrate the Cr barrier layer,[66] our storage aging (no bias, 175°C for 4000 hours) (see Section 3.2) should have allowed Au to penetrate a distance x = 280 μm, more than enough to cause degradation of the active region. No degradation occurred, however, which is in accord with storage aging experiments on quaternary LEDs (230°C for 2000 hours)[7] and lasers (100°C for 2000 hours).[33] Assuming that the constants of (16), which were determined at 600 to 820°C, are validly extrapolated to lower temperatures, it would appear that Au probably does diffuse large distances, however, in concentrations so low[65] (~10^{15} cm^{-3}) that no degradation is produced. With the application of forward bias, however, many orders of magnitude larger concentrations of Au are found[12,13,15] to have diffused inward, perhaps into vacancies created during annealing; this may correspond to the more shallow diffusion and distinctly different process for the migration of Au into InP, which was not characterized, but found.[65] (When Au/Ge contacts on GaAs are annealed with constant current, filaments of the contact material are found in the substrate. Annealing under

identical circumstances without current flowing does not produce contact penetration.[67])

In conclusion, we view the in-migration of Au (or perhaps some other contact element) from the p-contact region as the explanation, which best fits our observations and those of others,[23] for the initial saturable mode.

VIII. CONCLUSION

The results of our experimental study of the degradation characteristics of the 1.3-μm etched mesa BH laser may be summarized as follows.

1. Virtually all unscreened lasers possess to some degree an initially occurring mode of degradation that stabilizes with aging, i.e., the observed rate of degradation decreases in some time period from an initially high value to a constant lower value.

2. At all temperatures of interest, the CW light-current ($L-I$) curves are sublinear starting at threshold. As the terminal current is increased, an increasing fraction of that current is diverted into a shunt path around the active region; two current paths are always operative under lasing conditions. The CW $L-I$ curves show rollover behavior, i.e., at a fixed ambient temperature, an increase in the direct current produces a sublinear increase in output power until a maximum (rollover) power is reached, and for still larger currents the output power decreases until lasing is extinguished at some upper threshold current ($\sim$1 Adc). Above this point all injected current goes into a shunt path. If the injection current is held constant, then an increasing ambient temperature will cause an increasing fraction of this current to flow into a shunt path. At a high enough temperature ($\gtrsim$100°C) all current will flow into the shunt path.

3. In a long-duration ($\sim$3 kh), elevated temperature (60°C) burn-in (3 mW/facet), the initial mode will stabilize in 0.3 to 1 kh; in any lasing regime, some fraction of the terminal current will always flow in a shunt path. The initial mode will, however, stabilize more rapidly ($\sim$10 hours) (as observed by periodic monitoring at a temperature at which lasing occurs) in a high-temperature (150°C), high-current (250 mAdc) purge regime in which the current flows exclusively in a shunt path; there is no optical output in this regime. Thus, it is reasonable to conclude that the initial mode stabilization, which is often accompanied by a considerable fixed degree of degradation in laser performance, is caused by current flowing in a shunt path around the active region, whether or not the device is operated as a laser.

4. If, instead, it were concluded[23] that during *initial lasing* operation, prestabilization degradation occurred in the active region, while in an *initially* imposed T/I regime, the prestabilization degradation occurred

in the shunt path, then it would be prudent to refrain from any lasing operation until T/I purging compelled stabilization; it might prove beneficial to long-term aging if the relatively large prestabilization degradation were forced to occur outside the active region, rather than within. If it were shown to be beneficial, then the purge would actually produce a better laser than would an elevated temperature burn-in.

5. If the high-current, high-temperature purge regime is continued beyond the point of stabilization, then there is a long-term temporally linear degradation of the shunt path, which has a thermal activation energy, $(E_a)_{sh} \approx 0.85$ eV. The current is the driving mechanism for the degradation, and the degradation appears to be associated with a shunt path.

6. Thus, it is possible that when a purged (stabilized) device is operated as a laser, the always-existing shunt path fraction of the terminal current is the (or at least one of the) long-term wear-out degradation driving forces.

7. The degradation that occurs during the purge overstress aging may be considerably in excess of any subsequent degradation that would be experienced in the course of the system lifetime (25 years) of a laser operating at ocean bottom temperature. Alternately put, most of the degradation that would occur in an *unscreened, unpurged* laser during a system lifetime is compelled to occur "up front" by the purge.

8. Because the initial saturable mode has an incubation period, unscreened lasers can easily "pass the test" ($\Delta I/I \leq 5$ percent) of a conventional short-duration (10^2-hour), elevated temperature (60°C) burn-in (5 mW/facet). Thereafter, as we have witnessed, the degradation of a laser that is purge-induced, and which would occur eventually no matter how the device was operated, can be large enough to force a rejection of that laser as a subcable candidate. It is also possible that such a routine burn-in would cause a rejection of a laser that actually might have been a suitable subcable candidate, once stabilization had occurred. However rational short-duration burn-in aging might appear as a screening procedure to assure long-term reliability, it proved to be inadequate.

9. The unrecognized (by a conventional burn-in screen) existence of the initial saturable mode of degradation might have doomed the lightwave subcable project, either because of a significant number of postinstallation failures or because of the chaos and delay that its variable effects would have produced in making reliable lifetime predictions. Particular adverse consequences of using *unstabilized* lasers, i.e., survivors of a conventional burn-in, are:

    a. As long as the initial mode, which exists in varying degrees in all lasers, is controlling, it is impossible to determine credible rates

of degradation at any elevated lasing temperature for the long-term, wear-out mode.

b. As a result, no reliable estimate of activation energy for the wear-out mode may be made.

c. Consequently, lifetime predictions for lower use temperatures are impossible.

d. If the initial mode is mistaken for the long-term mode, the actual superlinear followed by sublinear increases in current as a function of time will result in an enormous spread in activation energies as determined by step temperature aging of individual lasers; whether the activation energy is large or small will depend upon the direction of the step, i.e., higher to lower temperatures, or the reverse.

e. A mistaken identification of modes, coupled with too small an apparent activation energy, would doom a potential system because of the excessive costs connected with redundancy and/or anticipated replacement. Similarly, the use of an erroneously large activation energy would doom a system after installation because the devices could fail substantially prior to the system lifetime.

f. A mistaken identification of modes might also reject perfectly good lasers, indeed doom a system, because the perceived rates of degradation would be too large; appearances might have been deceiving because the initial rapid rates would eventually stabilize at some lower entirely acceptable values.

g. A failure to detect the initial saturable mode might also cause excessive field failures because an initially occurring 50-percent increase in operation current, followed by another 50-percent wear-out increase over much longer times, could exceed the capacity of the transmitter compensator circuit. Such would not have been true, e.g., if a purge strategy were used to compel a timely stabilization that was accompanied by the initial 50-percent increase, because this would have occurred before the transmitter parameters were set up.

10. The purge compelled a timely (~10-hour) stabilization of the initial mode of degradation; thereafter, access to the long-term mode permitted its activation energy and temperature-activated degradation rates to be credibly determined so that reliable lifetime estimates could be made.

11. The long-term degradation rates have been found to be so low that, with some confidence, a subcable lightwave system using these lasers may be contemplated with either an extremely conservative definition for device failure or with a considerably smaller activation energy than has actually been measured.

12. The best guess about the origin of the initial saturable mode of degradation relates to the in-migration of Au, or some other p-contact metal or impurity.

IX. ACKNOWLEDGMENTS

REFERENCES

1. M. Hirao et al., "Fabrication and Characterization of Narrow Stripe InGaAsP/InP Buried Heterostructure Lasers," J. Appl. Phys., 51, No. 8 (August 1980), pp. 4539–40.
2. K. Mizuishi et al., "Accelerated Aging Characteristics of InGaAsP/InP Buried Heterostructure Lasers Emitting at 1.3 μm," Jap. J. Appl. Phys., 19, No. 7 (July 1980), pp. L429–32.
3. R. J. Nelson et al., "CW Electrooptical Properties of InGaAsP (λ = 1.3 μm) Buried Heterostructure Lasers," IEEE J. Quantum Electron., QE-17, No. 2 (February 1981), pp. 202–7.
4. K. Mizuishi et al., "Reliability of InGaAsP/InP Buried Heterostructure Lasers," Jap. J. Appl. Phys., 21, Supplement 21-1 (1982), pp. 359–64.
5. M. Ettenberg and C. J. Neuse, "Reduced Degradation in $In_xGa_{1-x}As$ Electroluminescent Diodes," J. Appl. Phys., 46, No. 5 (May 1975), pp. 2137–42.
6. S. Yamakoshi et al., "Degradation of High-Radiance $Ga_{1-x}AlAs$ LEDs," Appl. Phys. Lett., 31, No. 9 (November 1977), pp. 627–9.
7. S. Yamakoshi et al., "Reliability of High Radiance InGaAsP/InP LEDs Operating in the 1.2–1.3 μm Wavelength," IEEE J. Quantum Electron., QE-17, No. 2 (February 1981), pp. 167–73.
8. E. I. Gordon, F. R. Nash, and R. L. Hartman, "Purging: A Reliability Assurance Technique for New Technology Semiconductor Devices," IEEE Electron Dev. Lett., EDL-4, No. 12 (December 1983), pp. 465–6.
9. F. R. Nash et al., "Selection of a Laser Reliability Assurance Strategy for a Long-Life Application," AT&T Tech. J., this issue.
10. S. Yamakoshi et al., "Degradation of High Radiance InGaAsP/InP LEDs at 1.2–1.3 μm Wavelength," IEDM Technical Digest, Washington, D.C., 1979, p. 122.
11. Yeats et al., "Performance Characteristics and Extended Lifetime Data for InGaAsP/InP LEDs," IEEE Electron Dev. Lett., EDL-2, No. 9 (September 1981), pp. 234–6.
12. A. K. Chin et al., "Cathodoluminescence Evaluation of Dark Spot Defects in InP/InGaAsP Light-Emitting Diodes," Appl. Phys. Lett., 41, No. 6 (September 1982), pp. 555–7.
13. O. Ueda et al., "Mechanism of Catastrophic Degradation in InGaAsP/InP Double-Heterostructure Light Emitting Diodes and GaAlAs Double-Heterostructure Light Emitting Diodes With Pulsed Large Current," J. Appl. Phys., 53, No. 12 (December 1982), pp. 9170–9.
14. C. L. Zipfel, A. K. Chin, and M. A. DiGiuseppe, "Reliability of InGaAsP Light Emitting Diodes at High Current Density," IEEE Trans. Electron Dev., ED-30, No. 4 (April 1983), pp. 310–6.
15. A. K. Chin et al., "The Migration of Gold From the p-Contact as a Source of Dark Spot Defects in InP/InGaAsP LEDs," IEEE Trans. Electron Dev., ED-30, No. 4 (April 1983), pp. 304–10.
16. T. Yamamoto, K. Sakai, and S. Akiba, "10,000-h Continuous CW Operation of $In_{1-x}Ga_xAs_yP_{1-y}$/InP DH Lasers at Room Temperature," IEEE J. Quantum Electron., QE-15, No. 8 (August 1979), pp. 684–7.
17. M. Fukuda, K. Wakita, and G. Iwane, "Observation of Dark Defects Related to Degradation in InGaAsP/InP DH Lasers under Accelerated Operation," Jap. J. Appl. Phys., 20, No. 2 (February 1981), pp. L87–90.
18. M. Fukuda, K. Wakita, and G. Iwane, "Dark Defects in InGaAsP/InP Double

