Spectral Compensation for Flow Cytometry: Visualization Artifacts, Limitations, and Caveats

Mario Roederer*

Vaccine Research Center, NIH, Bethesda, Maryland

Background: In multicolor flow cytometric analysis, compensation for spectral overlap is nearly always necessary. For the most part, such compensation has been relatively simple, producing the desired rectilinear distributions. However, in the realm of multicolor analysis, visualization of compensated often results in unexpected distributions, principally the appearance of a large number of events on the axis, and even more disconcerting, an inability to bring the extent of compensated data down to "autofluorescence" levels.

Materials and Methods: A mathematical model of detector measurements with variable photon intensities, spillover parameters, measurement errors, and data storage characteristics was used to illustrate sources of apparent error in compensated data. Immunofluorescently stained cells were collected under conditions of limiting light collection and high spillover between detectors to confirm aspects of the model.

Results: Photon-counting statistics contribute a nonlinear error to compensated parameters. Measurement errors and log-scale binning error contribute linear errors to compensated parameters. These errors are most apparent with the use of red or far-red fluorochromes (where the emitted light is at low intensity) and with large spillover between detectors. Such errors can lead to data visualization artifacts that can easily lead to incorrect conclusions about data, and account for the apparent "undercompensation" previously described for multicolor staining.

Conclusions: There are inescapable errors arising from imperfect measurements, photon-counting statistics, and even data storage methods that contribute both linearly and nonlinearly to a "spreading" of a properly compensated autofluorescence distribution. This phenomenon precludes the use of "quadrant" statistics or gates to analyze affected data; it also precludes visual adjustment of compensation. Most importantly, it is impossible to properly compensate data using standard visual graphical interfaces (histograms or dot plots). Computer-assisted compensation is required, as well as careful gating and experimental design to determine the distinction between positive and negative events. Finally, the use of special staining controls that employ all reagents except for the one of interest (termed fluorescence minus one, or "FMO" controls) becomes necessary to accurately identify expressing cells in the fully stained sample. Cytometry 45: 194–205, 2001. © 2001 Wiley-Liss, Inc.

Key terms: flow cytometry; data analysis; compensation; multicolor immunophenotyping

Compensation is the mathematical process by which we correct multiparameter flow cytometric data for spectral overlap. This overlap, or "spillover," results from the use of fluorescent dyes that are measurable in more than one detector; this spillover is correlated by a constant known as the spillover coefficient. The process of compensation is a simple application of linear algebra, with the goal to correct for spillovers of all dyes into all detectors, such that on output, the data are effectively normalized so that each parameter contains information from a single dye.

In general, our ability to process data is most effective when the visualization of data is presented without unnecessary correlations. In other words, when displaying graphs of one or two parameters, we wish to be certain that there is no contribution of other (perhaps undisplayed) parameters to the distributions being shown. This problem is particularly acute when the number of interacting parameters is greater than two; currently, there are no tools to effectively display correlated multidimensional data. It is also unlikely that our brain's processing capabilities are sufficient to discern useful information from uncompensated multiparameter data displays, no matter how they are displayed. To date, the process of compensation has been achieved primarily on the instrument (in hardware).

This has been successful because of the use of fluorescent dyes with limited overlap, such that a few pairwise

*Correspondence to: Mario Roederer, Vaccine Research Center, NIH, 40 Convent Dr., Room 5509, Bethesda, MD 20892-3015.
E-mail: Roederer@drmr.com

corrections could render the data sufficiently orthogonal, and such that rectangular sort gates, quadrant analysis, and isotype controls could be used accurately. Under these circumstances, it has been relatively straightforward to achieve proper compensation. Proper or correct compensation is achieved when the compensated data in each detector have no bias in the fluorescence distribution that is related to the intensity measured in any other detector.

However, given the advent of the use of several new fluorochromes with multiple spectral overlaps, compensation is no longer simple enough to be achieved by manual interaction at the instrument. Thus, software that could automatically compensate the data postacquisition became the modus operandi for experiments utilizing more than 4 or 5 colors, and even some 3- or 4-color experiments (1,2). It is also important to note that correct compensation for more than 2 colors can almost never be achieved using the standard interface of adjusting compensation coefficients (rather than spillover coefficients), because of the interdependence of these values (3,4).

It is important to note that compensation cannot take into account dye-dye interactions or dye-cell interactions which change the fluorescence spectrum of a given dye. Under conditions where the spectrum of a dye may change due to experimental variables, then additional parameters must be collected to quantify this change, and more complex analysis than simple compensation must be utilized. Such analysis is beyond the scope of this paper, which deals exclusively with the case where dye spectra behave predictably and uniformly, as is the case for the vast majority of immunofluorescence experiments.

In the development of our 11-color flow cytometric technology (3,5,6), we discovered that even properly compensated data appeared to be undercompensated, despite the concomitant appearance of a large number of events on the axis. This was previously noted by Stewart and Stewart even for 4-color experiments (7). Indeed, what we observed was a significant "spreading" of the distribution of compensated data compared to the autofluorescence distribution of unstained cells.

One important contribution to this apparent error in compensation was the use of inaccurate log amplifiers. The analog circuits that convert linear input signals into log-scaled output signals, upon which the compensation is performed, are imperfect. The most serious problem with these circuits is that their dynamic range, purported to be 4.0 decades, actually ranges anywhere from 3.7–4.2 decades. The effect of this inaccuracy on compensation is insidious: data at fluorescence intensities below the compensation standard can be either under- or overcompensated (for log amps with less or greater than 4.0 decades, respectively); data at fluorescence intensities above the standard will be incorrectly compensated in the opposite direction. In addition, log amps typically have a variability as much as ±5% at different positions in the scale.

In order to overcome this problem, we carefully calibrated our log amplifiers, only to discover that while we significantly improved the accuracy of compensation, there was still significant spreading occurring (Roederer

and Bigos, unpublished findings). Another solution was to perform as much hardware compensation as possible, since hardware compensation occurs prior to the error-inducing logarithmic correction. This partial compensation reduced error significantly (8), but far from eliminated it.

In order to better understand the sources and impacts of measurement errors on compensation, I developed a mathematical model of the data collection, compensation, and visualization process that allowed for precise control of a number of important variables, including the relative amount of light at the detector, a base measurement error in the signal quantitation, the "binning" effects of data discretization, and the degree of spillover between detectors. In addition, I collected immunofluorescence data under different conditions to confirm the basic conclusions of this analysis.

Compensation functions perfectly well without needing to take into account autofluorescence or any other dye-independent background value (or logarithmic amplifier offsets). This was explained in more detail previously (4). Alternatively, autofluorescence can be modeled in compensation corrections (1); this process leads to similar absolute values of autofluorescence intensity before and after compensation. In the absence of this more complex correction, it is only important to have the autofluorescence intensity in all collected parameters at approximately the same signal level.

I will show that there are two distinct types of errors that contribute to imprecise compensation beyond the inaccuracies of log amplifiers discussed above: errors arising from photon-counting statistics, and measurement errors. These are distinct in that the former are nonlinear, while the latter are linear. I demonstrate that in the presence of detectable levels of these errors, it is not possible to properly set compensation by visual methods (i.e., relying on dot plots or histograms); nor is it possible to accurately analyze data using quadrant gates or control samples based on isotype controls in all channels. Importantly, these results hold true irrespective of the use of newer digital electronics that obviate the use of log amplifiers.

