The Effect of Histogram Size on Histogram-Derived Pulse Parameters

N. G. Paulter

***Abstract*— The effects of the number of histogram bins on histogram-derived pulse parameters of step-like waveforms is examined. An empirical method for selecting the optimal number of bins is described.**

***Index Terms*—Calibration, histogram, pulse parameters, reference standards, step-like waveforms, waveform recorders.**

I. Introduction

PULSE generators are frequently used as reference standards to check the operation of or to calibrate high-speed waveform recorders [1]–[4], such as oscilloscopes, or to compare the performance of different waveform recorders. In particular, step-like pulse generators (SLPG's) or impulse-like pulse generators (ILPG's) are used as reference standards because of the nominally static amplitude levels on either side of the transition (for SLPG's) or pulse (for ILPG's). We will focus our attention here on the step-like waveform (SLW) output of the SLPG's because SLPG's are more commonly used in pulse metrology than are ILPG's. Specific parameters of the measured SLW's, not the SLW's themselves, are typically used as the basis of an intercomparison or performance check. For high-speed waveform recorders, the SLW pulse parameters of interest are: transition duration (risetime or falltime), pulse amplitude, and aberrations, such as overshoot and undershoot (preshoot). Typically, these parameters are extracted from the SLW using histogram methods [5]. Most of these parameters are defined by the IEEE [6], however, for the purpose of this paper, the following is a brief description of the pertinent parameters.

- *Baseline*: The nominal steady-state amplitude of the waveform before the transition. As an example, the histogram of an ideal step is bimodal, where one mode defines the starting static amplitude level of the step (baseline) and the other mode defines the final static level of the step (topline).
- *Topline*: The nominal steady-state amplitude of the waveform after the transition.
- *Pulse amplitude*—The difference between the topline and baseline values of the step-like pulse.
- *Transition duration* $t_{a\text{-}b}$: The time it takes the pulse to go from level a to level b, where a and b are usually given as percentage of the pulse amplitude. For example, $t_{10\text{–}90}$ is the time it takes the pulse to go from a level that is 10% of its amplitude to a level that is 90% of its amplitude.
- *Overshoot*: The greatest differential level around the topline that is both near the transition and in a direction away from the baseline. Overshoot is usually given as a percentage of the pulse amplitude.
- *Undershoot*: The greatest differential level around the baseline that is both near the transition and in a direction away from the topline. Undershoot is usually given as a percentage of the pulse amplitude.

The histogram of a measured SLW is a discrete count-versus-bin relationship where each bin defines a unique amplitude interval within the amplitude range of the SLW and discrete count reflects the number of amplitude occurrences of the SLW in a given bin [5]. Every bin in a given histogram has the same size, that is, has the same amplitude range, and this amplitude range is given by the quotient of the SLW amplitude and the histogram size (number of bins). The effect of histogram size on histogram-derived pulse parameters is a consequence of the mode bins (topline and baseline bins) being dependent on the histogram size and that all the pulse parameters are derived from the topline and baseline values. This pulse parameter dependence on histogram size precludes accurate assessment and comparison of the behavior of high-speed waveform recorders. Accordingly, a histogram-based pulse parameter extraction method was developed that can reduce the effect of the pulse parameter histogram-size dependence.

II. Experiment

Four SLPG's (PG_a, PG_b, PG_c, and PG_d), representative of pulse generators typically used in pulse waveform metrology, and high-speed waveform recorder calibrations were used (see Fig. 1). The pulse generators PG_a, PG_b, and PG_c are similar (same model number). The measured SLW outputs of the four SLPG's consist of 1024 samples and have one transition region with nearly zero-slope regions on either side of the transition. The near-zero-slope region after the transition, however, may not occur immediately after the transition. The near-zero-slope regions are important because they are used to determine the topline and baseline values of the waveforms from which all the pulse parameters are derived. The zero-slope requirement is also important for certain waveform processing algorithms [7]. The histogram size was varied from 256 to 8192 bins for each of the four test waveforms, thus providing histogram-size-dependent pulse parameters for each of the waveforms. The size dependence for the pulse

Manuscript received May 18, 1997; revised December 11, 1998.

The author is with the National Institute of Standards and Technology, Gaithersburg, MD 20899 USA.

Publisher Item Identifier S 0018-9456(98)09893-3.