Heterostructure Lasers Under Accelerated Aging," J. Appl. Phys., *54*, No. 3 (March 1983), pp. 1246–50.
19. R. W. Dixon, "Derivative Measurement of Light-Current-Voltage Characteristics of (Al,Ga)As Lasers," B.S.T.J. *55*, No. 7 (September 1976), pp. 973–80.
20. T. L. Paoli, "Nonlinearities in the Emission Characteristics of Stripe-Geometry (AlGa)As Double-Heterostructure Junction Lasers," IEEE J. Quantum Electron., *QE-12*, No. 12 (December 1976), pp. 770–6.
21. D. C. Craft, N. K. Dutta, and W. R. Wagner, private communication.
22. N. K. Dutta and D. C. Craft, private communication.
23. B. W. Hakki, P. E. Fraley, and T. F. Eltringham, "1.3 μm Laser Reliability Determination for Submarine Cable Systems," AT&T Tech. J., this issue.
24. M. Fukuda and K. Wakita, "Degradation of InGaAsP/InP DH Lasers in Water Due to Facet Erosion," Jap. J. Appl. Phys., *19*, No. 11 (November 1980), pp. L667–70.
25. M. Morimoto and M. Takusagawa, "Accelerated Facet Degradation of InGaAsP/InP Double-Heterostructure Lasers in Water," J. Appl. Phys., *53*, No. 6 (June 1982), pp. 4028–37.
26. M. Fukuda, "Facet Oxidation of InGaAsP/InP and InGaAs/InP Lasers," IEEE J. Quantum Electron., *QE-19*, No. 11 (November 1983), pp. 1692–8.
27. K. Mizuishi et al., "Reliability of InGaAsP/InP Buried-Heterostructure 1.3 μm Lasers," IEEE J. Quantum Electron, *QE-19*, No. 8 (August 1983), pp. 1294–1301.
28. K. Fujiwara et al., "Aging Characteristics of Ga$_{1-x}$Al$_x$As Double-Heterostructure Lasers Bonded With Gold Eutectic Alloy Solder," Appl. Phys. Lett., *34*, No. 10 (May 1979), pp. 668–70.
29. H. Imai et al., "Aging Tests of Low Threshold Current InGaAsP/InP V-Grooved Substrate Buried Heterostructure Lasers Emitting at 1.3 μm Wavelength," Appl. Phys. Lett., *41*, No. 7 (October 1982), pp. 583–4.
30. H. Imai et al., "Accelerated Aging Test of InGaAsP/InP Double-Heterostructure Laser Diodes With Single Transverse Mode," Appl. Phys. Lett., *38*, No. 1 (January 1981), pp. 16–7.
31. A. K. Chin, F. Ermanis, and F. R. Nash, private communication.
32. M. Hansen, *Constitution of Binary Alloys*, New York: McGraw-Hill, 1958, p. 1107.
33. M. Fukuda et al., "Facet Degradation and Passivation of InGaAsP/InP Lasers," Appl. Phys. Lett., *41*, No. 1 (July 1982), pp. 18–21.
34. F. Bosch, private communication.
35. P. D. Wright, W. B. Joyce, and D. C. Craft, "Electrical Derivative Characteristics of InGaAsP Buried Heterostructure Lasers," J. Appl. Phys., *53*, No. 3 (March 1982), pp. 1364–72.
36. Y. Nakano et al., "Output Power Saturation of BH Laser Under High Current Operation," Electron. Lett., *18*, No. 12 (June 1982), pp. 501–2.
37. H. Namizaki et al., "Shunt Current and Excess Temperature Sensitivity of I_{th} and η_{ex} in 1.3 μm InGaAsP DH Lasers," Electron. Lett., *18*, No. 16 (August 1982), pp. 703–4.
38. N. K. Dutta et al., private communication.
39. D. A. Ackerman et al., private communication.
40. B. W. Hakki and F. R. Nash, "Catastrophic Failure in GaAs Double-Heterostructure Injection Lasers," J. Appl. Phys., *45*, No. 9 (September 1974), pp. 3907–12.
41. F. R. Nash and R. L. Hartman, "Accelerated Facet Erosion Formation and Degradation of (Al,Ga)As Double-Heterostructure Lasers," IEEE J. Quantum Electron., *QE-16*, No. 10 (October 1980), pp. 1022–33.
42. J. S. Bora and A. K. Babar, "Simplification of Base Failure Rate Models," Microelectron. Reliab., *20*, No. 4 (1980), p. 535.
43. W. B. Joyce, R. W. Dixon, and R. L. Hartman, "Statistical Characterization of the Lifetimes of Continuously Operated (Al,Ga)As Double-Heterostructure Lasers," Appl. Phys. Lett., *28*, No. 11 (June 1976), pp. 684–6.
44. W. B. Joyce et al., "Methodology of Accelerated Aging," AT&T Tech. J., this issue.
45. H. Kressel, M. Ettenberg, and I. Ladany, "Accelerated Step-Temperature Aging of Al$_x$Ga$_{1-x}$As Heterojunction Laser Diodes," Appl. Phys. Lett., *32*, No. 5 (March 1978), pp. 305–8.
46. D. S. Peck and C. H. Zierdt, "The Reliability of Semiconductor Devices in the Bell System," IEEE Proc., *62*, No. 2 (February 1974), pp. 185–211.
47. F. H. Reynolds, "Thermally Accelerated Aging of Semiconductor Components," IEEE Proc., *62*, No. 2 (February 1974), pp. 212–22.
48. T. Ikegami et al., "Stress Tests on 1.3 μm Buried Heterostructure Laser Diode," Electron. Lett., *19*, No. 8 (April 1983), pp. 282–3.
49. Y. Nakano et al., "Thermally Accelerated Degradation of 1.3 μm BH Lasers," Electron. Lett., *19*, No. 15 (July 1983), pp. 567–8.

50. J. Shimer and P. E. Fraley, unpublished work.
51. J. R. Pawlik, "Reduced Temperature Dependence of Threshold of (Al,Ga)As Lasers Grown by Molecular Beam Epitaxy," Appl. Phys. Lett., *38,* No. 12 (June 1981), pp. 974–6; see references within.
52. K. Fujiwara et al., "Analysis of Deterioration in In Solder for GaAlAs Lasers," Appl. Phys. Lett., *35,* No. 11 (December 1979), pp. 861–3.
53. R. G. Plumb, A. R. Goodwin, and R. S. Baulcomb, "Thermal-Impedance Ageing Characteristics of C. W. Stripe Lasers," Solid-State Electron. Dev., *3,* No. 6 (November 1979), pp. 206–9.
54. S. Ritchie et al., "The Temperature Dependence of Degradation Mechanisms in Long-Lived (GaAl)As DH Lasers," J. Appl. Phys., *49,* No. 6 (June 1978), pp. 3127–32.
55. C. J. Hwang, "Initial Degradation Mode of Long-Life (GaAl)As-GaAs Lasers," Appl. Phys. Lett., *30,* No. 3 (February 1977), pp. 167–9.
56. Y. Shima, N. Chinone, and R. Ito, "Effects of Facet Coatings on the Degradation Characteristics of GaAs-Ga$_{1-x}$Al$_x$As DH Lasers," Appl. Phys. Lett., *31,* No. 9 (November 1977), pp. 625–7.
57. H. Yonezu et al., "Mirror Degradation in AlGaAs Double-Heterostructure Lasers," J. Appl. Phys., *50,* No. 8 (August 1979), pp. 5150–7.
58. F. R. Nash et al., "Stabilization of Aging-Induced Self-Pulsations and the Elimination of an Initial Temporally Saturable Mode of Degradation in (Al,Ga)As Lasers by Means of Facet Coatings," Appl. Phys. Lett., *35,* No. 12 (December 1979), pp. 905–7.
59. F. Lukes, "Oxidation of Si and GaAs in Air at Room Temperature," Surf. Sci., *30,* No. 1 (March 1972), pp. 91–100.
60. E. Kuphal and H. W. Dinges, "Composition and Refractive Index of Ga$_{1-x}$Al$_x$As Determined by Ellipsometry," J. Appl. Phys., *50,* No. 6 (June 1979), pp. 4196–200.
61. A. J. Rosenberg, "The Oxidation of Intermetallic Compounds: III. The Room-Temperature Oxidation of A^{III}B^V Compounds," J. Phys. Chem. Solids, *14* (1960), pp. 175–80.
62. K. Endo et al., "Rapid Degradation of InGaAsP/InP Double Heterostructure Lasers Due to <110> Dark Line Defect Formation," Appl. Phys. Lett., *40,* No. 11 (June 1982), pp. 921–3.
63. O. Ueda et al., "Transmission Electron Microscope Observation of Dark-Spot Defects in InGaAsP/InP Double-Heterostructure Light-Emitting Diodes Aged at High Temperature," Appl. Phys. Lett., *36,* No. 4 (February 1980), pp. 300–1.
64. C. C. Chang et al., "Transmission Electron Microscopy of Au-Based Ohmic Contacts to n-Al$_x$Ga$_{1-x}$As," J. Appl. Phys., *50,* No. 11 (November 1979), pp. 7030–3.
65. S. I. Rembeza, "Diffusion of Au in InP Single Crystals," Sov. Phys. Semicond., *3,* No. 4 (October 1969), pp. 519–20.
66. P. R. Skeath, unpublished work.
67. C. J. Palmström, D. V. Morgan, and M. J. Howes, "Contact Degradation of GaAs Transferred Electron Devices," Nucl. Instrum. Methods, *150,* No. 2 (April 1978), pp. 305–11.