This paper is not meant to be a comprehensive analysis of errors in fluorescence measurements. Rather, I present a model of those errors that are sufficient to explain the artifacts observed for compensated data, and to alert researchers how to better analyze data in the context of such artifacts.

MATERIALS AND METHODS
Data Analysis

Flow cytometric data were gated, compensated, and displayed using FlowJo version 3.3.3 (Tree Star, San Carlos, CA). Data modeling was performed and displayed using JMP version 4 (SAS Institute, Cary, NC). A specially modified version of FlowJo was used to generate the output in Figure 1 (left).

Model

The simple model was developed using the JMP spreadsheet application. This model simulates an experiment using a primary "measurement" detector to measure the fluorescence of the experimental fluorochrome, and a secondary "spillover" detector into which the experimental fluorochromes exhibit spectral overlap. The specifications for this model are: 1) Events in the measurement channel are evenly distributed from the lowest to the highest value; in general, 40,000 events were modeled. 2) Autofluorescence (AF) in the spillover channel is normally distributed in the log domain, and is centered at $10^{0.5}$ with a standard deviation of 0.125 decades (Equation 1). 3) The absolute fluorescence in the measurement channel (F) is converted to a log-scaled channel value according to the number of channels C, given that the dynamic range is 4.0 decades (Equation 2). This channel value, CF, is an integer value. To generate the scale signal in the primary detector (DF), a random value between 0–1 (intrachannel randomization) is added to the integral channel value and converted back into the linear fluorescence domain (Equation 3). 4) The scale signal in the spillover detector (Ds) is the sum of AF, the spillover fluorescence from the primary detector (s × DF), an error term relating to the photon counting statistics (where the relative photon count is P), and an error term relating to measurement error E (Equation 4). 5) The compensated spillover parameter channel value (Cs) is then converted to log scale, and "binned" according to the number of channels C and number of decades of dynamic range N (in Figs. 2–4, N = 6) used to store the data (Equation 5). In accordance with specification 1, the channel values of the events in the measurement channel (CF) are evenly distributed from 0 to C. In these equations, the function Rn generates a random number with a normal distribution around zero and a standard deviation of 1, the function Ru generates a random number with a uniform distribution between 0–1, and the function int returns the integer portion of the expression:

$$AF = Log_{10}(Rn \times 0.125 + 0.5) \qquad (1)$$

$$C_S = int\left[Log_{10}(F) \times \frac{C + Ru}{4}\right] \qquad (2)$$

$$D_F = 10^{(Cs \times 4/C)} \qquad (3)$$

$$D_S = AF + s \times D_F + Rn \times \sqrt{\frac{s \times D_F}{P}} + Rn \times E \times s \times D_F \qquad (4)$$

$$S_S = int\left[Log_{10}(D_S) \times \frac{C}{N}\right] \qquad (5)$$

In the model, the values s, P, E, and C were parameters that were studied for this paper. In general, the output graphs shown in Figures 2–4 are graphs of C_S vs. C_F.

Figure 4 also shows a graph of D_S vs. C_F. Channel numbers less than zero were set to a value of zero.

Cell Staining and Flow Cytometric Analyses

Human PBMC were obtained by Ficol-Hypaque centrifugation; at least 10^6 cells were used for each stain. Cells were stained at room temperature for 15 min with FITC CD3, Cy5PE CD8, and PE CD4 (PharMingen, San Diego, CA) for Figure 1, or with APC-conjugated anti-human CD57 (purified antibody obtained from PharMingen and conjugated with APC obtained from ProZyme, San Leandro, CA, as previously described (9)) for Figure 5, and washed three times with staining medium (biotin, flavin-deficient RPMI supplemented with 4% newborn calf serum and 0.02% sodium azide). Data were collected on a modified FACSVantage SE (Becton Dickinson, San Jose, CA).

RESULTS
Striped Data in Compensated Bivariate Plots

Bagwell and Adams (1) previously described an artifact of visualization of compensated data, in which bivariate plots of compensated parameters appear "striped" (Fig. 1). As they describe, such striping is a direct consequence of the fact that the data have been divided into discrete "bins" for efficient storage purposes. Typically, the binning process is used to store intensity data in either 256 or 1,024 channels. For a 4-decade log scale, this means that each bin covers 3.7% (256 channels) or 0.91% (1,024) of the signal. In other words, for a 256-channel resolution, all events between the fluorescence intensities of [f] and [f × 1.037] will be grouped together in the bin starting at f.

The consequence of transforming the data into discrete values becomes visually apparent when compensation is applied. For example, using the data in Figure 1, consider the events with a CD4 fluorescence intensity of 40. There is a range of (uncompensated) CD8 fluorescences associated with these cells around a value of 10. Two vertically adjacent channels may have a CD4 intensity of 40, and CD8 intensities of 10.00 and 10.91 (for the 1,024-channel data). Given a 24% spillover coefficient, compensation removes 24% × 40 = 9.6 fluorescence units from each of these two events, resulting in compensated intensities of 0.40 and 1.31. In the first decade, these two fluorescence values are separated by nearly 84 channels. No events can exist in the intervening 83 channels, since the two events are as close as possible to begin with. Hence, there is the appearance of striations.

Bagwell and Adams (1) solved this problem by adding a uniform random value of between 0–1 to the channel number before compensation. As shown in Figure 1, this process leads to an even distribution of compensated events, with no evidence of striping in the data.

