

Statistical Methods for Large Flight Lots and Ultra-high Reliability Applications

R. Ladbury, *Member, IEEE*, and J. L. Gorelick

Abstract— We present statistical techniques for evaluating random and systematic errors for use in flight performance predictions for large flight lots and ultra-high reliability applications.

I. INTRODUCTION

Sampling strategies for radiation testing often represent a compromise between generality and economy.[1] The most general strategies (those using binomial statistics) assume little about parts' radiation response distribution (RRD). However such strategies require large samples to achieve high confidence of high success probability (e.g. 22 parts with no failures to have 90% confidence that at least 90% of parts would pass). The main reason for these large samples is that the schemes must work for pathological thick-tailed or multimodal failure RRDs as well as for those that are well behaved. (See Fig 1.) Moreover, because binomial sampling makes no assumptions about distribution form, even if we know with 90% confidence that, for example, 90% of parts pass at a dose D, we cannot say how many would pass at any other dose level.

More economical strategies argue based on technical and heuristic grounds that the RRD within a wafer lot is consistent from part to part,[2] and assume the RRD will approximate a particular form (usually Normal or Lognormal). This allows establishment of higher success probability and confidence with smaller test samples. Moreover, for small flight lots (<10 parts) and not-too-stringent reliability requirements, slight deviations of the actual RRD from the assumed form will not seriously alter RHA conclusions. However, if these conditions are violated, uncertainties in the RRD tails can dominate risk, and small sample do a poor job of constraining RRD tails or identifying pathological behavior therein.

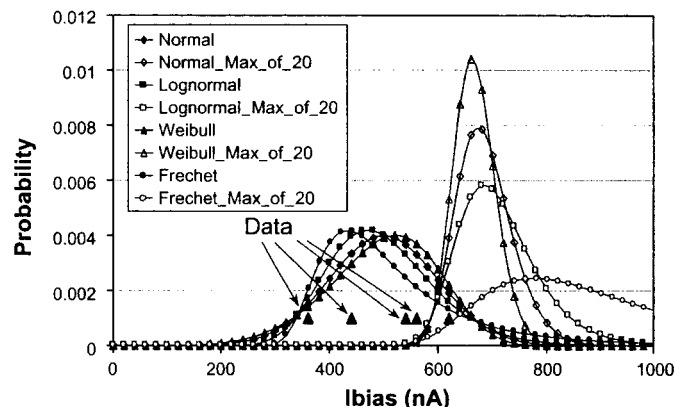


Fig. 1. Although Normal, Lognormal, Weibull and Frechet distributions yield good fits to the data shown, they behave very differently in their tails. This becomes clear in looking at the worst performer in flight lots as small as 20 parts.

Implications of distribution pathologies have been treated in discussions of “maverick” devices[3] and of bimodality in the ELDRS response of the National Semiconductor LM111 voltage comparator.[4],[5] Reference 3 considers the implications of occasional outliers seen in large-lot tests of 108A op amp. Specifically, about 1% of parts show abnormally large changes in offset voltage. The authors conclude that such maverick devices would likely not be detected by small-sample tests and this could have significant hardness assurance implications for some applications. Reference 4 considers the hardness assurance implications of bimodal radiation response in LM111 voltage comparators. The authors concluded that multimodality within a single wafer lot precludes sampling strategies that assume a particular distribution for any mode, since the existence of further modes at lower probability cannot be ruled out. Such parts require binomial (or, in the parlance used in reference 4, “distribution-free”) sampling. The authors suggested the bimodality in the LM111 resulted from subtle differences in post-processing that shifted the balance between competing mechanisms. The authors of reference 5 found evidence for such a competition occurring in the nitride passivation of the LM111s and concluded that the balance was shifted by differences in the pre-irradiation elevated thermal stress (PETS) to which the parts were exposed after wafer fabrication. The elucidation of this mechanism meant that bimodality in the LM111 could be controlled. This reduced the urgency of developing RHA methodologies capable of dealing with pathological RRDs. However, the issue of pathological RRDs has not gone away,

Manuscript received October 9, 2001. (Write the date on which you submitted your paper for review.) This work was supported in part the NASA James Webb Space Telescope (JWST) Program and the NEPP Electronic Radiation Characterization (ERC) Project.