AUTHORS

David A. Ackerman, A.B. (Physics), 1976, Cornell University; M.S. and Ph.D. (Physics), University of Illinois at Urbana-Champaign, in 1978 and 1982, respectively; AT&T Bell Laboratories, 1982—. Since 1982 Mr. Ackerman has been working at AT&T Bell Laboratories in the area of semiconductor lasers. Member, APS, Sigma Xi.

Richard W. Dixon, A.B. (Engineering and Applied Physics), 1958, Harvard College; M.S. and Ph.D. (Engineering and Applied Physics), Harvard University, in 1960 and 1964, respectively; AT&T Bell Laboratories, 1965—. Mr. Dixon is currently the Director of the Lightwave Devices Laboratory of AT&T Bell Laboratories.

Niloy K. Dutta, B.Sc. (Physics, Honours), 1972, and M.Sc. (Physics), 1974, St. Stephen's College, New Delhi, India; Ph.D. (Physics), 1978, Cornell University, Ithaca, NY; AT&T Bell Laboratories, 1979—. Mr. Dutta had a

Postdoctoral Associate at Cornell University for one and a half years. During his doctoral and postdoctoral work at Cornell, he worked on spin-flip Raman lasers, infrared spectroscopy, soft x-ray lasers, and nonlinear optics with nonmonochromatic waves. Since 1979 he has been a Member of Technical Staff at AT&T Bell Laboratories, where he has contributed extensively to the research and development of InGaAsP semiconductor lasers. He has authored or coauthored more than 60 publications on his work. Member, IEEE, American Physical Society, Optical Society of America.

Robert L. Hartman, B.A. (Mathematics), 1958, Columbia College; B.S. (Engineering Physics), M.S. and Ph.D (Physics), Columbia University, in 1959, 1961, and 1968, respectively; AT&T Bell Laboratories, 1968—. At AT&T Bell Laboratories, Mr. Hartman is presently engaged in studying the physics of III-V compound semiconductor devices, particularly InGaAsP/InP and (Al,Ga)As-GaAs double-heterostructure injection lasers. He has contributed extensively to the understanding of the degradation mechanisms in these devices and authored or coauthored several key papers in this field, especially those involving the discovery and analysis of dark-line defects; the discovery of strain effects upon laser reliability; the attainment of long operating lifetimes in LEDs and injection lasers; the first measurement of the thermal acceleration of injection laser operating lifetime; and effects of surfaces on LED and laser reliability. In 1980 Mr. Hartman was appointed Supervisor of the Semiconductor Lightwave Lasers Group. Member, Sigma Xi, American Physical Society.

Franklin R. Nash, B.S. (Physics), 1955, Polytechnic Institute of New York; Ph.D. (Physics), 1962, Columbia University; AT&T Bell Laboratories, 1963—. At AT&T Bell Laboratories, Mr. Nash has been involved in research on gas lasers (modulation and mode locking), detection of subnanosecond pulses using photodiodes, nonlinear optical materials, parametric oscillators, neoclassical theory of radiation, and semiconductor lasers (mode guidance, material defects, electro-optic anomalies, and degradation mechanisms). Currently, he is engaged in the reliability assessment of long-wavelength semiconductor lasers. Member, American Physical Society, IEEE.

Jonathan R. Pawlik, B.S.E. (Engineering Physics) and M.S. (Physics), University of Michigan, in 1969 and 1971, respectively; Ph.D. (Applied Physics), Harvard University; AT&T Bell Laboratories, 1979—. While at Harvard University, Mr. Pawlik studied the infrared and magnetic properties of hydrogenated amorphous silicon and germanium. As a Member of Technical Staff at AT&T Bell Laboratories, he has been involved in InGaAsP and GaAlAs injection laser characterization.

William J. Sundburg, B.S.E.E., 1968, Newark College of Engineering; M.S.E.E., 1972, Rutgers University; AT&T Bell Laboratories, 1959—. From 1959 to 1965, Mr. Sundburg participated in experiments studying surface phenomena of semiconductors materials. From 1965 to 1969, Mr. Sundburg studied oxide growth on single crystal metals. From 1969 to 1972, he established processing techniques for fabricating single cell MOS nonvolatile memories. He designed integrated memory circuits utilizing nonvolatile memory cells from 1972 to 1977. Mr. Sundburg worked on the design of custom logic large-scale integrated circuits from 1977 to 1981. Since 1981 he has worked in the laser reliability group studying modes of degradation of semiconductor lasers.

Reliability of InGaAs Photodiodes for SL Applications

By R. H. SAUL,* F. S. CHEN,* and P. W. SHUMATE, JR.†

(Manuscript received March 6, 1984)

A major objective of the TAT-8 submarine cable development program is the reliability assurance of its critical components. A relatively new device, the InGaAs photodiode, is used as the detection element in optical receivers and as the monitor of laser output in optical transmitters. In this paper, we describe a comprehensive reliability program aimed at assessing and assuring the reliability of InGaAs photodiodes. A major portion of this work has involved device operation at overstress conditions. Results to date indicate that for receiver photodiodes a device design exists which is predicted to meet a 1-FIT reliability objective. Tests of monitors of similar design are in progress. Critical to the success of the reliability assurance program is the ability to identify weak devices which are likely to fail early in the system life cycle. A conventional high-temperature burn-in is shown to be impractical, and it would not necessarily remove devices which fail by low activation energy processes. Overvoltage provides a means of accelerating such failure mechanisms. A "purge" using a combination of accelerants (high temperature and overvoltage) is shown to have considerable promise in failing weak devices, while inducing significant changes in those devices which subsequently fail early in a life test (elevated temperature at normal bias conditions). Devices which are unaffected by the purge are shown by extended life tests to be robust devices which have a high probability of meeting TAT-8 requirements.

I. INTRODUCTION

Photodiodes fabricated with an InGaAs absorbing region have high

---

* AT&T Bell Laboratories. † AT&T Bell Laboratories; present affiliation Bell Communications Research, Inc.

---

quantum efficiency over the 1.0- to 1.6-μm wavelength region. These devices are used in all long-wavelength lightwave transmission systems being developed by AT&T Bell Laboratories. The Submarine Cable (SL) application utilizes InGaAs photodiodes as the photosensitive element in receivers and as monitors of laser output for power stabilization in transmitters. The high repair costs associated with an undersea installation put great emphasis on assuring that the reliability requirement of each critical component is met. While initial reliability information on InGaAs photodiodes, either generated in the laboratory[1-3] or through our terrestrial applications,[4] led us to be optimistic about the long-term stability of these devices, it was clear that a comprehensive reliability evaluation was required. In addition, although there is extensive AT&T experience with short-wavelength (silicon) photodiodes in FT3 lightwave systems, the photodiode screening procedures for assuring reliability are not directly applicable because of substantial differences in manufacturing technologies and semiconductor properties.

The first phase of our reliability program involved substantial testing of InGaAs photodiodes under various overstress conditions of temperature, humidity, bias, etc. Coupled with failure mode analysis, this phase revealed design and fabrication flaws, whose subsequent elimination led to devices with enhanced reliability. Extensive *qualification* testing then established that the design met system reliability objectives. Since in manufacture it is expected that some devices will be flawed and will therefore be susceptible to premature failures, a *certification* process is required which "purges" out the flawed devices, leaving a surviving population of robust devices which have a high probability of meeting SL reliability objectives. A *surveillance* phase of this process involves the accelerated testing of a portion of the survivors to establish that the manufactured devices match the reliability statistics of the previously qualified devices. The remaining survivors are then incorporated into receivers or laser packages and subjected to further extended operational testing. Final selection of receivers or laser packages for undersea installation is then made. Long-term tests are continued on a portion of these packages to reveal the true reliability statistics of the manufactured population; the results help determine maintenance strategy.

Key to the success of the certification process is the development of a purge procedure. In general this procedure will include thermomechanical tests (temperature cycling, mechanical shock, vibration, hermeticity, etc.) to weed out packaged devices that have assembly flaws. Additionally, and most importantly, the photodiode chips must be operated under carefully selected overstress conditions which promote failure or cause detectable instabilities or changes in devices that

contain manufacturing flaws. It is desirable that the life of the surviving (robust) devices should not be deleteriously affected, however. The critical issue is to assure that devices which survive the purge procedure do not fail during the service life of the system. Ideally, the type of stress and the failures induced should be the same as those encountered in normal operation so that relevant failure modes are stimulated. However, a variety of stresses can be employed to eliminate all but the most robust devices. We have found that simultaneous application of high temperature and high reverse bias is a promising method for achieving the desired objectives.

In the following section we review the photodiode device design from the standpoint of reliability. Critical device parameters and potential failure modes are discussed in Section III. Results of accelerated aging tests, used to qualify the device design, are described in Section IV. The development of a purge is then discussed in Section V.