At first glance, it may seem that the process of intrachannel randomization adds error to the data, and therefore should be avoided as it does not faithfully reproduce the "true" distribution (10). However, the amount of error it adds is unbiased and small: the position of an event
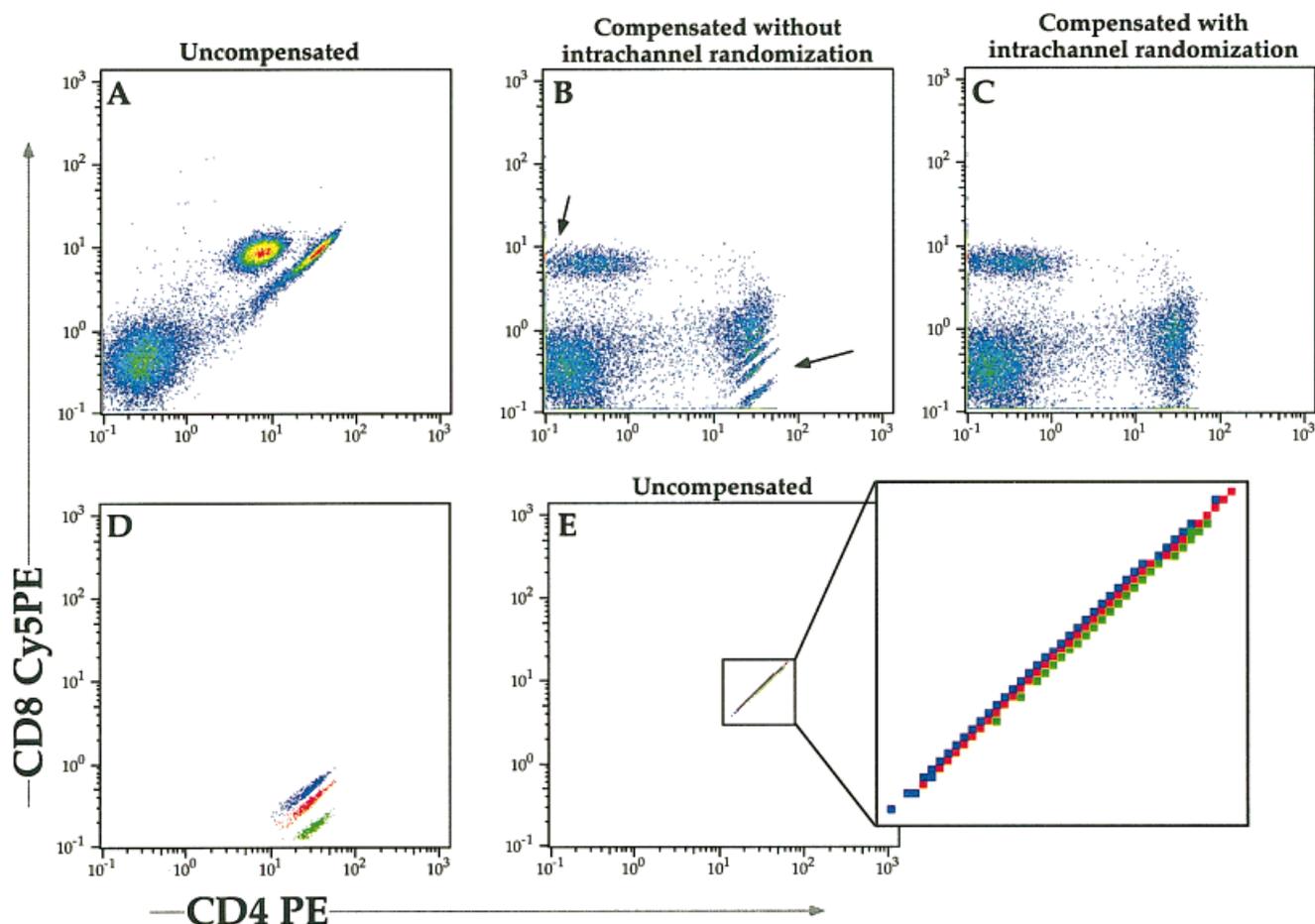
FIG. 1. Intrachannel randomization improves visualization of compensated data. Left: Three-color immunofluorescence data were collected uncompensated. Single-stained compensation controls were used to calculate the proper compensation matrix. Graphs were gated for lymphocytes. **A:** Uncompensated pseudocolor plot of CD8 versus CD4. **B, C:** Same data after proper compensation, using no intrachannel randomization (B) or randomizing every measured event within a single channel (1) prior to compensation (C). Arrows in B indicate striations resulting from discretization of data prior to compensation. **D:** Three of the striations in B were individually gated and displayed in separate colors. The slight spread in these striations is due to the effects of compensation from the undisplayed fluorescein (CD3) channel; without the contribution of that third channel, each striation would be a thin line of one bin's width. **E:** Same data as D, but shown uncompensated. **Inset:** Each of the three striated groups results from essentially a single-channel ratio within the original data. Thus, it is impossible to have events between the striated distributions in B and D, since there are no events in between the data in the original data: they represent adjacent channel ratios.

without intrachannel randomization is known to within $\pm\sqrt{(1/12)}$ channels; randomization leads to an error of $\pm\sqrt{(1/6)}$ channels (data not shown). In a typical 1,024-channel (4-decade) collection, this corresponds to adding an average error of $\pm0.1\%$, which is small compared to other unavoidable errors present in the measurement. Therefore, the intrachannel randomization, which significantly aids visualization of compensated data, is highly recommended.

Effect of Binning and Measurement Error on Compensation

For most current flow cytometry instruments, where complete compensation of all potential spillovers is not possible because of a limitation on the number of corrections that can be applied, postacquisition compensation must still be performed. Therefore, it is of interest to determine the effect of binning the data on the accuracy of compensation. As noted above, the average error introduced by binning the data is ±0.29 channels (or, with intrachannel randomization, this increases to ±0.41 channels). To determine the effect of this and other errors on compensation, I developed a model of compensation (see Materials and Methods) in which the parameters related to measurement error, number of bins used to store the data, spillover percentage, and others could be varied in a defined way. Note that the binning error is only applicable to data *stored* with logarithmic scaling, where the error is proportional to intensity; if data are *stored* in the linear domain, the binning error is inversely proportional to intensity and thus becomes unimportant (assuming that linear domain data are stored with at least 18 bits (262,144 channel) resolution to achieve the required low-signal precision).
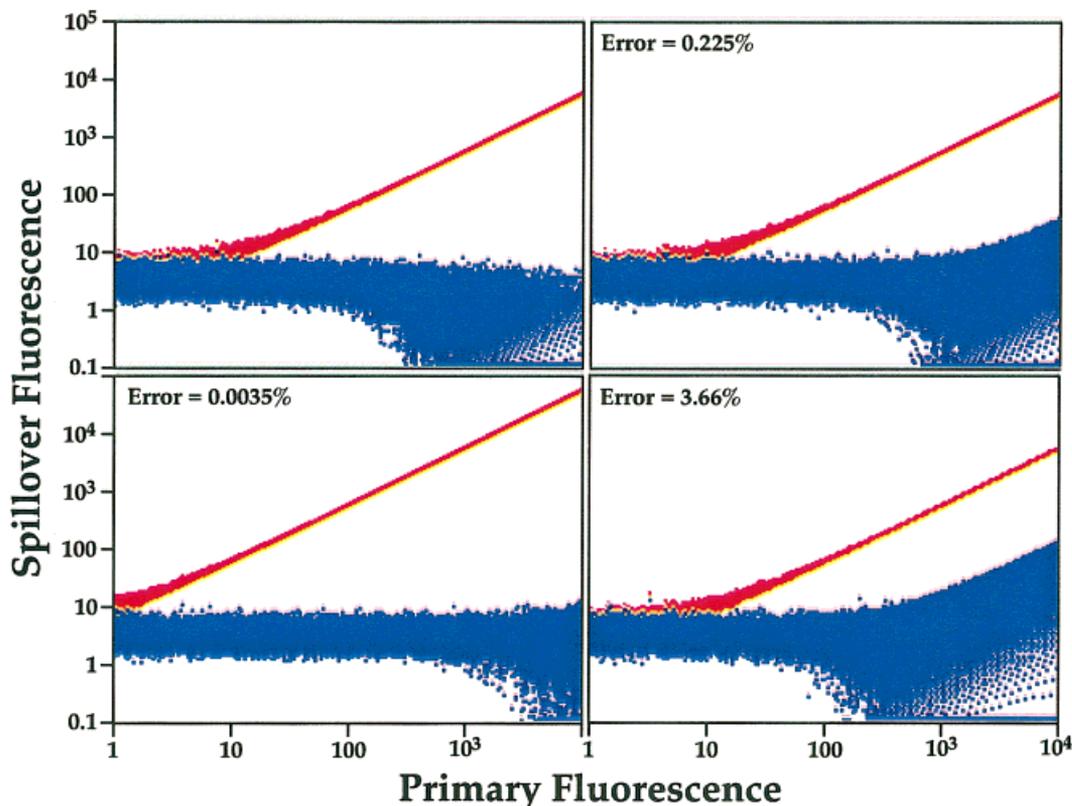
FiG. 2. Effect of measurement error and log-scale binning on compensation. A model of the compensation process was derived to examine the effect of photon-counting statistics, measurement error, and degree of spillover on compensation (see Materials and Methods). In this hypothetical model, cells are stained such that the entire population covers a continuum of intensity from autofluorescence to the top of the measurement scale. Here, the distributions shown in red are the uncompensated data; those shown in blue are the properly compensated data. The measurement channel (abscissa) is a 4-decade log scale. The spillover channel (ordinate) is shown with 6-decade scaling; one can consider the range from 1–104 to be the typical collected output range. The purpose of showing 6 decades is to better illustrate deviations in the distribution. In these evaluations, the number of photons was considered infinite, such that there was no photon-counting error; the spillover was 20% for all cases. **Top left**: This model shows the effect of "proper compensation" when log-scale binning effect is ignored. By binning data into discrete values on a log scale and using the scale intensity value of the left edge of the bin as the intensity for all events in that bin, the average error for each event is -0.5 ± 0.29 channels, which translates to a constant percentage of intensity anywhere in the scale (for 1,024 channels of 4-decade data, it is $-0.45 \pm 0.26\%$). This leads to an overcompensation of nearly 0.5%, which is evident in the data. Therefore, the intensity assigned to each bin must be the value for the center of the bin. In that case, the minimum average error for any given measurement value is 0 ± 0.29 channels (or $0 \pm 0.26\%$). The error for any given measurement must also take into account instrumentation errors; typically, these are much smaller. **Top right:** A model where the minimum measurement error is 0.23% (similar to that occurring with postacquisition compensation of 1,024-channel, 10-bit data). **Bottom left:** Minimum measurement error is 0.0035% (for 65,536-channel, 16-bit data). Bottom right: Minimum measurement error is 0.90% (for 256-channel, 8-bit data).