R. Ladbury is with the Muniz Engineering, Inc, Greenbelt, MD 20771, USA (phone: 301-286-1030; fax: 301-286-4699; e-mail: Raymond.L.Ladbury.1@gsc.nasa.gov).

J. L. Gorelick is with Boeing Space Systems, Los Angeles, CA USA

as we show below, and particularly for commercial parts, remediation of such response may not be feasible. RHA methods for dealing with such parts are still needed.

In addition to such pathologies, there is also the issue of systematic errors that may be introduced when the actual RRD—though well behaved—varies from the form assumed in the analysis. As Fig. 1 shows, even well behaved RRDs yield systematic errors for large flight lots under such circumstances. Similar errors occur for ultra-high-reliability applications, since here, too, uncertainties in the behavior of RRD tails can dominate risk to the application. Usually, the sample sizes used in radiation characterization and RLAT are too small to constrain RRD tails. Here we discuss use of representative archival data to constrain distribution pathologies and provide sufficient statistics that bounding behavior for the part can be inferred. We also investigate the influence of assumed distribution form by fitting the data to several well behaved distribution forms with different symmetries—thereby estimating the distribution dependence of the analysis.

II. DATA SOURCE

The data used in this study were mined from the BSIS radiation database and were compiled in the course of normal lot testing at the Raytheon Component Evaluation Center using their gammacell 220 Co-60 irradiator. Most parts were procured to internal Boeing Specifications similar to Standard Military Drawings (SMDs). Tests were conducted per Mil-Std 883 Method 1019 at dose rates from 50-300 rads(Si)/s. Parametric and functional shifts were measured after each dose step. Because the test conditions and methodologies were consistent across the entire dataset, we were able to combine data for parts across many different wafer lots. Individual RLAT samples ranged from 4-20 parts.

III. METHODOLOGY

For each part, we calculated statistics at each dose step for a representative parameter for each lot and for the ensemble of all lots combined. We used the ensemble's greater statistics to characterize distribution pathologies—or if no pathologies were evident, we used binomial statistics to bound the proportion of parts that exhibited such pathologies. After Namenson,[6] we varied our analysis method depending on whether variability from lot to lot greatly exceeded that typically seen within a wafer lot, or whether inter-lot and intra-lot variation were roughly commensurate. For the commensurate case, we used the ensemble to infer flight-lot behavior. When intra-lot distributions were tight, but lot-to-lot variability were large, we combined data from the flight-lot RLAT sample and the ensemble, using the ensemble to determine required RLAT sample sizes and potential flight-lot variability and relying on flight-lot RLAT results to estimate mean flight-lot performance.

To gauge dependence of analysis conclusions on the assumed distribution form, we fit data to three different forms. We chose the Normal, Weibull and Lognormal distributions

for their different symmetries, their familiarity and the fact that they can be motivated on physical (Weibull or Lognormal) or mathematical (Normal) grounds:

We chose the Normal distribution

$$f(x, \mu, \sigma) = (2\pi\sigma)^{-1/2} \exp(-(x - \mu)^2 / \sigma^2) \quad (1)$$

because it is symmetric about its mean, and the sample-mean behavior for small samples is known to follow Student's *t* distribution. While it is unphysical for some problems because it is defined from $-\infty$ to $+\infty$, analysis conclusions will not be affected as long as $\mu \gg \sigma$.

We chose the Weibull distribution

$$f(x, w, s) = (s/w^s) x^{s-1} \exp(-(x/w)^s) \quad (2)$$

since for shape parameter, $s > 3.68$ (corresponding to $\sigma/\mu < 0.3$), it is skewed left. For $s < 3.68$, the effects of distribution breadth will generally dwarf those of negative skew in any case. While no analogue to the Student's *t* distribution exists for the Weibull, we can numerically estimate the distribution's small sample behavior as a function of s .

We chose the Lognormal distribution

$$f(x, \mu, \sigma) = (2\pi\sigma^2 x^2)^{-1/2} \exp(-(\ln(x) - \mu)^2 / \sigma^2) \quad (3)$$

because it is skewed slightly to the right. It can be physically motivated when damage is proportional to damage already sustained. As with the Weibull, it is necessary to calculate numerically the small sample behavior distribution-parameter estimators as a function of σ .