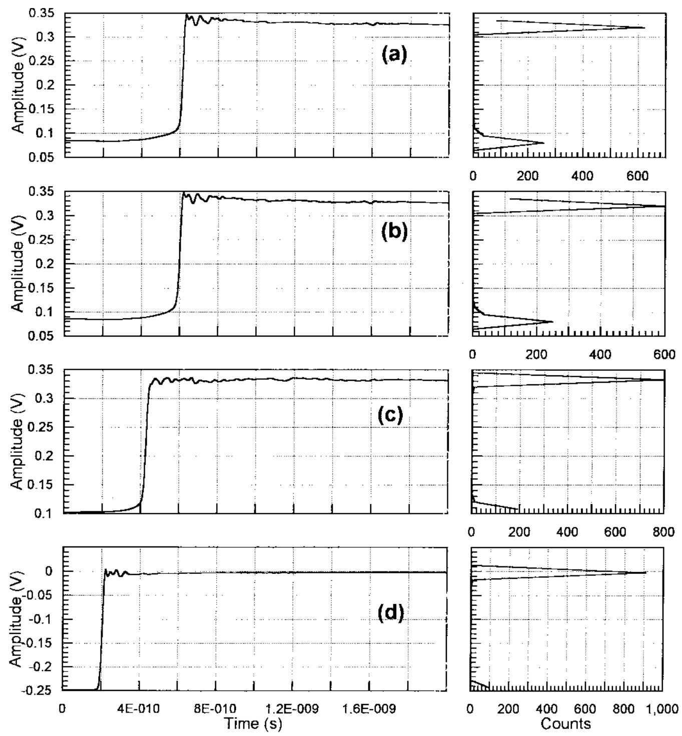

Fig. 1. Recorded data for four different pulse generators: PG_a, PG_b, PG_c, and PG_d, and their histograms (right side).

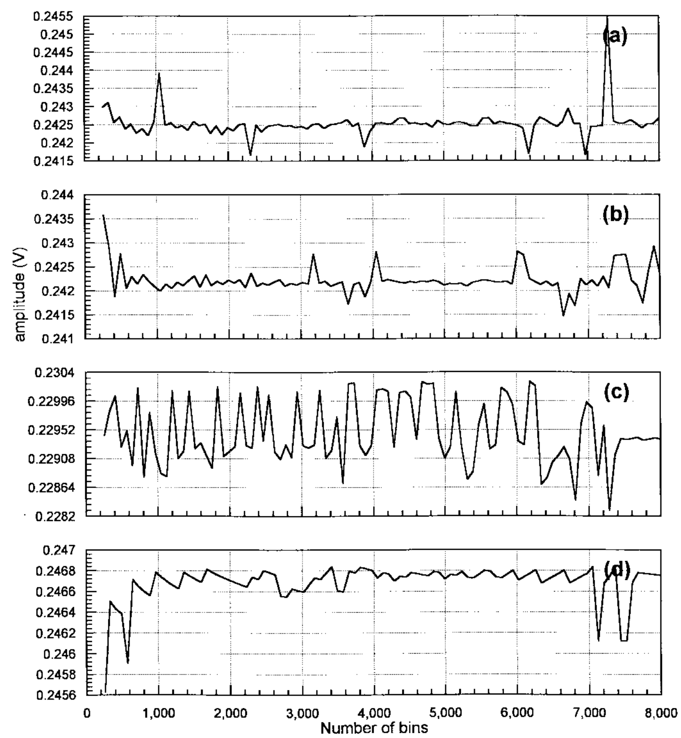

Fig. 2. Histogram-derived amplitude for waveforms shown in Fig. 1. The data in panels (a)–(d) correspond to the waveforms in panels (a)–(d) of Fig. 1.