Figure 2 illustrates the effect of adding an intensity-proportionate error to the measurement. (Both log-scale binning errors and signal processing errors fall into this category and can be treated identically for the purposes of modeling.) There are two important conclusions that can be drawn from these illustrations. First, the spread downward in the compensated distribution (below autofluorescence, towards the axis) occurs at a much lower-stained intensity than the spread upwards. This is because the effect of the error on the distribution is unbiased (equal in both directions), but the log-scaling makes it much more apparent at the lower intensity range than at the higher intensity range. Hence, there is a significant buildup of events in the lowest channel of the compensated spillover parameter, even at measurement intensities well below where the distribution is rising. Second, because the measurement error (or binning error) is proportional to the

measurement intensity, the spread upward occurs at a relative log-log slope of 1:1. It is important to note that by decreasing the spillover, the relative positions of the minimum error slopes (both above and below the ideal autofluorescence distribution) move to the right. A 10-fold reduction in spillover decreases the error at any given intensity by an identical factor of 10.

This upward spreading in compensated data has an important implication for distinguishing "dim-positive" from negative events in the spillover (compensated) channel. Because this error process results in the compensated (negative) distribution increasing at the same rate as the uncompensated distribution, the ratio of uncompensated distribution to the top of the compensated distribution is a constant, equal to the spillover multiplied by the error in the measurement (e.g., 20% spillover $\times$ 0.5% error = 0.1% ratio). In this case, it will be difficult or impossible to
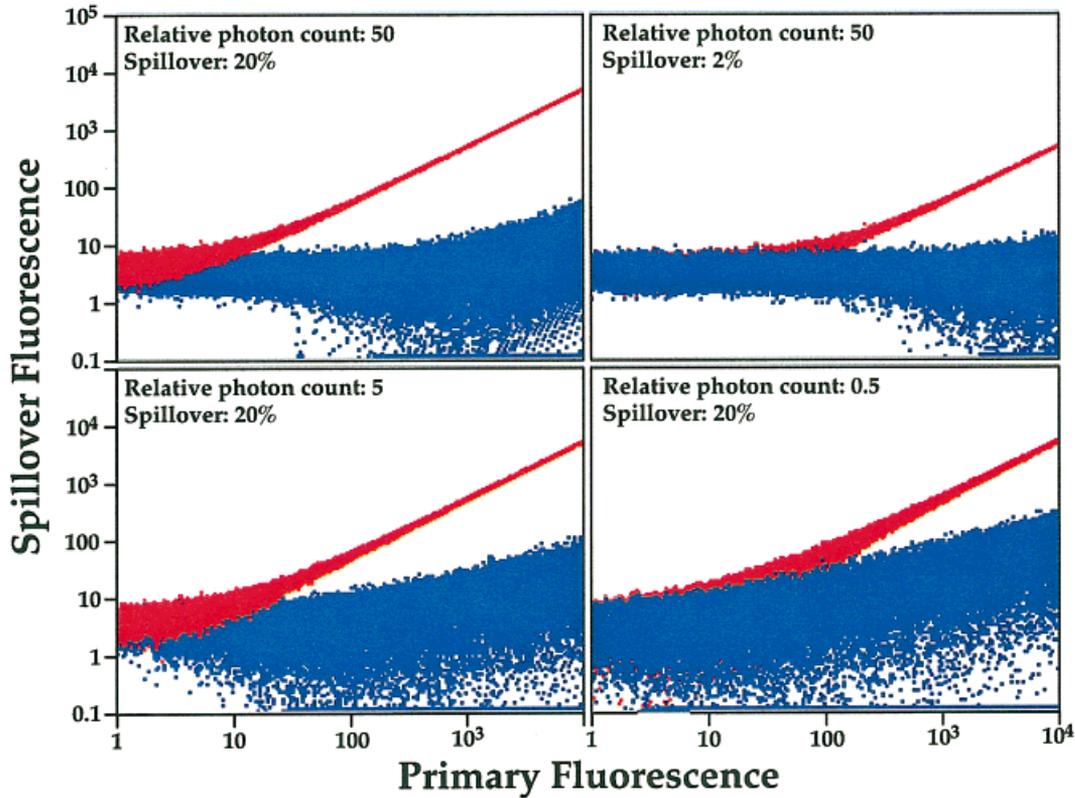
FIG. 3. Effect of low-emission intensity on compensation. Model and displays are as for Figure 2. For each of these distributions, an intrachannel correction of +0.5 channels was used (see text), and the data were considered to have been collected in 1,024-channel mode. In each graph, the relative number of emitted photons and the relative spillover percentage are shown.

distinguish events having less than 0.1% of the uncompensated spillover fluorescence from autofluorescent cells.