We fit data simultaneously to these three distributions using maximum likelihood (ML) techniques, since ML fits yield confidence interval as well as best-fit parameter estimates. By taking the fit parameters that yield worst-case results and fall on the bounds of these intervals, we can bound degradation at the given confidence level.

IV. DISTRIBUTION PATHOLOGIES

Although serious pathologies in RRDs are relatively rare, we found examples of both thick-tailed and bimodal distributions. These pathologies pose problems for conventional small-sample RHA methodologies because the RRD does not approach zero in a predictable way even for very large degradation levels. This effectively invalidates strategies such as increased margin or derating, since no matter how much margin one incorporates, one cannot be confident that degradation will not exceed that level.

A. OP484

The Analog Devices OP484 quad op amp exhibits a high and a low mode of radiation-induced increased bias current, *I*_{bias}, with at least some lots spanning both modes (see Fig. 2). While such bimodality can pose significant difficulties for RHA methodologies, the Standard Military Drawing (5962-???) version of this part allows *I*_{bias} to degrade up to 3000 nA after 100 krad(Si)—much higher than the levels seen here. A study of the saturation trends in the data and of the correlations between prerad and post-rad behavior suggests

that like the LM111, the bimodality here may derive from competing mechanisms and that the same mechanism responsible for the introduction of the high degradation mode may also affect the prerad leakage current.

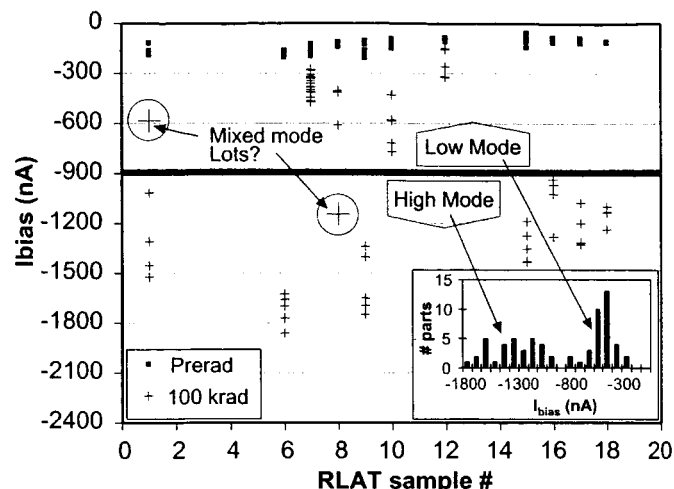


Fig. 2. OP484 Ibias vs. Lot and histogram after 100 krad(Si).

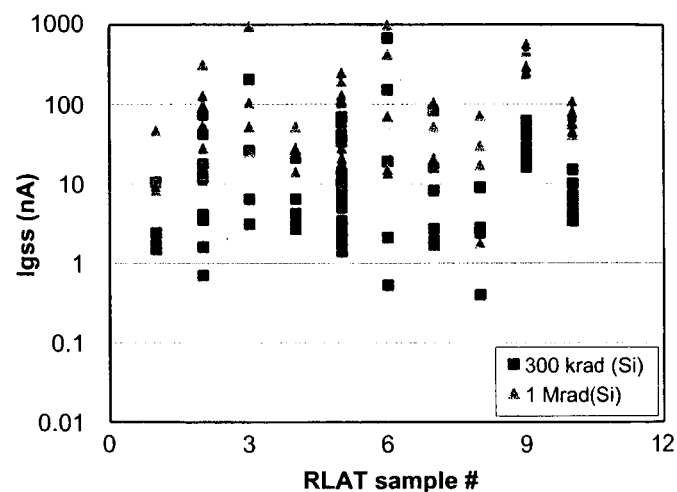


Fig. 3. Histogram of Ibias for RH1014 op amps after 200 krad(Si) along with Norm, Lognormal and Weibull fits to the data.

B. 2N5019

Gate-to-source leakage current I_{gss} after 0.3 and 1 Mrad(Si) in 2N5019 Junction Field Effect Transistors (JFET) varies so broadly that a log scale is needed to plot it. Several methods show the data may be fit by a thick-tailed (Frechet-type) extreme-value or very broad lognormal distribution (see Fig. 4). Predicting on-orbit performance for such a part is difficult, since a finite probability exists even for very high I_{gss} .