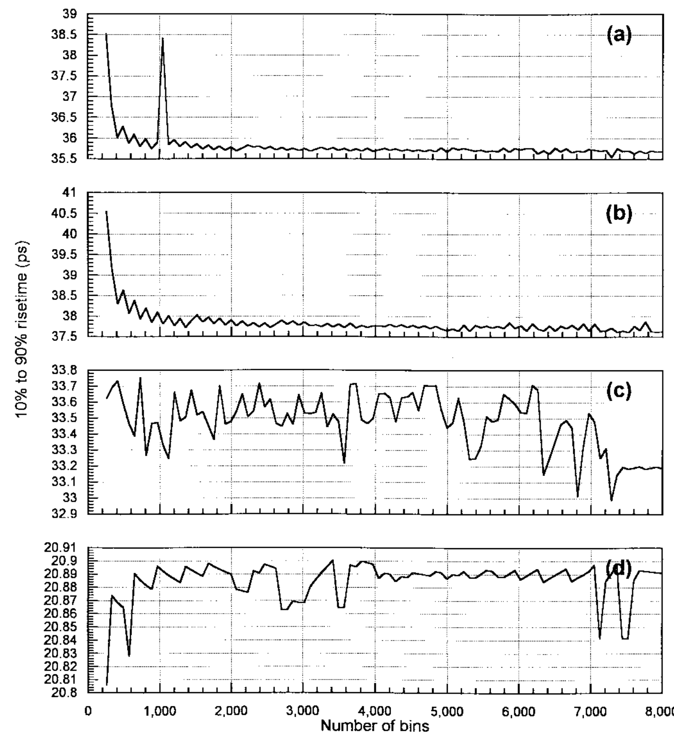

Fig. 3. Histogram-derived t10-90 risetime for waveforms shown in Fig. 1. The data in panels (a)–(d) correspond to the waveforms in panels (a)–(d) in Fig. 1.

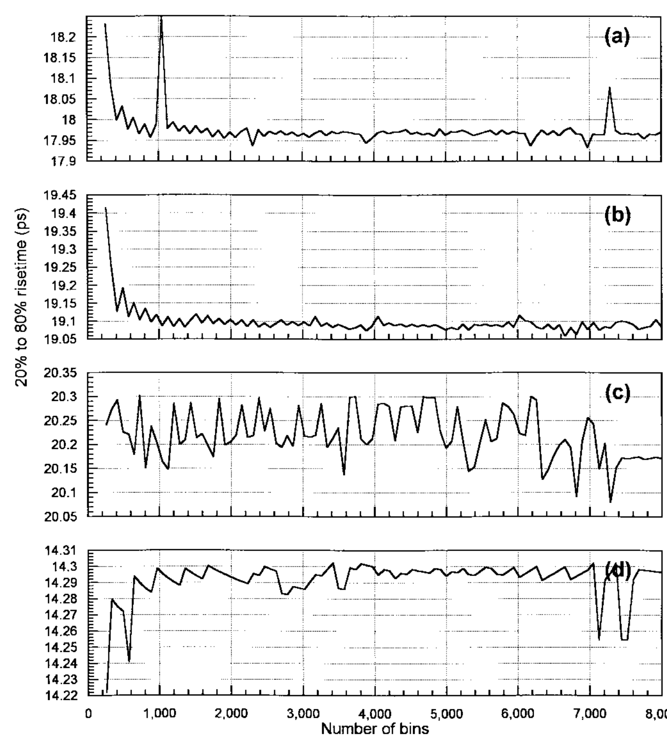

Fig. 4. Histogram-derived t20-80 risetime for waveforms shown in Fig. 1. The data in panels (a)–(d) correspond to the waveforms in panels (a)–(d) in Fig. 1.

parameters, pulse amplitude, $t_{10\text{-}90}, t_{20\text{-}80}$, overshoot, and undershoot are shown in Figs. 2–6.

The range chosen in this study for the histogram sizes 256–8192 was based on the results shown in Fig. 7. This figure shows the counts in the topline and baseline bins and the difference in counts between these bins and their adjacent bins as a function of histogram size. The reason for considering the count difference is that it gives an indication of the stability of the histogram shape. From the results of Fig. 7, it can be seen that the histograms have not settled to any nominally steady