Effect of Limiting Emission Light on Compensation

Binning errors can be overcome by storing the data in the linear domain (as new, "digital" electronics cytometers can, although it is important to recognize that such data require at least 18-bit resolution to achieve the necessary precision at low signal levels). Other proportional measurement errors are typically extremely small by comparison. However, a fundamental measurement error that can never be overcome is one arising from counting statistics. For most cytometry applications, the number of photoelectrons in the photomultiplier tube (PMT) detector is typically in the range of 1–10^5, depending on the signal intensity. For example, autofluorescence in the fluorescein or phycoerythrin detectors (for lymphocytes) is typically below 10 photoelectrons. The error in this measurement must be at least as great as the counting error, which is the square-root of the count, i.e., 10 ± 3.2 ($\pm 32\%$). Even at 10^4 photoelectrons, which puts the signal into the third decade of fluorescence, the counting error is $\pm 1\%$. Given that measurement errors contribute to the spread in compensated parameters (Fig. 2), I included this type of error in my model to model the effect of limiting light levels on compensated parameters.

Figure 3 illustrates the effect of decreasing the number of photons (or photoelectrons at the first PMT dynode) on compensation. There are four conclusions that can be drawn from Figure 3: 1) Reducing spillover (Fig. 3, top left versus top right) decreases the "error" in the compensated distribution concomitantly. 2) As the number of photons available to the primary detector decreases, the error in the distribution increases concomitantly. 3) As is the case for proportionate errors (Fig. 2), the spread downward occurs at a much lower intensity than the spread upwards. 4) Finally, because the photon-counting error is proportional to the square root of the measurement intensity (i.e., nonlinear), the spread upward occurs at a relative log-log slope of 1:2. Thus, this visualization artifact cannot be corrected by overcompensating the data, because compensation is a linear process.

Visualization Artifacts in the Log Scale Domain

The log scaling that is commonly used to display flow cytometric data causes an apparent bias (asymmetry) in the spread of compensated data due to the sources of error noted above. For example, as shown in Figures 2 and 3, the spread towards the axis is apparent at much lower intensities than the spread upwards. In addition, it seems that the bulk of the distribution of the compensated data moves towards higher intensity, giving the impression
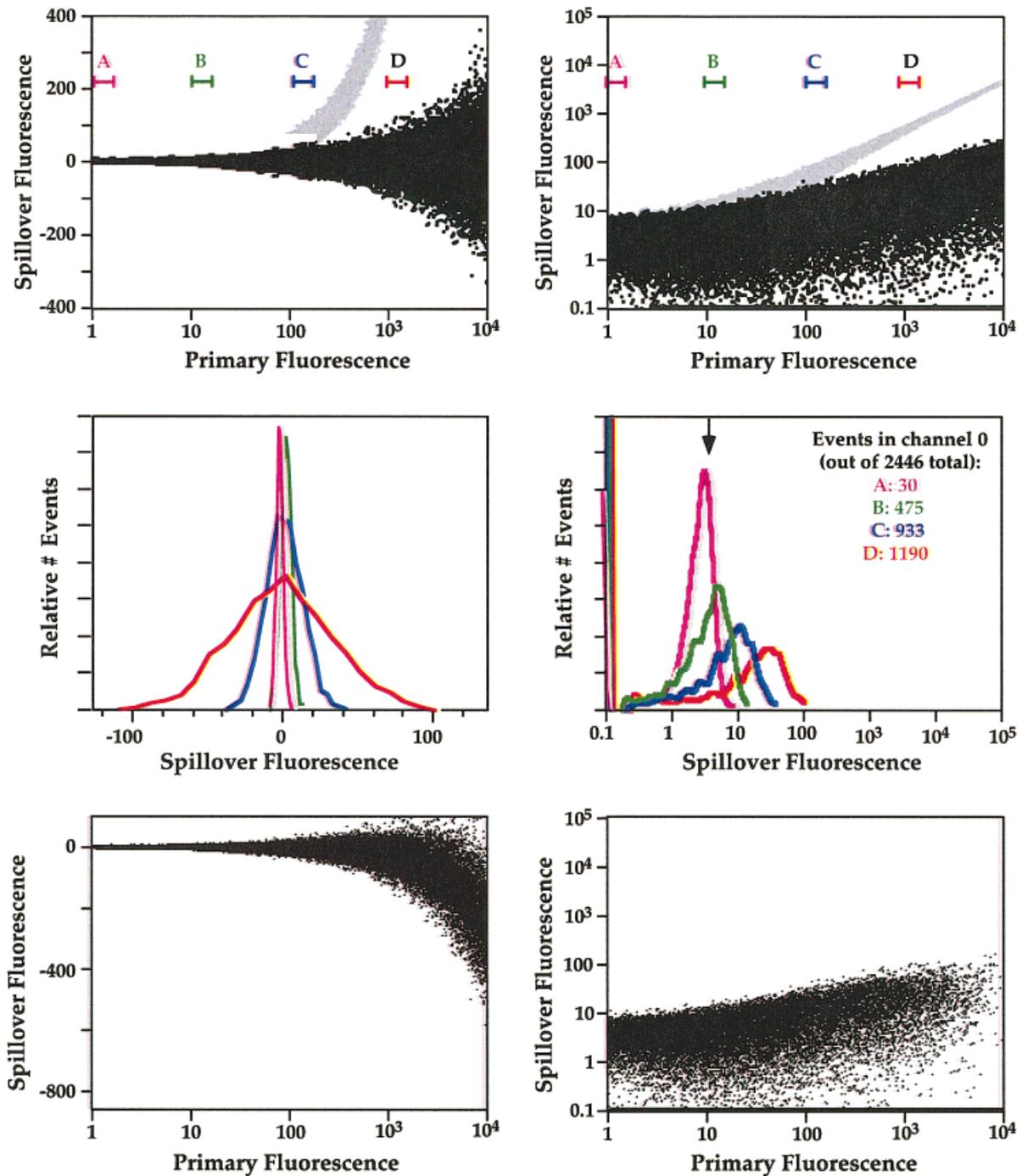
FIG. 4. Visual estimation of proper compensation is impossible in the presence of detectable error. The data in one of the models of Figure 3 (relative photon count = 0.5, spillover = 20%) are shown in different forms here to illustrate the cause of the apparently asymmetric spreading of compensated data. **Top right:** Log-log plot of compensated (black) and uncompensated (gray) data. **Top left:** Same data in linear scaling, in which the compensated parameter has not been "pinned" to a minimum value of 1.0 (minimum value on the log plot). Here it is evident that the spreading of compensated data is in fact symmetric in linear scale, and that the error contribution is therefore unbiased. Below are overlaid four histograms of the compensated spillover fluorescence parameter from regions A–D as shown. Histograms at left and right represent exactly the same data, and differ only in the scaling (linear vs. log). The change in scaling results in a change in the apparent shape of the histogram; the shape is what humans use to judge the central tendency. Therefore, we estimate that the log-scaled histogram D would have the highest median fluorescence, and A the lowest, whereas in fact all four have the same median fluorescence (as is obvious from the linear distributions at left). Arrow at right is positioned over the median of all four distributions. Since setting proper compensation requires adjusting the spillover correction until histograms A–D have the same median, visual adjustment will always lead to overcompensation of the data as we try to force the distribution in D to align with that in A. **Bottom:** Not even improperly overcompensating the data can correct for this artifact. Here, the applied compensation was 30% (i.e., 50% overcompensated). **Left:** Data in the linear domain. **Right:** In the log domain.

that the data are actually undercompensated. Figure 4 illustrates that this impression is a visual illusion created by the logarithmic scaling, and arises from our (incorrect) perception that the central tendency of a population moves with its mode rather than its median.