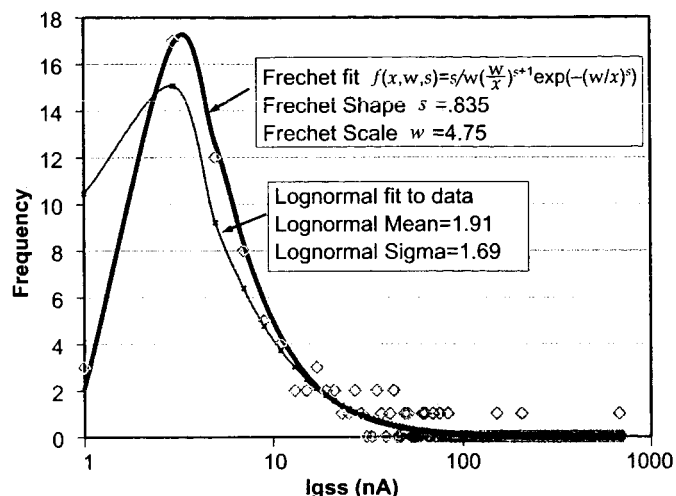


Fig. 4. I_{gss} for 2N5019 FETs fits either a thick-tailed extreme-value (Frechet-type) distribution or a very broad Lognormal.

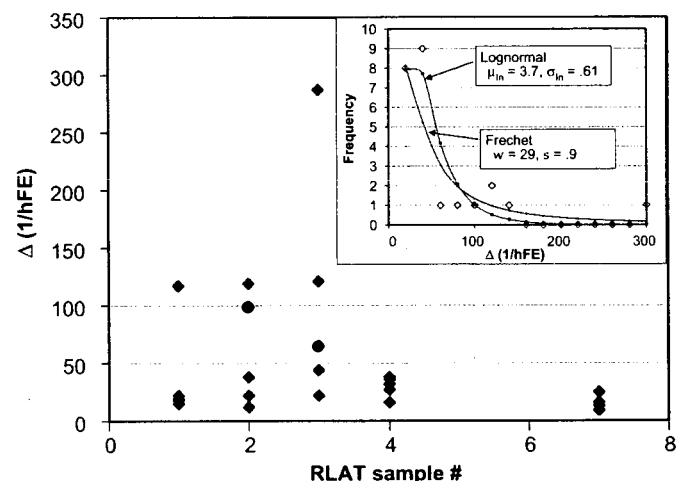


Fig. 5. Changes in gain of 2N2658 BJTs after 300 krad(Si) span more 32x. The inset shows the histogram of gain change, along with Lognormal and Frechet fits to the data.

C. 2N2658

The radiation-induced gain change in the 2N2658 bipolar junction transistor (BJT) is also thick-tailed, following either a broad lognormal or Frechet distribution. These trends suggest it is unwise to dismiss "outliers" that crop up in RLAT. With the added statistics of the ensemble, such "outliers" may resolve into pathologies rather than remaining isolated occurrences.

The difficulty of predicting degradation makes it important to understand how susceptible the application may be to the observed degradation.

V. WELL BEHAVED DATA

A. RH1014 Op amp

To illustrate use of well behaved archival data we use 38 RLAT samples (158 parts total) of the Linear Technologies RH1014 quad op amp tested at 60, 100 and 200 krad(Si). The fact that the 158-part dataset yielded no sign of RRD pathology establishes at the 90% CL (using binomial statistics) that any such pathologies comprise less than 1.5% of the parts. Fig. 6 shows I_{bias} after 200 krad(Si), along with Normal, Lognormal and Weibull fits to the data.

ML fits to the 200 krad data yield confidence intervals (the shaded ellipses in Fig. 7) as well as best-fit values (black). The fit parameters that yield worst-case distributions and are consistent with a given CL establish the bounding distribution for that CL.

Using the distributions with WC fit parameters of confidence level CL and for desired success probability, P_s , we define the design-to I_{bias} as the value where there is less than a $1 - P_s$ probability of exceeding that value. As shown in Table I, the contributions of both random and systematic errors become more significant as CL and P_s increase. Repeating the process for the 60 and 100 krad(Si) steps yields Fig. 8 for 99.9999% success probability at the 95% CL.