II. REVIEW OF DEVICE DESIGN: CONSIDERATIONS FOR RELIABILITY

Initial InGaAs photodiodes were fabricated in a mesa geometry.[5-7] These devices were simple to fabricate and capable of low dark currents. While stable operation has been reported for several thousand hours in limited samples,[2,3] we believe that long-term device stability is compromised in the mesa geometry since the junction perimeter is exposed. Contaminants on the mesa walls can result in surface leakage; the slow ingress of contaminants along the p-n junction can also increase generation-recombination current. Planar photodiodes, on the other hand, are more amenable to surface passivation which seals the junction and (ideally) stabilizes surface states. Figure 1 schematically illustrates the planar InGaAs photodiodes developed at AT&T Bell Laboratories.[8] Device characteristics for SL receiver and monitor applications are given in Table I. Several design features are noteworthy. Back-face illumination[6] eliminates carrier loss due to surface recombination, thus enhancing quantum efficiency. A quarter-wave SiN_x film on the substrate side provides an Antireflection (AR) coating. On the junction side of the chip a double dielectric layer is used: the first SiN_x layer serves as a mask for selected area diffusion; the second SiN_x layer defines the p-contact area and serves as an additional passivation layer which completes the sealing of the surface. The entire junction area is therefore protected from the ambient. Further, the metal contact to the semiconductor surface is restricted so that the stress, associated with the metallization-semiconductor interface, is removed from the junction perimeter. Initially, plated p-contacts were used; however, bias aging at high humidity showed a fivefold improvement in mean times to failure with e-beam evaporated
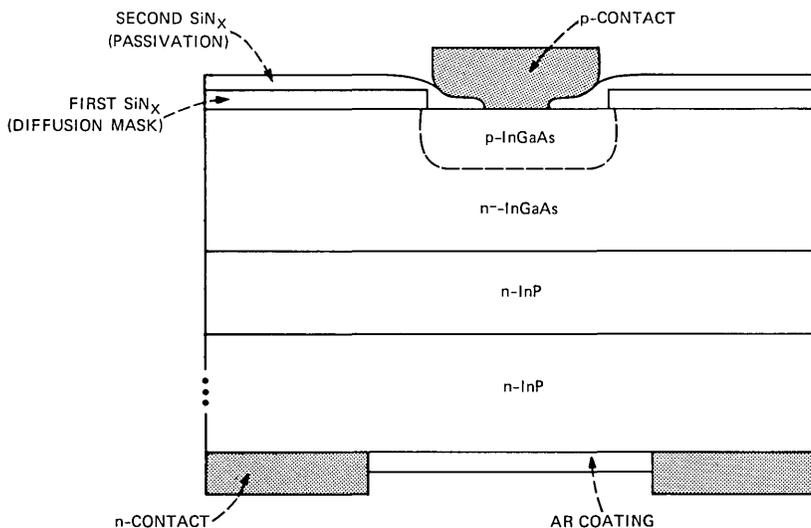
Fig. 1—Schematic illustration of planar InGaAs/InP photodiode. Typical characteristics are given in Table I.

Table I—Photodiode characteristics for SL applications

	Receiver ($V_R = 10V$)	Monitor ($V_R = 5V$)
Junction diameter (μm)	78	250
Maximum dark current (nA)	40	100
Maximum capacitance (pF)*	0.42	7.0
Minimum quantum efficiency	0.8	0.6
Maximum rise/fall time (ns)	0.5	0.5

* Chip only.

contacts.[9] The improvement is thought to be the result of reduced contamination and the absence of pores in the metallization. This design results in low dark current, high quantum efficiency, and, as shown below, excellent reliability.

III. CRITICAL DEVICE PARAMETERS AND POTENTIAL FAILURE MODES

In this section we briefly consider the device parameters which are critical to system operation, discuss possible failure modes, and suggest methods for probing sensitivity to the associated failure mechanisms. In general, increased dark current, increased capacitance, and reduced quantum efficiency deleteriously affect receiver sensitivity. Additionally, pulse rise and fall times must not increase to the point that they become comparable to the data time slot. For the monitor application, quantum efficiency and its spatial uniformity are the principal parameters. Electrical continuity is obviously critical to both applications.

Table II—Critical device parameters and potential failure
mechanisms

Parameter	Potential Failure Mechanisms	Accelerants
• Dark current	Bulk leakage (junction degrada- tion, local breakdown) Surface leakage	High-temperature bias Overvoltage High humidity Thermal cycling
• Capacitance	Diffusion, doping density changes	High temperature
• Quantum efficiency	"Darkening" of InP Degradation of AR coating	High temperature
• Rise/fall times	Formation of traps	High-temperature bias
• Electrical continuity	Open wire or chip bond Fractured chip	Thermal cycling Centrifuge Vibration Shock

Table II summarizes, for these parameters, potential failure mecha-
nisms and accelerants. A detailed discussion is given below.

3.1 Receiver photodiode

Failure of the photodiode in a digital lightwave receiver can impact
operation of a repeater section in two ways, either through catastrophic
failure or gradual degradation evidenced by a slowly increasing error
rate. Catastrophic failures can occur in these devices either from opens
due to die- or wire-bond failures, fractured chips, or from shorted
devices. Metallization-related failure modes, viz., electrolytic corrosion
and electromigration, are not expected because of the low current
density and low electric field in these devices.

Gradual degradation in the detector electro-optical properties can
lead to decreased receiver sensitivity. Clearly, a decrease in photodiode
quantum efficiency results in a proportional decrease in sensitivity.
Such degradation could, for example, arise from changes in the AR
coating applied to the device. Quantum efficiency could also decrease
if absorbing regions were to develop in the transparent InP substrate,
or if recombination centers were to develop in the InGaAs epitaxial
material.

Absorbing or recombination regions ("dark-spot" or "dark-line"
defects) observed in high-radiance LEDs and lasers usually result from
nonradiative recombination processes.[10] In detectors, photo-excited
carriers are rapidly separated by the field in the depletion layer, i.e.,
there is no carrier recombination. Moreover, the current densities are
several orders of magnitude lower than for LEDs and lasers. Conse-
quently, these phenomena are not expected to occur. In fact, after
1500 hours of accelerated aging at 200°C, the observed changes in
quantum efficiency of (AR-coated) photodiodes are less than ±0.1 dB.
Assuming an activation energy greater than 0.5 eV (see Section 4.1),
this is equivalent to $\gtrsim 2 \times 10^6$ hours of operation at TAT-8 conditions

($\leqslant 30^\circ$C). Similar findings have been reported for mesa InGaAs photodiodes.[3]

Other parameters contribute to receiver sensitivity in increasingly subtle but well-understood ways.[11] For example, the detector dark current contributes shot noise at the receiver input, degrading the signal-to-noise ratio. Consequently, the usual failure mode in photodetectors is taken as increased dark current[12] with End-Of-Life (EOL) determined by a prescribed degradation in receiver sensitivity (e.g., 1 dB). The increase in dark current can arise from either surface and/or bulk contributions. As mentioned above, poor or no passivation can lead to contamination of the junction, resulting in surface leakage. Furthermore, field-assisted migration of impurities from the surface into the diode's active region can result in increased generation-recombination current. Also, field-assisted drift of charge in the dielectric can result in inversion or accumulation on the semiconductor surface which, by increasing the depletion layer volume, would increase generation-recombination current.

The dark-current shot noise, however, is not necessarily the most important noise parameter. The dependence of the mean-squared shot-noise current on bandwidth or bit rate B (receiver bandwidth is approximately $0.5 \cdot$ bit rate) is linear, whereas other noise terms vary as higher powers of B. For example, for the SL receiver that uses a bipolar transistor as the first gain element, the capacitance of the detector multiplies transistor noise terms that vary as B^2 and B^3. Thus, at the SL bit rate of 296 Mb/s, dark current is not a major noise source until it reaches the order of 1 μA. As described in Section IV, we find that in accelerated aging tests, the dark current (I_d) can increase very rapidly, eventually resulting in a virtually shorted device. We have taken $I_d \geqslant 100$ nA as the operational definition of "end-of-life"; experimentally we find that this definition is essentially equivalent because of the catastrophic increase in I_d.

As suggested above, detector capacitance strongly influences SL receiver sensitivity and it conceivably could increase with time to reduce sensitivity. For example, thermally activated changes in the diffusion profile could alter capacitance. Since the capacitance is determined by the large width of the "intrinsic" region in a pin diode, plus relatively small depletion widths in the p and n material on either side, and since the operating temperature is only about one-third of the absolute temperature at which diffusion is performed, it is highly improbable that such changes will occur. On the other hand, charge accumulation near the surface could induce local changes in the depletion layer width. As in the case of quantum efficiency, no significant changes ($< \pm 0.02$ pF) have been observed in accelerated tests, consistent with results reported elsewhere.[3]

Finally, the intrinsic rise time of the detector can affect receiver sensitivity through Intersymbol Interference (ISI). The equalizer in the linear channel of a receiver is normally adjusted at the time of manufacture. If subsequent changes in the detector alter the pole location(s) of the receiver, the equalizer is no longer correctly compensating for these poles and intersymbol interference appears. Within the resolution of our rise-time measurement (±30 ps), no changes in this parameter have been observed after 1500 hours of accelerated aging at 200°C.

3.2 Monitor photodiodes

The discussion of the preceding section largely applies in the case of the laser monitors. Since the quantum efficiency directly scales the current level used as an analog of the laser output for control and stabilization purposes, quantum efficiency must remain stable at least to the extent that the laser output must remain stable. In present applications of these monitor detectors, capacitance and rise time play relatively unimportant roles. Dark current in the laser monitor is limited by its possible effect on the laser feedback-control circuit, being interpreted equally with true photocurrent as a "light level." The circuit-imposed limit on dark current is 5 μA, which is taken as the operational definition of end-of-life.

IV. ACCELERATED AGING TESTS

Long-term accelerated aging tests have been conducted on over 500 receiver and monitor photodiodes. These results have produced iterations in the device design for enhancing reliability and provide the basis for qualifying the device design for TAT-8. We first consider tests that utilize high temperature as a means for accelerating failure.

4.1 High temperature-bias aging

Photodiodes, assembled in hermetic packages, were operated at the normal bias condition at various elevated temperatures in the range 85 to 260°C. In situ measurements of I_d were made. Periodically, devices were cooled (while biased) to room temperature for measurements of I_d, reverse breakdown voltage V_{BR}, and, for some devices, rise and fall times, quantum efficiency, and capacitance. Figure 2 illustrates a typical I-V characteristic as a function of aging time. I_d (defined at the operating voltage) remains nearly constant, in general for an extended period, then may slowly increase, and finally increases rapidly to failure—defined as 100 nA for receiver photodiodes and 5 μA for monitor photodiodes. The slow degradation is not always present and its rate is not correlated to the onset of the rapid rise in
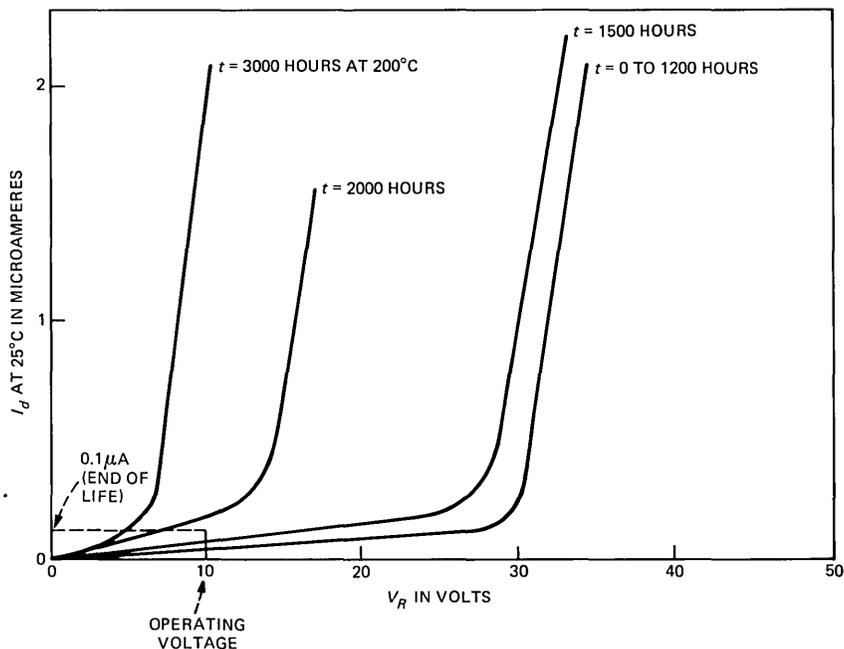
Fig. 2—Typical change in I-V characteristic for photodiodes as a function of aging time.