It is evident in Figure 4 that the error-induced spread in the compensated parameter is symmetrical in the linear domain: it is unbiased. Proper compensation is achieved when the compensated data in the spillover detector have no intensitydependent bias in distribution, i.e., the median of the events with high primary detector fluorescence is the same as that for "negative" or low primary fluorescence.

It is crucial to recognize that all of the histograms in the bottom right of Figure 4 have the same median fluorescence, and all represent properly compensated distributions. Visual setting of compensation (whether based on gated histograms such as these, or visual estimates of the distribution based on dual parameter graphs such as those at the top of Fig. 4) is impossible, because we are not capable of accurately identifying the median (or mean) of these distributions in the log-scale (see also Wood (11)). Indeed, the human tendency would be to align either the tops of the distributions or the modes of the distributions, either of which would result in significant overcompensation.

Note that no amount of overcompensation can lead to rectilinear displays by virtue of the fact that the spread in the data is nonlinear (Fig. 4). Overcompensating by 50% still results in a spreading of the compensated data for a portion of the distribution; above a certain intensity, all of the compensated data end up on the axis, with the information content completely lost.

Visualization Artifacts in Immunofluorescence Data

In order to ascertain whether the conclusions from this model of compensation are applicable to "real" data, I collected immunofluorescence stains under conditions that would limit the number of photons available to the primary detector (Fig. 5). The effect of limiting light (errors from photon-counting statistics) are apparent just by examining the uncompensated data: the correlation between the primary and spillover detectors decreases dramatically as the amount of light measured in the APC detector decreases. The impact on compensation is substantial: the compensated spillover detector signal shows increasing spread as the photon-counting errors dominate. Interestingly, even at full collection intensity, there was a significant spread in the compensated signal: a spread which increased at a 1:2 log:log slope, consistent with photon-counting statistics as the unavoidable source of the spread.

The data in Figure 5 also illustrate the log-scaling visualization problem modeled in Figure 4. For example, the contour plots reveal that the number of events on the bottom axis is substantial. Of course, this is not very evident when viewing dot plots of the data: since all the events are within a single channel, the dot plots "hide" the

large fraction of events that is present there. It is noteworthy that the histograms shown in Figure 5 have the same median fluorescence (thus represent properly compensated data), and reinforce the concept that proper compensation is not achievable by visual inspection of log-scaled data.

Summary

Figure 6 summarizes the conclusions of the models and data analysis. One aspect that was not explicitly included in the model was the contribution of photon-counting and binning/measurement errors in the spillover channel itself. These errors do not change the quality of the spread in the compensated data (i.e., the presence of both a nonlinear and a linear component)—they simply increase the magnitude of the error. Indeed, the errors modeled here should be considered the sum of the errors in all detectors that are used in the compensation of signal from a given detector. In our 11-color experiments (3,6), we found that the APC detector received significant fluorescence spillover from more detectors than any other; hence, the magnitude of errors and the concomitant spread in the compensated signal were greatest for the APC signal (5). Ironically, even though APC is one of the "brightest" fluorochromes (measured as the measured signal to autofluorescence ratio), the compensation process reduces its utility significantly in our polychromatic flow cytometry analysis.

DISCUSSION

Unfortunately, compensation is and will probably remain the least well-understood process in flow cytometric data analysis that significantly impacts the visualization and interpretation of experimental data. Many of the complexities and apparent artifacts that arise from the compensation process are difficult to intuit, largely because of the fact that most data display is in the log domain. It is crucial to understand the effects of different types of errors on compensation in order to appreciate the inherent limitations of our data, as well as how these limitations might be best addressed by changes in experimental design.

Perhaps the greatest source of "measurement error" in today's systems is the log amplifier. This analog circuitry is imperfect, in that the conversion can be accompanied by a ± 2–5% error (from "true" log conversion, which is the assumption used by the data analysis). In addition, the dynamic range of the amplifiers is rarely exactly 4.0 decades (another assumption); deviations from this assumed dynamic range lead to significantly imprecise quantitation. The degree of imprecision linearly affects the compensated data, as illustrated in Figure 2. Fortunately, the advent of digital electronics that collect and store data in the linear domain with 18-bit resolution obviates this problem. These electronics also eliminate the "binning" error, since data binning occurs in the linear domain (and thus error is inversely proportional to signal levels, rather than a constant proportion).
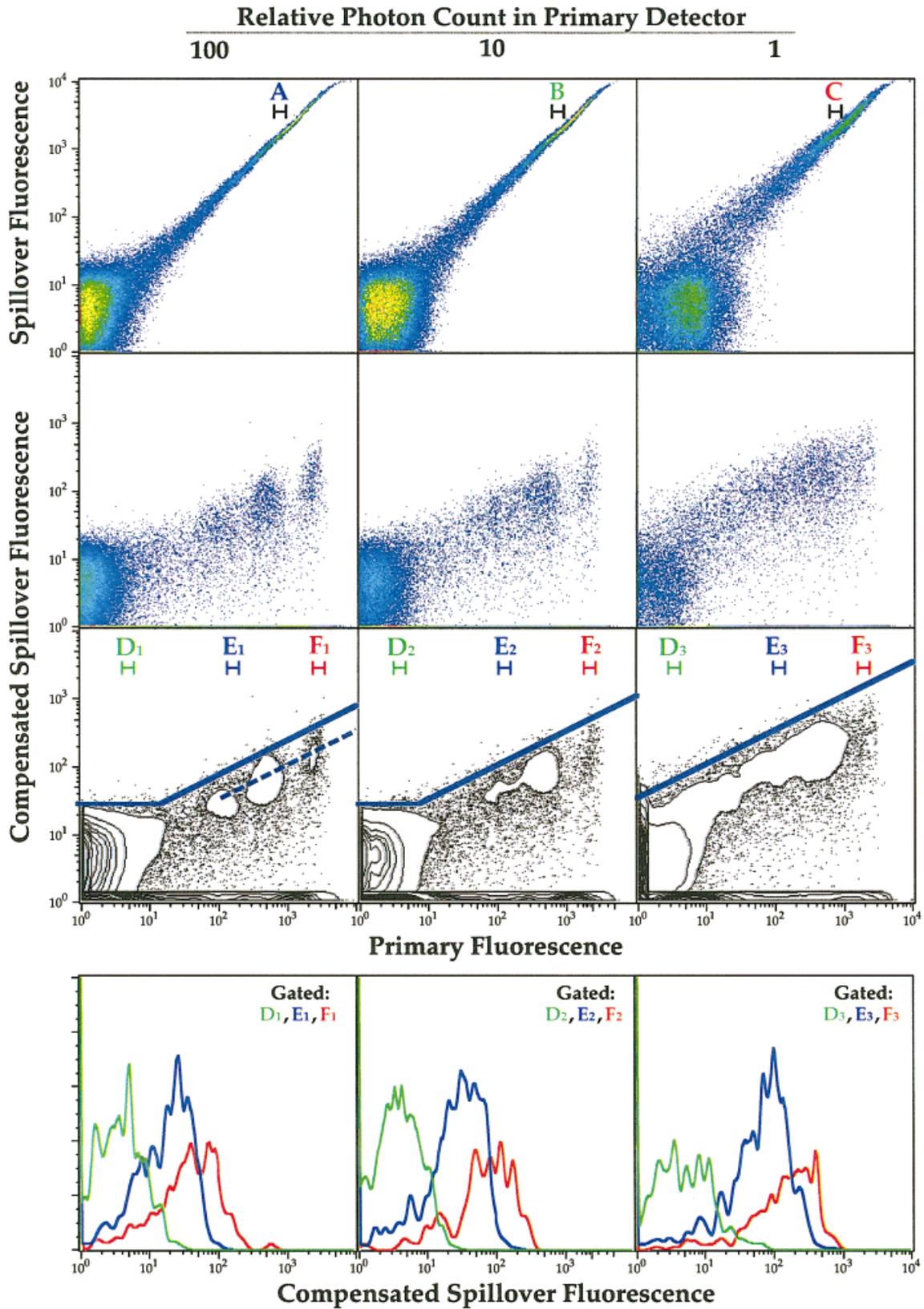
Relative Photon Count in Primary Detector