TABLE I: DESIGN-TO-IBIAS AFTER 200 KRAD (Si)

P(success)>	Fit	Normal	Lognormal	Weibull
99%	Best	76.8 nA	80.6 nA	73.7 nA
99%	95%WC	80.4 nA	86.4 nA	77.2 nA
99.9999%	Best	96.9 nA	116 nA	84.4 nA
99.9999%	95%WC	103 nA	130.5 nA	90.1 nA

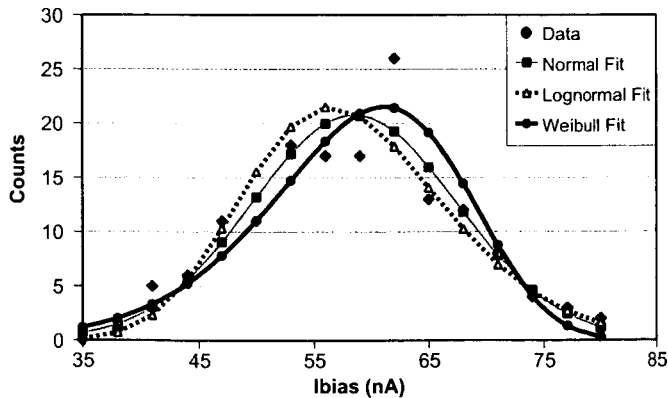


Fig. 6. I_{bias} after 200 krad(Si) and best fits to the data.

μ	7.3	7.5	7.7	7.9	8.1	8.3	8.5	8.7	8.9	9.1	9.3	9.5	9.7
55.6	-7.65	-6.30	-5.26	-4.51	-3.99	-3.40	-3.59	-3.66	-3.89	-4.25	-4.73	-5.32	-6.01
55.8	-6.65	-5.36	-4.37	-3.65	-3.18	-2.42	-2.86	-2.96	-3.21	-3.60	-4.11	-4.73	-5.45
56	-5.77	-4.52	-3.58	-2.90	-2.46	-2.14	-2.20	-2.34	-2.62	-3.03	-3.57	-4.21	-4.95
56.2	-5.01	-3.80	-2.89	-2.25	-1.84	-1.45	-1.64	-1.80	-2.10	-2.54	-3.10	-3.76	-4.51
56.4	-4.36	-3.18	-2.31	-1.69	-1.32	-1.15	-1.16	-1.34	-1.67	-2.12	-2.70	-3.37	-4.14
56.6	-3.83	-2.68	-1.83	-1.24	-0.88	-0.74	-0.77	-0.97	-1.31	-1.78	-2.37	-3.06	-3.84
56.8	-3.42	-2.29	-1.46	-0.89	-0.55	-0.42	-0.46	-0.67	-1.03	-1.51	-2.11	-2.81	-3.61
57	-3.13	-2.01	-1.19	-0.64	-0.31	-0.19	-0.25	-0.47	-0.83	-1.32	-1.93	-2.64	-3.44
57.2	-2.95	-1.84	-1.03	-0.48	-0.16	-0.05	-0.11	-0.34	-0.71	-1.21	-1.82	-2.53	-3.34
57.4	-2.86	-1.76	-0.98	-0.43	-0.14	-0.04	-0.09	-0.32	-0.67	-1.17	-1.78	-2.50	-3.30
57.6	-2.94	-1.84	-1.03	-0.48	-0.16	-0.04	-0.09	-0.32	-0.67	-1.17	-1.78	-2.53	-3.33
57.8	-3.11	-2.00	-1.18	-0.63	-0.30	-0.18	-0.25	-0.47	-0.83	-1.32	-1.92	-2.63	-3.43
58	-3.40	-2.28	-1.44	-0.87	-0.53	-0.40	-0.46	-0.67	-1.02	-1.50	-2.10	-2.80	-3.60
58.2	-3.81	-2.66	-1.81	-1.22	-0.87	-0.72	-0.75	-0.95	-1.29	-1.77	-2.35	-3.04	-3.83
58.4	-4.33	-3.16	-2.28	-1.67	-1.29	-1.12	-1.14	-1.32	-1.65	-2.10	-2.68	-3.36	-4.13
58.6	-4.97	-3.76	-2.86	-2.22	-1.81	-1.62	-1.61	-1.77	-2.08	-2.52	-3.07	-3.74	-4.49
58.8	-5.73	-4.48	-3.54	-2.86	-2.43	-2.21	-2.17	-2.31	-2.59	-3.01	-3.54	-4.19	-4.92
59	-6.61	-5.31	-4.33	-3.61	-3.14	-2.89	-2.82	-2.93	-3.18	-3.57	-4.09	-4.70	-5.42
59.2	-7.60	-6.25	-5.22	-4.46	-3.95	-3.66	-3.55	-3.63	-3.85	-4.21	-4.70	-5.29	-5.98