I_d. Therefore, the devices appear to have a "sudden-failure" characteristic with respect to I_d. A qualitatively similar behavior is observed for silicon photodiodes.[13] Measurements of V_{BR} do not provide an earlier precursor to failure than I_d, and the time to failure is not correlated to the initial value of V_{BR}. In situ measurements of I_d provide, at least for some diodes, an earlier indication of failure than room temperature measurements. The latter is not of significant advantage, however, since failures that occur in several thousand hours (under accelerated conditions) can be identified as impending failures only ~200 hours earlier. Consequently, room temperature measurements are principally used. As indicated previously, it has been found that there are no systematic changes in rise and fall times, quantum efficiency, and capacitance over long-term, high-temperature aging within our experimental error. Increased dark current is therefore the only observed chip-related failure mode.

Figure 3 illustrates several (dark-current) failure distribution plots for receiver photodiodes of early design that were aged at various temperatures. Two modes of failures are observed: early failures of a freak population, which in these devices is a high proportion (0.15) of the sample population; and the main population. At the lowest tem-
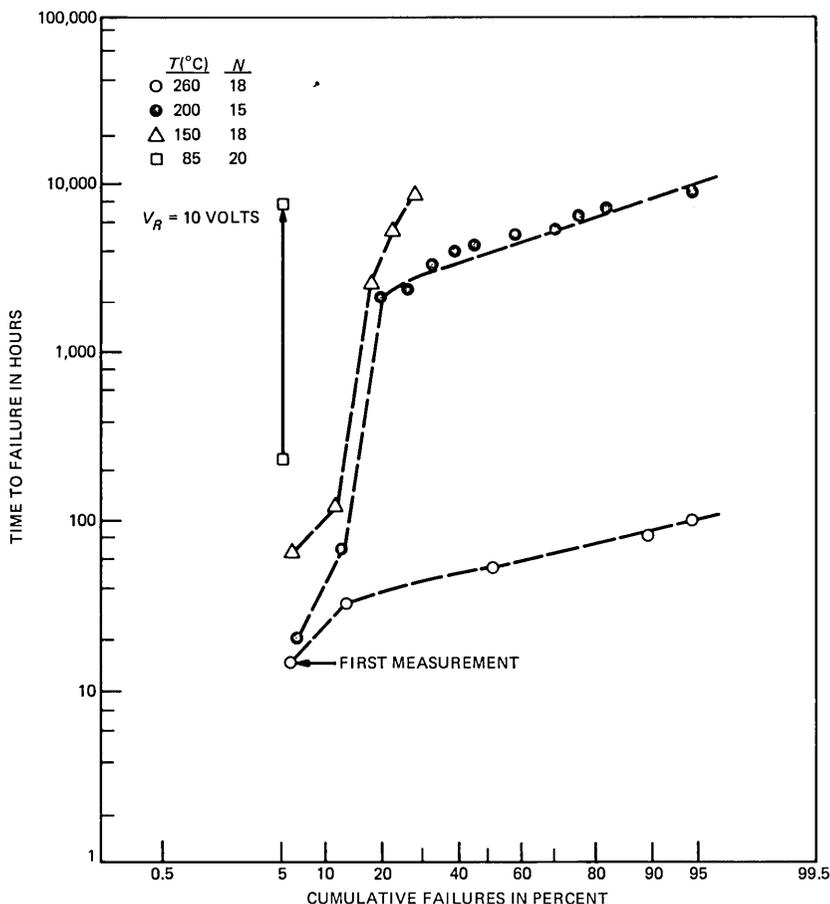
Fig. 3—Dark-current failure distributions for receiver photodiodes of early design after high-temperature aging. The vertical arrow indicates that no further failures have occurred.

peratures ($T < 150°$C), the main population is not clearly defined even after nearly 10^4 hours of aging. In fact, at 85°C the observed changes in I_d (neglecting the first early failure) are within experimental error. The data in Fig. 3 indicate that failures in both the freak and main populations are thermally activated, although not necessarily with the same activation energy. Assuming that the freak population can be removed by some appropriate means (see below), the Median Life (ML) and standard deviation, σ, for the main population can be estimated using conventional censoring techniques.[14]

Figure 4 shows the temperature dependence of ML for 116 receiver photodiodes; the corresponding values of σ are given in parentheses. For 150°C, a worst-case estimate was made assuming that the last
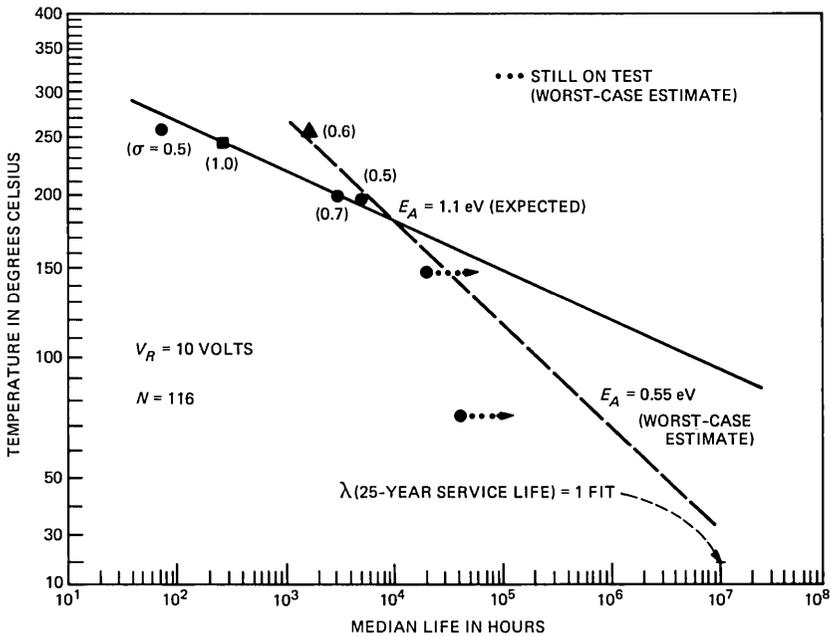
Fig. 4—Median life as a function of temperature for receiver photodiodes. Activation energy estimates vary from 1.1 eV (expected value) to 0.55 eV (worst case). The median life (for $\sigma = 1$) required for a maximum failure rate of 1 FIT over a 25-year service life is shown for reference.

observed failure in the transition region from freak to main population was the first failure of the main population. An average value of $\sigma = 0.8$ was used to extrapolate to ML. In a similar fashion, a worst-case estimate at 85°C indicates ML $\gg 2 \times 10^4$ hrs. Note that at 250 to 260°C there is a spread in estimated values of ML by a factor of ~140. The progressive increase in ML, in fact, reflects systematic improvements in device design and processing. The "probable" activation energy, shown as the dashed line, is 1.1 eV, which is reasonably consistent with the data; a "worst-case" activation energy, shown by the dotted line, is 0.55 eV. On the abscissa is a reference mark at the ML corresponding to a maximum failure rate of 1 FIT over the 25-year service life for TAT-8. The accelerated aging data collected in this phase of the receiver photodiode qualification program predict that even the worst-case estimates are consistent with a 1-FIT reliability objective at SL operating conditions, viz., 10°C undersea and 30°C continental shelf. Similar studies have begun for monitor photodiodes. As these latter devices use identical fabrication technology, operate at a lower bias, and have a higher dark current at end-of-life, it is likely that the 30-FIT reliability objective will be met.

As an adjunct to the above test program, failure mode analysis was

(a)

(b)
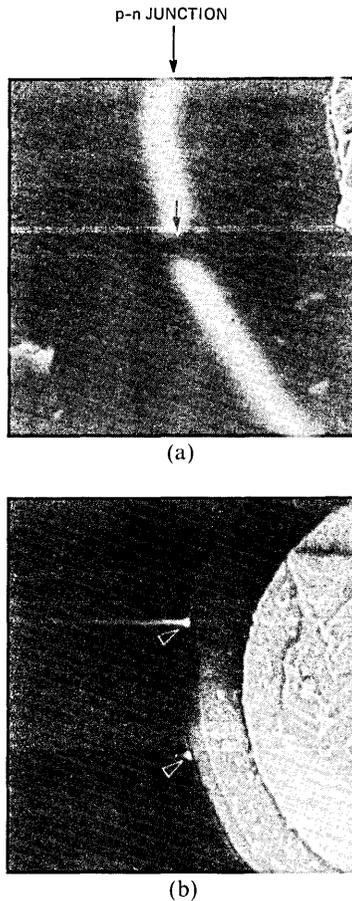
Fig. 5—Electron-beam-induced current signals superimposed on secondary electron images of failed photodiodes. A photodiode with $V_{BR} = 0$ exhibits a leakage path shorting the junction, designated by the arrow in Fig. 5a. A photodiode with a reduced V_{BR} of 8V exhibits localized breakdown, designated by arrows in Fig. 5b, when a reverse bias of 8V is applied.

performed using the Electron-Beam-Induced Current (EBIC) mode of a scanning electron microscope.[15,16] In particular, a correlation was sought between failures and anomalies in the region where the p-n junction intersected the SiN_x passivation layer. For devices for which $V_{BR} = 0$ after aging, a local leakage path shorting the p-n junction is evident, as in Fig. 5a. In other cases, local premature breakdown sites were evident in the EBIC signals when the applied reverse bias was equal to the (reduced) photodiode breakdown voltage, as in Fig. 5b. The cause of the large reduction in V_{BR} is not known but presumably involves the field- and temperature-assisted drift of some impurity species and/or defects to localized sites in the p-n junction. The

localized failure sites have been correlated[16] with microplasmas that are initially present.