FIG. 5.

However, no electronics can overcome the fundamental counting error inherent in measuring signal levels. We will always have this nonlinear error contribution in our data, leading to the spread of compensated data as shown in Figures 3 and 5.

The degree to which the errors are apparent depends principally on two factors: the degree of spillover, and the brightness of the signal. Minimizing these errors can be accomplished by using fluorescent dyes that are as bright as possible, with as little spectral overlap as possible. Likewise, optimizing light collection will improve signal detection, e.g., by slowing the stream velocity to allow for longer interrogation times for peak-area measurements, using higher numerical aperture (NA) light collection objectives, and using interference filters that are as wide as possible without substantially increasing spectral overlap. If postacquisition compensation is necessary, then storing the data in as many channels as possible (and, if possible, in the linear domain with at least 18 bits or 262,144 channels) will also minimize error.

It is important to recognize that the visualization artifacts shown in Figures 2–6 are not simply aesthetic problems. The spreading of compensated data can significantly impact on the analysis and interpretation of data. For example, where this error is present, the use of linear "quadrant" gates, or *any* gate based on a completely unstained sample, would lead to erroneous results, since at higher intensities, the autofluorescence distribution will spread up into the "positive" gate. The best control is to stain cells with *all* reagents *except* for the one of interest, in order to determine the exact range of the negative population. This type of control can be termed "fluorescence minus one," or FMO. A nonlinear gate can be drawn based an FMO control, and applied to the fully stained sample to determine which events are positive. (FMO controls are more fully discussed elsewhere (5,12), although they were not defined by name.) Such a gate would have the same general shape as those shown in Figure 6.

Because of the propagation of errors, the more parameters that are measured, the more pronounced will be this effect. (And of course, the error is more easily detected at higher intensities, where the absolute magnitude of the error is large.) These effects undoubtedly account for the (incorrect) conclusion that the compensation setting itself depends on the intensity and presence of other stains (7); it is the exacerbated spread in properly compensated data that makes such data appear undercompensated. The propagation of errors through the compensation process can also significantly impact on the sensitivity and resolution of flow cytometry measurements: the discussions by Wood (11) and Wood and Hoffman (13) are particularly relevant when the limitations of multiparameter compensation are superimposed.

The other important conclusion is that in the presence of any significant errors of the types discussed here, it is impossible to visually set compensation properly when viewing data on a log scale. Because of our inability to correctly estimate the central tendency of the distributions such as those shown in Figure 4, manual adjustment will always result in overcompensation (which cannot, in any case, correct the apparent error). Therefore, computer-aided adjustment of compensation is necessary. At a minimum, such aid would be to provide median statistics or a graphical representation of the median as a function of intensity (as WinList does) of positive and negative populations during any user-controlled adjustment; alternatively, the computer can make all of the adjustments automatically without user intervention (as FlowJo does).

New software tools to aid in the analysis of compensated data are also desirable: tools that can automatically create "background" versus "positive" gates based on FMO control samples would be valuable. These gates would have nonlinear shapes (Fig. 6). Of course, it is important for such tools to avoid common pitfalls such as those illustrated in Figure 1 (leading to "striped" compensated data) and Figure 2 (leading to automated overcompensation).

Even with such tools, proper data analysis, interpretation, and presentation on the part of flow cytometry users will require a high level of understanding of the intricacies of compensation and the inherent limitations of compensated data. Without a proper understanding, there will continue to be a surfeit of improper conclusions based on assumptions about the rectilinearity of cytometry data.