Fig. 7. ML Best fit (magenta) 90% (yellow) and 95% (green) CL intervals for the Normal distribution for data in Fig. 6.

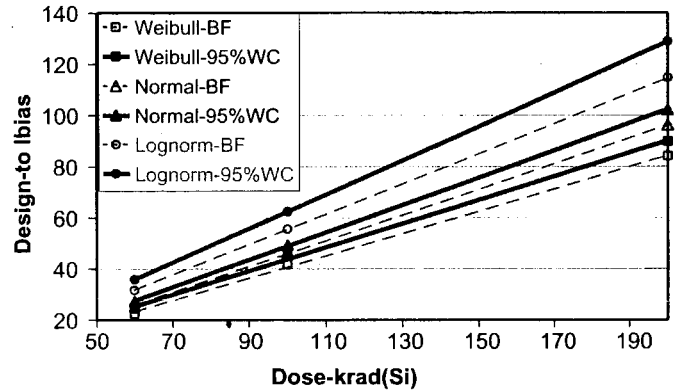


Fig. 8. Design-to I_{bias} for 99.9999% P(success)—Best Fit (open symbols) and 95% CL (solid symbols).

The linearity between dose and increased I_{bias} makes it possible to correlate the design-to I_{bias} es with equivalent radiation design margins (RDM). The 99.9999/95 values are roughly $2.5\times$ the mean values for each dose step. Factoring in the usual $2\times$ margin, this means a $5\times$ RDM ought to establish the same reliability level. While this procedure carries risk (e.g. due to process changes), the above process provides an empirical basis for the RDM, and is preferable to an across-the-board, ad hoc requirement.

Like high reliability requirements, large flight lots emphasize the distribution tails. Failure in a flight lot of N parts with no redundancy is driven by the worst performing part. For failure probability density function (pdf) $p(I_{bias})$ and cumulative distribution $Pf(I_{bias})$, the pdf for the worst-case I_{bias} is

$$p(\text{Max } I_{bias}) = \binom{N}{1} p(I_{bias}) \times (1 - P(I_{bias}))^{N-1} \quad (4)$$

Table II gives best-fit and 95% WC design-to values for flight lots of 20 and 100 RH1014s.

TABLE II: DESIGN-TO-IBIAS AFTER 200 KRAD (Si)

P(s)/Fit	# of parts	Normal	Lognormal	Weibull
99/BF	20	84.7	93.1	78.
99/95	20	89.3	101.8	82.8
99/BF	100	88.3	99.3 nA	80.2 nA
99/95	100	93.3	109.5 nA	85 nA

VI. EFFECTS OF REDUNDANCY

Effectively implemented redundancy can significantly increase reliability not just for random failures, but in some cases even for wear-out type failure mechanisms such as TID degradation. If instead of requiring all N parts in the flight lot to pass some criterion (e.g. $I_{bias} < \text{design-to value}$), only $N-M$ parts must pass (N for M redundancy), failure will be driven by the $(N-M+1)$ th worst part, and the failure pdf will be given by:

$$p(I_{bias}_{N-M+1}) = \binom{N-M}{1} \binom{N}{N-M} (P(I_{bias}))^M \times p(I_{bias}) \times (1 - P(I_{bias}))^{N-M-1} \quad (5)$$

By ensuring that the failure rate is no longer driven by the worst parts, added redundancy de-emphasizes the importance of distribution extremes. As Fig. 9 shows, when the system approaches 2:1 redundancy, the failure distribution of the redundant system is driven by the median behavior of the single-part distribution—even if the single-part distribution is bimodal or thick-tailed. A small test sample is much more likely to provide adequate understanding of a distribution's central behavior than of its tails. If an application requires large flight lots or high reliability, but testing large samples is not practical, redundancy may be the best way to increase confidence in mission success—and for simplifying qualification.