4.2 Overvoltage aging at high temperature

To determine the extent to which operating voltage affects reliability, 100 monitor photodiodes were aged at 200°C with reverse bias V_R ranging from 0 to 30V. Figure 6 illustrates the pronounced effect of V_R on ML. It is noted that only I_d failures are observed and no changes in capacitance or quantum efficiency are detected prior to I_d failure. The mode of failure induced by overvoltage is therefore similar to that induced by thermal acceleration. The importance of reverse bias has been previously established in silicon phototransistors where ML was inversely related to V_R.[13] In a device with homogeneous junction characteristics, the leakage current under conditions of high temperature and overvoltage flows uniformly through the junction. Power dissipation is sufficiently low that devices are not impaired. The shortened ML with increased V_R is likely owing to enhanced field-assisted drift of an impurity species to localized sites, as mentioned above. However, if the device is flawed, so that a disproportionally
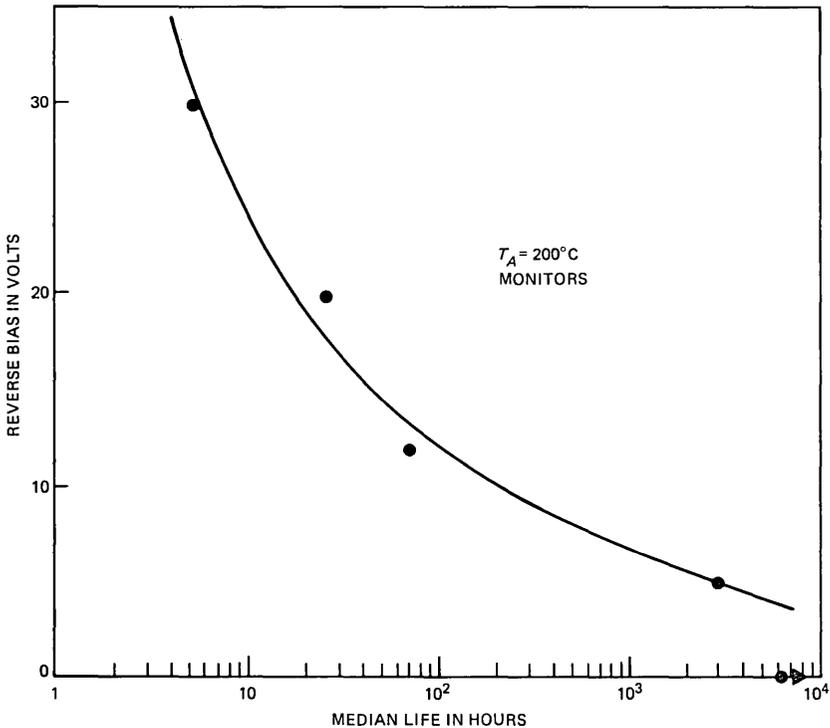
Fig. 6—Effect of reverse bias on the dark-current median life for monitors.

large fraction of the current flows through a small region of the junction, the temperature will rise locally; this will further increase current flow, leading to a runaway process which will ultimately inflict permanent, localized damage, perhaps because of melting or defect generation. At the highest overvoltage, local semiconductor damage was readily visible at the perimeter of the p-n junction.

The previous section showed that temperature at fixed bias accelerates photodiode degradation, while the data presented in Fig. 6 showed that reverse bias at fixed temperature accelerates degradation. Consequently, both parameters can be utilized to formulate an optimum method for purging weak devices. This subject will be treated more fully in Section V.

4.3 Humidity-bias testing

Humidity is known to provide surface leakage paths in reversed-biased diodes, which can ultimately lead to device failure.[13,17] For this reason our photodiodes are hermetically packaged. To assess the sensitivity of the photodiodes to humidity, over 100 unpackaged photodiodes were bias-aged at 85°C with relative humidity levels ranging from <3 to 85 percent.[9] Consistent with studies on other devices[17], ML of I_d failures was reciprocally related to the humidity level, and bias substantially enhanced the degradation rate. However, the mechanism of failure was surprising. Scanning electron microscopy of failed devices revealed that the n-type InGaAs, in the vicinity of the p-n junction, was severely etched. It was postulated that when moisture penetrated the dielectric and bridged the p-n junction, rapid electrochemical oxidation of the n-type semiconductor occurred which damaged the junction, effectively shorting the device. From the humidity data, the critical concentration of moisture at the p-n junction which phenomenologically resulted in device failure was calculated. From the observed time dependence of failure, the corresponding critical ambient moisture level was estimated to be ~3×10^5 ppm. Consequently, this failure mechanism is eliminated in the hermetic packages used in SL applications, as the entrapped moisture level is ≲10 ppm.

4.4 Temperature cycling

Initial temperature cycling tests of first-generation photodiodes, which used plated p-contacts, revealed an alarmingly high failure level resulting from wire bond opens. For example, 10 percent of the diodes developed open bonds at the p-contacts after 10 cycles from −55° to 125°C. Bond adherence was substantially improved with the introduction of evaporated contacts. Thus, in 406 receiver photodiodes that were similarly cycled, only two opens developed and no failures developed in 296 monitors. The low incidence of bond failure results from

the fact that the wires are not subject to mechanical constraint in the hermetic packages used here, but it might be the case in a molded or filled package. In the course of this study we found that the thermal cycling can induce significant increases in dark current in ~10 percent of the devices, providing a means of screening out flawed devices. The increase in dark current is presumed to arise from thermally induced stress at the InGaAs-SiN$_x$ interface.

V. PURGE: A MEANS OF ELIMINATING EARLY FAILURES AND IDENTIFYING WEAK DEVICES

It is common with various semiconductor devices to use a high-temperature "burn-in" at the maximum-rated bias to eliminate early failures associated with a freak population of devices in which the failure mechanism is thermally activated. For the burn-in to be effective, the ML of the freak population must be orders of magnitude smaller than the main population, so that the burn-in can be long enough to eliminate (fail) the freak devices without adversely affecting the life of the main population. Figure 3, however, shows that at 200°C, a burn-in of 500 to 1000 hours is needed to eliminate freaks. Such a burn-in is prohibitively long and will significantly shorten the useful life of the remaining devices (ML = 4000 hours). Because of the sudden-failure characteristic, initial degradation rate cannot be used to identify early failures. Finally, a conventional burn-in does not identify flawed devices which may fail by very low activation energy mechanisms.

In Section 3.2, we showed that overvoltage provided a nonthermal means of accelerating photodiode degradation. We have determined that a *combined* stress of high temperature and overvoltage provides an effective purge for failing freak devices and for identifying flawed devices which would likely fail early in the TAT-8 life cycle.[18] Ideally, the purge will result in a population of robust devices that will meet or (more likely) exceed the reliability expected of the main population. To evaluate a trial purge, photodiodes are subjected to a relatively short but strong purge. Weak devices will tend to fail as a result of the purge. For the survivors, the change in I_d provides a useful measure of device stability, which is then used to eliminate additional weak devices. The efficiency of the purge is judged by life testing (high temperature but normal bias) *all* of the survivors and comparing the early failures in the life tests with those identified in the postpurge measurements.

Figure 7 is a histogram of the change in I_d for 33 receiver photodiodes following a 10-hour, 200°C purge at a 2X overvoltage. The purge was sufficient to result in four failures. In general, reverse-bias aging at elevated temperatures results initially in a decrease in I_d, perhaps as
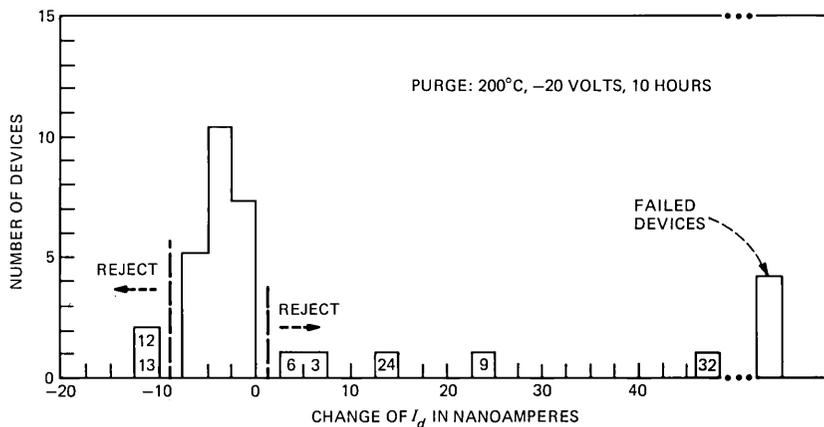
Fig. 7—Histogram of change in I_d for receiver photodiodes following a 200°C purge at 2X overvoltage.
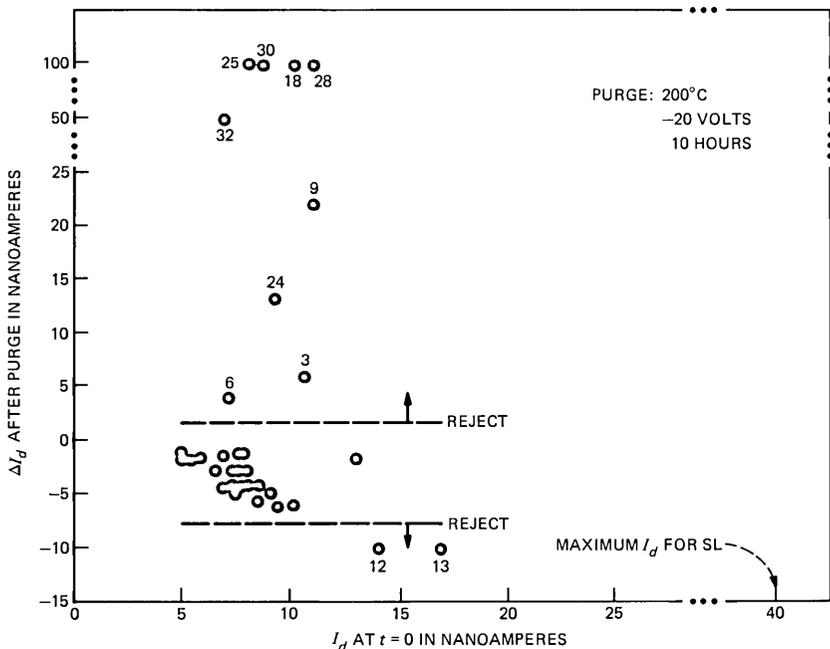
Fig. 8—Change in I_d for receiver photodiodes following a 200°C purge at 2X overvoltage as a function of initial I_d.

a result of stress relief or "stabilization." However, in Fig. 7 we see that several devices show significant increases in I_d, while several others show a higher than nominal reduction in I_d. Large changes in I_d, both positive and negative, are considered to be characteristic of

unstable devices, and consequently those devices should be eliminated. In the histogram these devices are identifed with the device number. Figure 8 shows the change in I_d resulting from the purge as a function of initial I_d. Weak devices, identified in the purge, are indicated with the device number. While there is some tendency for high I_d devices to exhibit the largest change in I_d, there are numerous exceptions. On the other hand, most devices for which I_d is initially greater than 10 nA are culled out in this purge. This result, however, is not general. Consequently, purge yield is not strongly correlated to the initial value of I_d. Similarly, purge yield is not correlated to initial V_{BR}, although devices with very low V_{BR} (<10V) are likely to fail.