ACKNOWLEDGMENTS

FIG. 5. Example of photon-counting and measurement errors in compensation of immunofluorescence data. Here, human PBMC were stained with APC-conjugated anti-CD57. Fluorescence was collected in the primary APC detector (using a 660/30 filter), as well as a Cy5.5APC detector (using a 710/50 filter, with a 680 dichroic splitting the emission beam). **Left column:** Fluorescence was collected with no-light reducing filter. **Middle column:** An ND1 filter was placed in front of the APC detector only. **Right column:** An ND2 filter was used. PMT voltages were concomitantly increased so that the CD8+ peak had the same fluorescence; the difference is the number of photons at the detector (the ND1 reduces by ~10-fold; the ND2 by ~100-fold; data not shown). Top row shows uncompensated data; other rows show properly compensated data. Second row (pseudocolor) and third row (contours) are different representations of the same data; contour plots illustrate the significant accumulation of events on the X axis (see Fig. 4). Overlaid on the contour plots, solid blue lines indicate approximate extent of the photon-counting error contribution (sloped lines are drawn at exactly a 1:2 log:log slope). Under conditions of lowest light (highest error), the error (blue line) is apparent even for autofluorescent (CD57−) cells. With 10-fold more light (middle column), the error line is shifted right exactly 10-fold. With the greatest amount of light (left column), the error line is only shifted about 3-fold higher (dotted blue line indicates where it would be if it were shifted 10-fold higher). This indicates that the photon-counting statistics present in the spillover detector are contributing significantly to the error term: this error term is identical for all three collections, since no neutral density filter was placed in front of the spillover detector. However, at lower light levels (middle and right columns), the counting error in the primary fluorescence detector was much higher and was the principal contributor to the spreading of compensated data. Bottom-most histograms are distributions of the gated spillover parameter for three different levels of intensity (D–F). Each of these histograms has the same median value (indicated by arrows, just above the lower axis); the apparent bias in the distribution is an artifact of the log-scaling of this data (see Fig. 4).
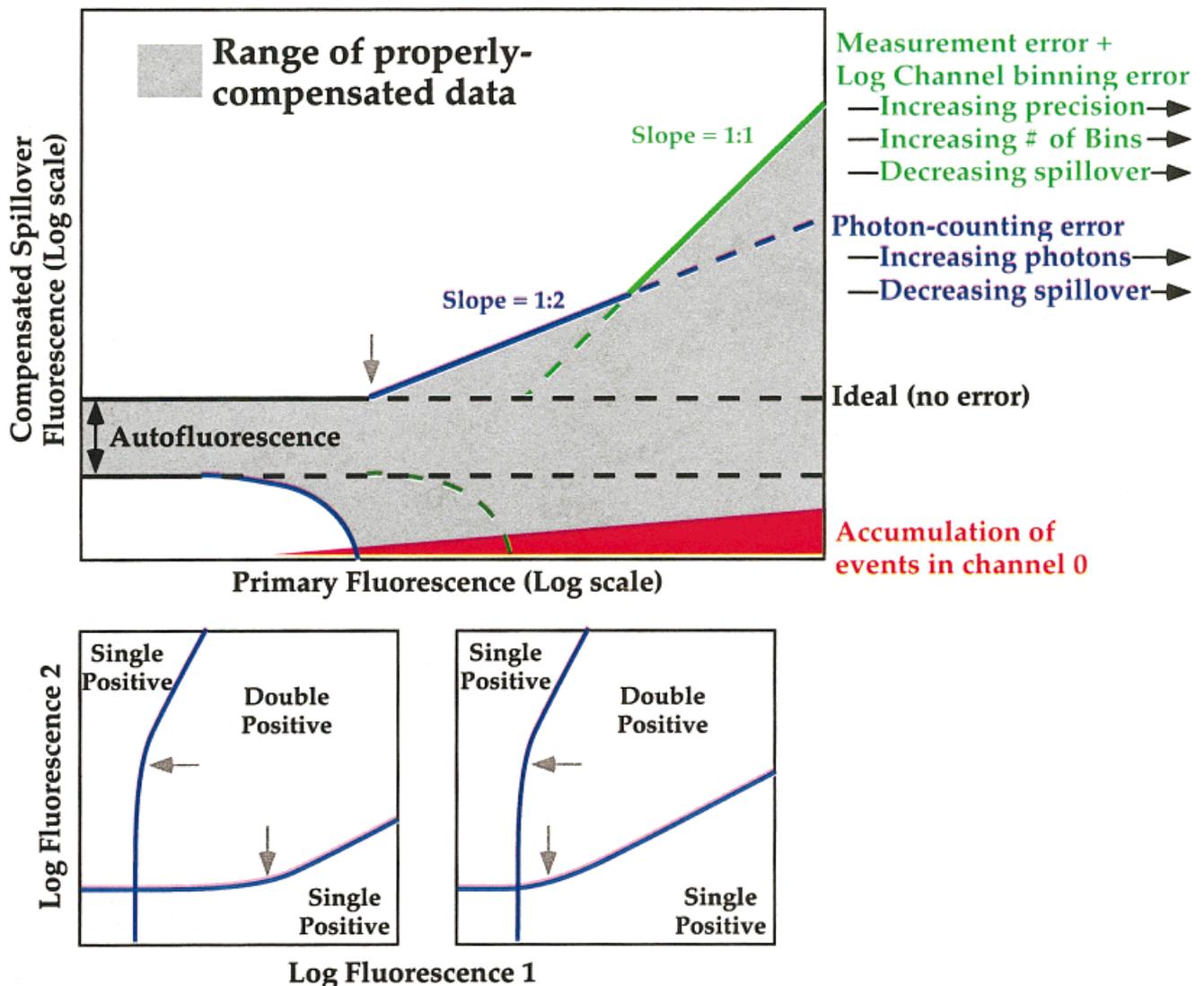
Fig. 6. **Above:** Model of compensation visualization. As shown in Figures 2, 3, and 5, there are (at least) two distinct sources of error leading to imprecise compensation, irrespective of type of electronics and instrumentation; other errors will only exacerbate these effects. Black lines show the ideal properly compensated distribution in the absence of error. The first type of error, shown in blue, is due to photon-counting statistics, and increases by the square root of the measurement intensity. The measurement error (and log-scale binning error), shown in green, increases proportionately with measurement intensity. Both asymptotically drop below the autofluorescence distribution, causing a buildup of events in channel zero of the compensated parameter (see Fig. 2). The position of the deviation from autofluorescence (gray arrow) is equal to the product of the spillover and the sum of all contributing errors. In multicolor applications, this product must be summed over all measurement channels. **Below:** Example of appropriate "quadrant" gating lines that correctly divide bivariate distributions into single- and double-positive regions. There are two variables for each measurement that control the positioning of these quadrants: the autofluorescence distribution(which defines the intersection point of the two blue lines), and the summed errors in each channel (which define the deviation from autofluorescence noted by the gray arrows). In the hypothetical example at right, the measurement errors contributing to fluorescence parameter 1 measurements are larger than those on the left.

development of the theoretical and practical concepts underlying compensation for high-end multicolor analyses, and for advice in the development of models such as the one I used here.

LITERATURE CITED

1. Bagwell CB, Adams EG. Fluorescence spectral overlap compensation for any number of flow cytometry parameters. Ann NY Acad Sci 1993;677:167–184.

2. Roederer M, Murphy RF. Cell-by-cell autofluorescence correction for low signal-to-noise systems: application to epidermal growth factor endocytosis by 3T3 fibroblasts. Cytometry 1986;7:558–565.

3. Roederer M, De Rosa S, Gerstein R, Anderson M, Bigos M, Stovel R, Nozaki T, Parks D, Herzenberg L. Eight-color, 10-parameter flow cytometry to elucidate complex leukocyte heterogeneity. Cytometry 1997;29:328–339.

4. Roederer M. Compensation. In: Robinson JP, Darzynkiewicz Z, Dean PN, Dressler LG, Rabinovitch PS, Stewart CC, Tanke HJ, Wheeless LL, editors. Current protocols in cytometry. New York: John Wiley & Sons, Inc.; 1999.

5. Baumgarth N, Roederer M. A practical approach to multicolor flow

cytometry for immunophenotyping. J. Immunol. Methods 2000;243: 77–97.

6. DeRosa SC, Herzenberg LA, Roederer M. Eleven color, 13 parameter flow cytometry: identification of human naive T cells by phenotype, function, and T cell receptor diversity. Nat Med 2000; 7:245–248.

7. Stewart CC, Stewart SJ. Four-color compensation. Cytometry 1999; 38:161–175.

8. Bigos M, Baumgarth N, Jager GC, Herman OC, Nozaki T, Stovel RT, Parks DR, Herzenberg LA. Nine color eleven parameter immunophenotyping using three laser flow cytometry. Cytometry 1999;36:36–45.

9. Roederer M. Conjugation of monoclonal antibodies. http://www. drmr.com/abcon

10. Shapiro HM, Perlmutter NG, Stein PG. A flow cytometer designed for fluorescence calibration. Cytometry 1998;33:280–287.

11. Wood JC. Fundamental flow cytometer properties governing sensitivity and resolution. Cytometry 1998;33:260–266.

12. Kantor A, Roederer M. FACS analysis of lymphocytes. In: Herzenberg LA, Weir DM, Blackwell C, editors. Visualization of compensated data. Handbook of experimental immunology (fifth edition). Cambridge: Blackwell Science; 1997. p 49.41–49.13.

13. Wood JC, Hoffman RA. Evaluating fluorescence sensitivity on flow cytometers: an overview. Cytometry 1998;33:256–259.