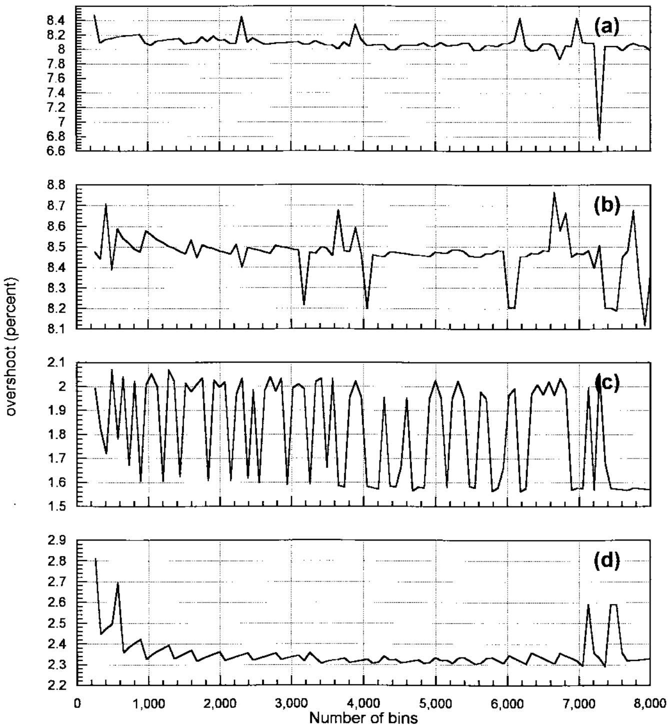

Fig. 5. Histogram-derived overshoot for waveforms shown in Fig. 1. The data in panels (a)–(d) correspond to the waveforms in panels (a)–(d) in Fig. 1.

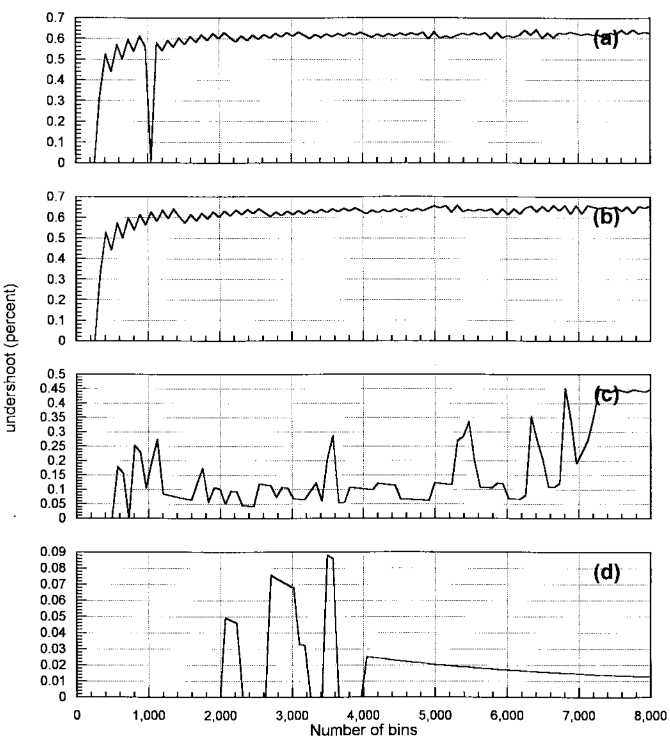

Fig. 6. Histogram-derived undershoot for waveforms shown in Fig. 1. The data in panels (a)–(d) correspond to the waveforms in panels (a)–(d) in Fig. 1.

behavior until after about 2000 bins. Since we want to find a region with steady behavior, starting the search at a histogram size of 256 bins is conservative. Beyond about 7000 bins the histogram behavior is relatively steady. There are other reasons for not using more than 8192 bins. First, the reduced number of counts per bin may cause the histogram to display structure