Figure 9 shows I_d of the purged population (excluding devices that failed) after 1900 hours of life testing at 200°C and $V_R = 10$V. The numbered devices are those identified as weak devices in the purge. Note that two of these identified devices failed in the life test, at 310 hours (device 24) and 600 hours (device 32). Also, identified devices 9 and 13 are nearest to failure, i.e., have the highest dark current. Consequently, the purge at 2X overvoltage appears to fail successfully the weakest devices and to identify for removal devices that subsequently fail early. The remaining devices are the most robust portion
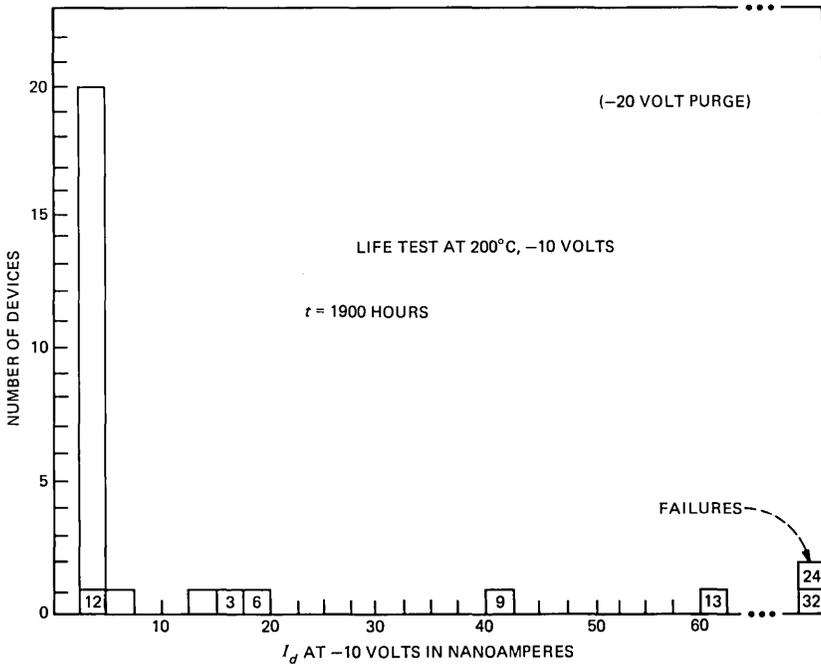
Fig. 9—Histogram of I_d values for the purged devices in Fig. 7 following a 1900-hour life test.
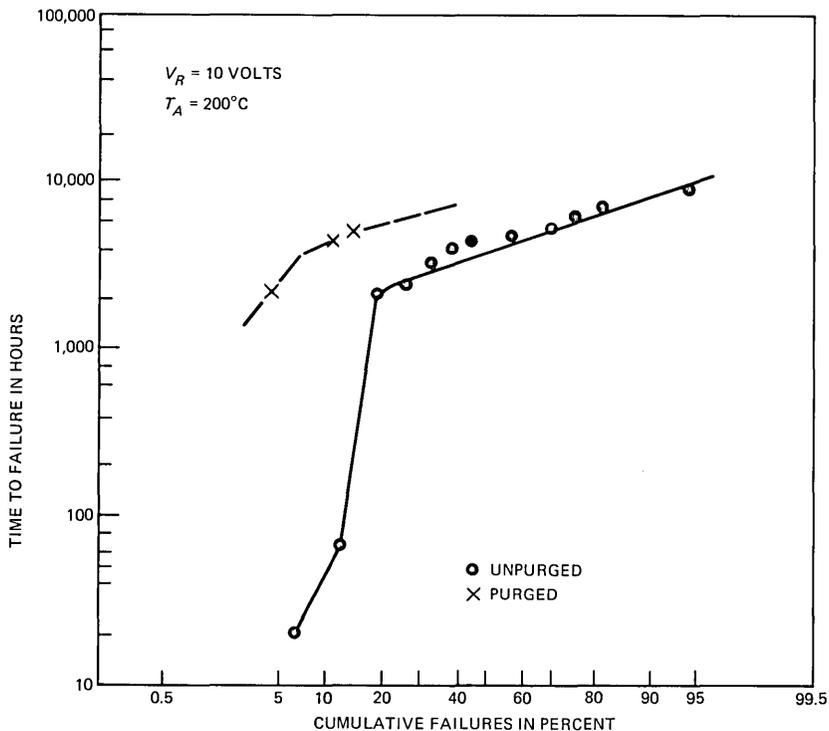
Fig. 10—Dark-current failure distribution for purged and unpurged receiver photo-diodes.

of the main population. This is confirmed by Fig. 10, which gives the failure distribution after extended ($7 \cdot 10^3$ hours) life testing of the robust devices (not numbered in Fig. 9) that were identified by the purge. Shown for comparison are unpurged devices. The dramatic reduction in early failures for the purged devices is evident.

One of the drawbacks of the aforementioned trial purge is the lack of wide separation between some of the weak devices and the robust main population. For example, in Fig. 7, devices 12 and 13, and 3 and 6, appear to be part of the tails of the main population. One expects that a purge at higher overvoltage may induce greater changes in I_d in weak devices, thus providing the desired spread. Tests at 3X overvoltage look promising. Figure 11 is a histogram of the change in I_d for receiver photodiodes following purge. Figure 12 is the corresponding histogram of I_d following a subsequent 1800-hour life test. Although the sample size is small, it appears that the stronger purge results in two distinct populations, viz., failed devices and robust devices. Note, for example, comparing Figs. 9 and 12, that after 1800-hour life testing, the purge survivors show less spread in I_d for the stronger purge. In
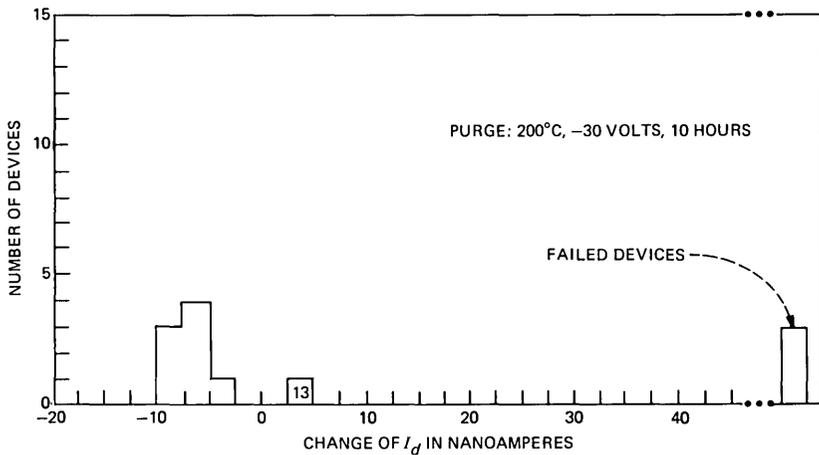
Fig. 11—Histogram of change in I_d for receiver photodiodes following a 200°C purge at 3X overvoltage.
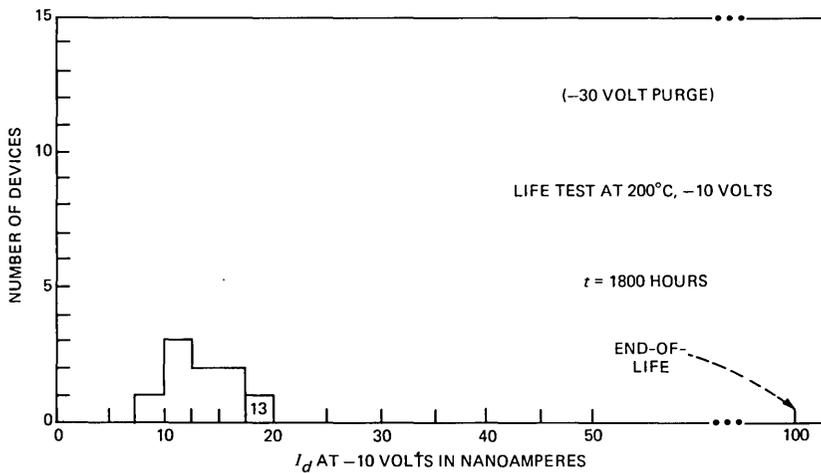
Fig. 12—Histogram of I_d values for the purged devices in Fig. 10 following an 1800-hour life test.

fact, Fig. 13 shows that the change in I_d after the 1800-hour life test is uniformly small for the strong purge, suggesting that the survivors are robust devices.