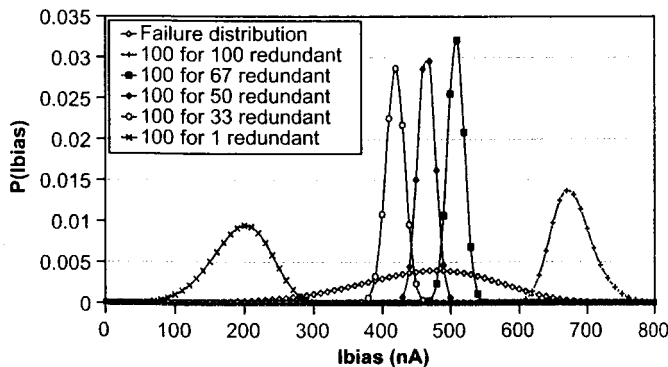


Fig. 9. For a hundred-part flight lot with a Weibull single-part failure distribution ($w=500$, $s=5.3$), increased redundancy causes the system failure distribution to approach and then surpass the median behavior of the single-part failure distribution (median behavior is achieved at 2:1 redundancy).

VII. LARGE LOT-TO-LOT FLUCTUATIONS

When lot-to-lot variability greatly exceeds intra-lot variations, the ensemble exaggerates flight-lot variability, but sampling errors due to small RLAT samples preclude meaningful inference. The first step in such cases is to characterize the variability and use trends in the data (e.g. between distribution mean and width) to infer flight-lot performance. If no trends emerge, it may be useful to assume that the mean and width are independent, estimating width with archival data and inferring flight-lot mean with lot-specific data. There are several advantages to this procedure. First it is physically reasonable, since one can view shifts in the mean radiation response as arising from lot-specific process changes, while intra-lot variations could be due to tolerances within the process. Second, while the sample mean for small samples converges rapidly to the parent-distribution mean, convergence of the variance of the distribution of sample variances (s^2) has a more complicated dependence on sample size n

$$\text{var}(s^2) = (1/n) \left[M_4 - \frac{n-3}{n-1} \sigma^4 \right] \quad (6)$$

where M_4 and σ^2 are the 4th and 2nd central moments of parent distribution. As n increases, expression 3 approaches convergence as n^{-1} . For a Normal RRD, sample sizes required to determine the sample mean within an error of σ are 4 parts for 90% confidence, 5 parts for 95% confidence and 9 parts for 99% confidence. Required samples scale as the inverse square of the allowable error. For other distributions, the sample size can be estimated numerically, but the same rule applies: parameters that largely affect where the distribution is centered (e.g. normal mean, lognormal mean and Weibull width) will generally converge rapidly even for small sample sizes, while those that affect mainly the distribution width and shape (e.g. normal and lognormal σ or Weibull shape) will benefit most from the additional statistics supplied by the ensemble distribution.

Mean shifts in gain Δh_{FE} for 2N2907 PNP transistors after 300 krad(Si) vary significantly from lot to lot, but are grouped tightly within a lot (especially for lots 1-4, 9 and 11, for which all parts are from a single wafer). (See Fig. 10). We fit each lot to a Normal distribution using an ML fit. We also fit the data allowing the mean to vary from lot to lot, but assuming a common value for σ across all lots. (See table III.) As the required confidence level increases, the reduced random sampling errors due to the greater statistics of the ensemble provide a clear advantage.

TABLE III: COMPARISON OF ENSEMBLE AND SINGLE-LOT σ .