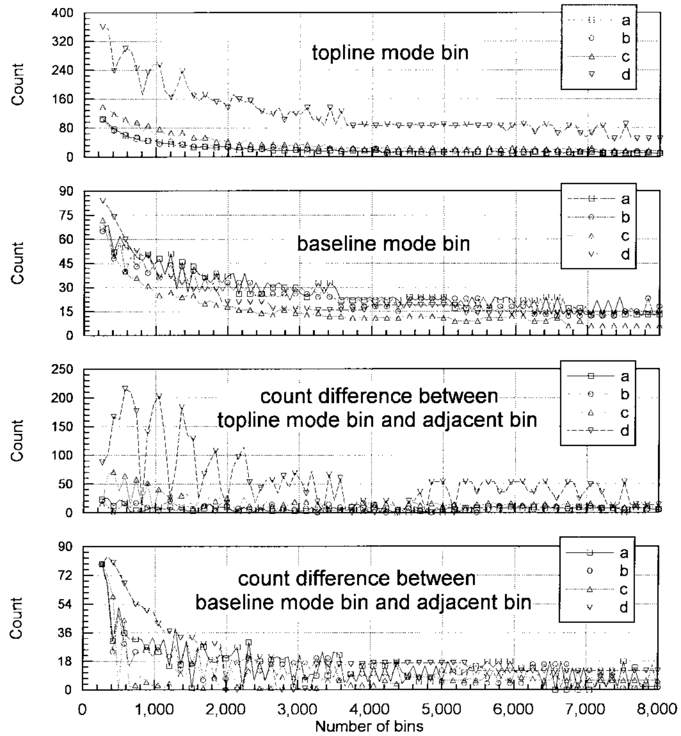

Fig. 7. Various histogram artifacts as a function of the number of histogram bins. The curves labeled (a)–(d) correspond to the curves found in panels (a)–(d) in Fig. 1.

that would make mode identification difficult or ambiguous. For example, the zero counts in the bottom two panels of Fig. 7 indicate that there are at least two adjacent bins in either the topline or baseline regions that have the same count. Furthermore, statistical significance is lost when the mode bin count is small. In a more practical sense, measurement noise limits the realistic amplitude resolution of the histogram and, thus, the number of useful histogram bins. From these figures it can be seen that it would be difficult to pick an arbitrary histogram size that would be useful for all SLW's, even though the SLW's appear similar. Consequently, a method was developed that can automatically generate a histogram that is optimal for SLW pulse parameter extraction.

To develop the new histogram method a pulse amplitude (or difference between topline and baseline bins) criterion was selected to determine the optimal histogram because all of the SLW pulse parameters are dependent on pulse amplitude. The optimal histogram is obtained from a region of contiguous histogram sizes in the pulse-amplitude-versus-histogram-size curve where the pulse amplitude dependence on histogram size is the least.

The search algorithm was written to find an optimal histogram that had between 256 and 8192 bins. To expedite the search, only 101 different histograms were generated from the possible pool of 7937 histograms (8192 − 256 + 1) therefore, making the histogram size increment equal 79.36 ($[8192 - 256]/100$). The actual number of bins, N_n for each histogram, $H[N_n]$ used in this study was computed by adding $79.36n$ and 256 where n is the histogram-size index ($0 \leq n \leq 100$) and then rounding the result to the nearest integer. The search for the optimal $H[N_n], H[N_{\rm opt}]$ was performed in three steps, where each step uses the pulse-amplitude-versus-histogram-size $\mathrm{PA}[N_n]$ curve. In the first step, the search looks for a region in $\mathrm{PA}[N_n]$ spanning five contiguous N_n where the relative change in $\mathrm{PA}[N_n]$ was 1% or less. The requirement for five N_n was chosen empirically to ensure at least one $PA[N_n]$-stable region would be found. (The wider the initial search range, the less likely an $PA[N_n]$-stable region would be found because of the 1% tolerance.) If more than one $PA[N_n]$-stable region is found, the search is repeated using a reduced (halved) relative change in amplitude. The reason for reducing the acceptable relative amplitude change in the search is that the number of $PA[N_n]$-stable regions discovered is also reduced. If still more than one $PA[N_n]$-stable region exists, the search range is expanded to include two additional N_n (one on either side of the original five). Expansion of the search range makes meeting the amplitude criterion more difficult thereby, reducing the number of discovered $PA[N_n]$-stable regions. Finally, if more than one $PA[N_n]$-stable region still exists, the smallest differential amplitude that has satisfied the previous searches is used to obtain the N_{opt}.

III. Results

For the data shown in Fig. 1, the N_{opt} based on this search are: (a) 5032, (b) 5111, (c) 7639, and (d) 4084. Figs. 2–6 show that these N_{opt} correspond to regions of pulse parameter stability. In addition, these N_{opt} also occur in quiescent or relatively quiescent zones of the direct histogram parameters (Fig. 7).