Additional experiments have been performed to establish optimum purge conditions. Figure 14 summarizes the effect of increased overvoltage, at constant temperature, on receiver photodiode failures. As the overvoltage is increased, the fraction of devices that are failed by
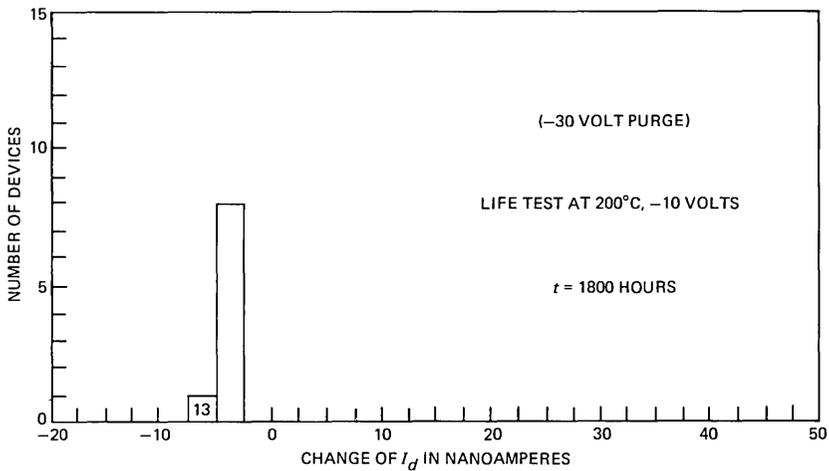
Fig. 13—Histogram of change in I_d for the purged devices in Fig. 10 following an 1800-hour life test.

the purge increases and the fraction of devices that fail a subsequent life test (initially) decreases. However, for the highest overvoltage, the failure level increases rapidly, indicating that "good" devices are being damaged. This is borne out by the corresponding increase in life test failures. Note also the increase in the total (purge plus life test) failure level. The results in Fig. 14 suggest that a 200°C purge at 30V (3X overvoltage) is near optimum for our receiver photodiodes. Similar studies are in progress for monitor photodiodes.

VI. CONCLUSIONS

Testing of planar InGaAs photodiodes under a variety of accelerated conditions reveals that increased dark current is the only detected failure mode. High-temperature aging results to date indicate that, for receiver photodiodes, a device design exists (main population of failures) that is predicted to be consistent with a 1-FIT reliability objective at TAT-8 operating conditions. High reliability (<30 FITs) is expected for monitor photodiodes of similar design. Moreover, a purge using a combination of accelerants (temperature and overvoltage) has been developed, which has considerable promise of eliminating flawed or weak devices that would otherwise fail early in the TAT-8 life cycle. Extended life tests verify that the purge survivors are the most robust devices. Consequently, purge testing coupled with appropriate inspection and thermomechanical screening is expected to provide a means for assuring that the selected devices have a high probability of meeting TAT-8 reliability objectives.
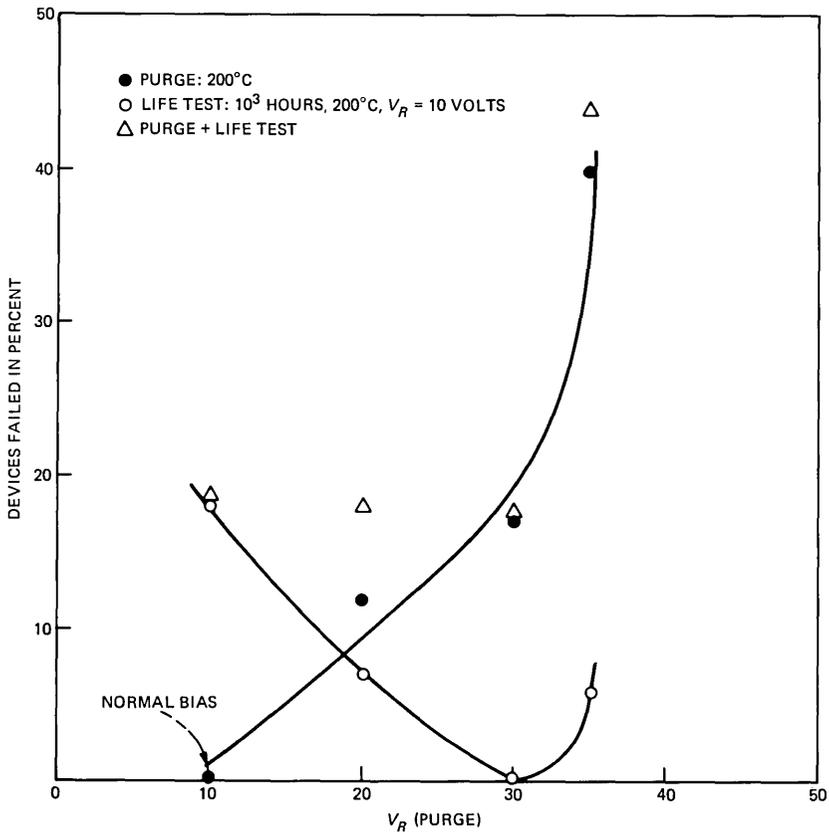
Fig. 14—Effect of reverse bias on resulting receiver photodiode failures for the purge and subsequent life test. Purge time is ten hours except at $V_R = 35V$, for which a time of one hour was used.

VII. ACKNOWLEDGMENTS

It is a pleasure to acknowledge the technical assistance of C. F. King in acquiring the aging data, the cooperation of O. G. Lorimor and W. S. Trimmer in supplying the devices used in these studies, and the encouragement and support of A. A. Bergh. The manuscript has benefited from many useful suggestions of R. L. Hartman and E. I. Gordon.

REFERENCES

1. F. S. Chen, unpublished work.
2. G. H. Olsen and M. Ettenberg, "High Temperature Reliability of Vapor-Grown InGaAs Photodiodes," topical meeting on Optical Fiber Communication, Phoenix, Ariz., April 13–15, 1982.
3. D. G. Jenkins and A. W. Mabbit, "The Reliability of GaInAs Photodiodes and GaAs FETs for Use in PIN-FET Fiber Optic Receivers," IEEE Specialist Conference on Light-Emitting Diodes and Photodetectors, Ottawa-Hull, September 15–16, 1982.

4. J. M. Nemchik and M. J. Buckler, "Design of the SLC^{TM} 96 System Loop Lightwave Feature," IEEE Int. Conf. Commun., Boston, Mass., June 1983, Conf. Rec., *I,* pp. 91–4.
5. T. P. Pearsall et al., "A GaInAs/InP Heterophotodiode With Reduced Dark Current," J. Quant. Elec., *QE-17* (February 1981), pp. 255–9.
6. T. P. Lee et al., "Small Area InGaAs/InP p-i-n Photodiodes," Elec. Lett., *16* (February 1980), pp. 155–6.
7. G. H. Olsen, "Low-Leakage, High Efficiency, Reliable VPE InGaAs 1.0–1.7 μm Photodiodes," IEEE Elec. Device Lett., *EDL-2* (September 1981), pp. 217–9.
8. O. G. Lorimor et al., unpublished work.
9. L. S. Lichtmann, P. A. Kohl, and R. H. Burton, "Degradation of Silicon Nitride-Protected InGaAs p-i-n Photodiodes by an Electrochemical Etching Reaction," IEEE Specialist Conference on Light-Emitting Diodes and Photodetectors, Ottawa-Hull, Canada, September 1982.
10. R. H. Saul, "Recent Advances in the Performance and Reliability of InGaAsP LEDs for Lightwave Communication Systems," IEEE Trans. Elec. Devices, *ED-30* (April 1983), pp. 285–95.
11. R. G. Smith and S. D. Personick, *Topics in Applied Physics, 39,* New York, N.Y.: Springer-Verlag, 1980, pp. 89–160.
12. D. P. Schinke, R. G. Smith, and A. R. Hartman, *Topics in Applied Physics, 39,* New York, N.Y.: Springer-Verlag, 1980, pp. 63–87.
13. R. H. Saul et al., unpublished work.
14. D. S. Peck, "The Analysis of Data From Accelerated Stress Tests," Proc. Reliability Phys. Symp., Las Vegas, Nev., March 31, 1971, pp. 69–78.
15. F. S. Chen, P. S. D. Lin, and F. Ermanis, unpublished work.
16. A. K. Chin, F. S. Chen, and F. Ermanis, "Failure Mode Analysis of Planar Zn-Diffused $In_{0.53}Ga_{0.47}As$ PIN Photodiodes," J. Appl. Phys., *55* (March 1984), pp. 1596–606.
17. D. S. Peck and C. H. Zierdt, Jr., "Temperature-Humidity Acceleration of Metal-Electrolysis Failure in Semiconductor Devices," Proc. Reliability Phys. Symp., Las Vegas, Nev., April 3, 1973, pp. 146–52.
18. R. H. Saul and F. S. Chen, "Reliability Assurance for Devices With a Sudden-Failure Characteristic," IEEE Elec. Device Lett., *EDL-4* (December 1983), pp. 467–8.

AUTHORS

Fong S. Chen, B.S. (Electrical Engineering), 1951, National Taiwan University; M.S. (Electrical Engineering), 1955, Purdue University; Ph.D. (Electrical Engineering), 1959, Ohio State University; AT&T Bell Laboratories, 1959—. Mr. Chen has worked on the development of ferrite devices, masers, optical modulators, lightwave transmitters, and on the reliability of photodiodes. He is currently engaged in the development of GaAs integrated circuits. Member, IEEE.

Robert H. Saul, Ph.D (Metallurgy and Materials Science), 1967, Carnegie-Mellon University; AT&T Bell Laboratories, 1967—. Initially, Mr. Saul was engaged in characterization and development of epitaxial growth techniques for opto-electronic materials and devices. In 1972 he became Supervisor of a group responsible for developing visible light-emitting diodes. That work pioneered the use of multislice epitaxial techniques for achieving state of the art performance in a manufacturable growth system. Since 1975 Mr. Saul has supervised a group responsible for developing a variety of infrared light-emitting diodes that are used in optical-isolators and fiber optic systems. His current interests include development of long-wavelength sources for lightwave transmission systems and reliability of detectors. Mr. Saul holds eight patents in the area of materials growth and opto-electronic devices. Senior member, IEEE; member, American Physical Society, Sigma Xi, and Tau Beta Pi. Chairman, 1982 IEEE Specialist Conference on Light-Emitting Diodes and Photodetectors.

Paul W. Shumate, Jr., B.S. (Physics), 1963, College of William and Mary; Ph.D. (Physics), 1968, University of Virginia; Bell Laboratories, 1969–1983. Present affiliation Bell Communications Research, Inc. At Bell Laboratories, Mr. Shumate's work included research on the physical properties of magnetic bubble materials and on memory devices. In 1975 he became Supervisor in the Lightwave Devices and Subsystems Department, where he was subsequently responsible for FT3 laser transmitter development, optical data links for *5ESS*™ switching equipment, laser- and LED-based video transmitters, and p-i-n/FET receivers. Other work included detector packaging and reliability studies for SL, and laser packaging for FT4E. Since transferring to Bell Communications Research in October 1983, where he is District Manager of Lightwave Device Technology, he is responsible for research on very-high-bit-rate digital and heterodyne laser transmitters and receivers. Member, American Physical Society, Phi Beta Kappa, Sigma Xi; Senior Member, IEEE.

CONTENTS, APRIL 1985