Type of Lot	Best-Fit σ Range	95% CL WC σ Range	Best-Fit σ Across Lots	95% CL WC σ Across Lots
Single-Wafer	0.15-5.3	1.7-11.8	2.1	3.6
Multi-Wafer	952.7-11.8	6-23.3	6.8	9.8

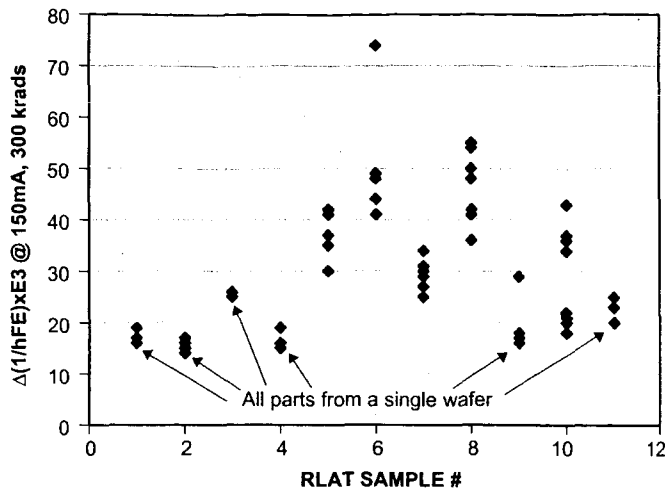


Fig. 10. Gain shifts in 2N2907 PNP transistors are consistent within wafer lots, but vary significantly across lots. Parts in lots 1-4 and 9 are from a single wafer.

VIII. CONCLUSION

Large flight lots and high reliability requirements pose problems for RHA methods that rely on small sample sizes. Sampling errors arising from the small sample size and systematic errors arising from the assumed form of the RRD can invalidate RHA analyses. In this note we have shown that sampling errors can be bounded by augmenting RLAT results with archival data to characterize radiation response variability and reduce random errors. Systematic errors from the assumed distribution form may be estimated by fitting the data to multiple forms—as we have done here for the Normal, Lognormal and Weibull distributions, mainly because they are well behaved, but have different symmetries. The results for the different distributions gauge the sensitivity of analysis to the assumed distribution form. If there is no reason to favor one distribution over the others, it is prudent to assume the distribution that yields worst-case results.

Using archival data to bound distribution pathologies and infer performance is simplest when inter-lot and intra-lot variability are comparable in size. Then the greater statistics of the ensemble distribution allow inference of flight-lot performance with greater confidence than would be possible from RLAT data alone. When lot-to-lot variability dominates, the first step is to look for trends in the data that are useful for prediction. Otherwise, using RLAT data to estimate the mean (or other parameter most affecting distribution central behavior) and the ensemble to estimate the standard deviation (or other parameter most affecting distribution shape and width).

In looking at the effects of large flight lots, it is important to use a distribution that considers flight lot size (equation 1) and system redundancy if present (equation 2).

We have also noted that redundancy not only may increase system robustness to random failures, but also to degradation and wear-out type failure mechanisms. Redundancy can simplify part qualification by driving system performance toward the central portion of the radiation response distribution—a region much more easily characterized by small sample sizes. Early in a part's history when archival data are not available, the only options for use in high reliability or large-flight-lot application may be qualification using very large samples or conservative design including significant (2:1 or more) redundancy.

REFERENCES

- [1] R. Pease, "Microelectronic Piece Part Radiation Hardness Assurance for Space Systems", 2004 NSREC Short Course.
- [2] MIL-HDBK-814, "Ionizing Dose and Neutron Hardness Assurance Guidelines for Microcircuits and Semiconductor Devices."
- [3] A. Johnston and C. Lancaster, "A Total Dose Homogeneity Study of the 108A Operational Amplifier," IEEE Trans. Nucl. Sci., NS-26, No. 6, Dec. 1979.
- [4] J. Krieg et al., "Hardness assurance implications of bimodal total dose response in a bipolar linear
- [5] M. Shaneyfelt et al., "Impact of passivation layers on enhanced low-dose-rate sensitivity and pre-irradiation elevated-temperature stress effects in bipolar linear ICs," IEEE Trans. Nucl. Sci., 49, No. 6, Dec. 2002.
- [6] A. Namenson, "Lot Uniformity and Small Sample Sizes in Hardness Assurance", IEEE Trans. Nucl. Sci. 35, No. 6, Dec. 1988.
- [7] Resources for distributions, likelihood methods and inference include NIST/SEMATECH e-Handbook of Statistical Methods, www.itl.nist.gov/div898/handbook/, 2005, mathworld.wolfram.com, www.Weibull.com, and www.resacorp.com