A correlation between pulse amplitude versus histogram size and the other pulse parameters (except undershoot) versus histogram size can be seen in Figs. 2–6. Also, for PG_a and PG_b, the variability of $t_{20\text{-}80}$ with N_n is greater than that of $t_{10\text{-}90}$. The reason for this observation is that the duration of the low-slope component (small $\Delta V/\Delta t$ or large $\Delta t/\Delta V$) of the transition is more sensitive to bin size (or ΔV) than the high-slope component and that the $t_{20\text{-}80}$ has a smaller low-slope component that does $t_{10\text{-}90}$: $(\Delta t/\Delta V)\Delta V$ is larger for $t_{10\text{-}90}$ than for $t_{20\text{-}80}$. The transition durations of PG_d are least affected by histogram size because the low-slope region and the concavity in the near-zero-slope region prior to the transition (the pretransition concavity) (see Fig. 1) are absent. The pretransition concavity in PG_a and PG_b affects pulse parameters because the baseline determination is ambiguous. This concavity is not seen for PG_c because of the time delay used in acquiring the data. Overshoot shows a strong correlation to pulse amplitude for all four SLW's (compare Figs. 2 and 5). On the other hand, undershoot does not show such a strong correlation (see Figs. 2 and 6).

The effects of noise on the histogram-derived pulse parameters should also be considered. Noise is a concern because the contribution of noise n_t to the signal at each discrete time instance t for different measurements is not identical. That is, $n_{t,i} \neq n_{t,j}$ where i and j denote different measurements. For averaged signals, the average noise value for each $t, \langle n_t \rangle$ is also not identical for different measurements, that is, $\langle n_t \rangle_i \neq \langle n_t \rangle_j$. Consequently, the same pulse generator and oscilloscope should not be expected to give identical results from measurement to measurement. However, the pulse parameters, which are statistically based, should be the same or very close. Therefore, it is important that the histogram accommodate the variable SLW's by using a variable number of bins. For example, if we look at panel (a) of the figures we see a sharp spike at around 1020 bins. If the spike is related to some spurious event and our histogram was fixed to use only 1020 bins, then on one day we may see a $t_{10\text{-}90}$ of 38.5 ps and, on another day, 36 ps. This variation in the observed parameter may mislead someone into thinking that either the pulse generator or oscilloscope characteristics had changed when, in fact, the observed pulse parameter variation may have been caused by histogram size. Note, this spike does not occur in panel (b), which represents data from a SLPG similar to that used for the data in panels (a).

A successive approximation method has been proposed and claimed to provide more accurate pulse amplitude determinations than the histogram method [8]. However, no details are provided in [8] as to why the histogram method is inadequate. As we have seen, the pulse parameters are dependent on histogram size and inattention to that fact can cause errors in the values of the histogram-derived pulse parameters.

IV. Conclusion

Histogram-based pulse parameters are sensitive to the histogram size (number of bins) and can result in size-dependent pulse parameters. Therefore, using a fixed histogram size to compute pulse parameters is not recommended. An empirical criterion for obtaining an optimal histogram size was found, which is based on pulse amplitude stability, and was shown to coincide with regions of stability for transition duration, pulse overshoot, and counts in the mode bins.

Acknowledgment

The author would like to thank J. P. Deyst and J. W. Allen for technical comments, and B. A. Bell for editorial and administrative support.

References

[1] D. Henderson and A. G. Roddie, “Calibration of fast sampling oscilloscopes,” *Meas. Sci. Technol.*, vol 1, pp. 673–679, 1990.
[2] W. L. Gans, “Present capabilities of the NBS automatic pulse measurement system,” *IEEE Trans. Instrum. Meas.*, vol. 25, pp. 384–388, Dec. 1976.
[3] D. Henderson, A. G. Roddie, and A. J. A. Smith, “Recent developments in the calibration of fast sampling oscilloscopes,” *Proc. Inst. Elect. Eng.*, vol. 139, pp. 254–260, Sept. 1992.
[4] W. L. Gans, “Dynamic calibration of waveform recorders and oscilloscopes using pulse standards,” *IEEE Trans. Instrum. Meas.*, vol. 39, pp. 952–957, Dec. 1990.
[5] *IEEE Standard on Pulse Measurement and Analysis by Objective Techniques*, IEEE Standard 181, 1977.
[6] *IEEE Standard Pulse Terms and Definitions*, IEEE Standard 194, 1977.
[7] N. G. Paulter and R. B. Stafford, “Reducing the effects of record truncation discontinuities in waveform reconstructions,” *IEEE Trans. Instrum. Meas.*, vol. 42, pp. 695–700, June, 1993.
[8] J. Kisela, “Traceable measurement technique for oscilloscope calibrator functions,” in *Proc. Nat. Conf. Standards Laboratories Workshop Symp.*, Monterey, CA, Aug. 1996, pp. 381–398.

N. G. Paulter, for a photograph and biography, see this issue, p. 608.