states that, in our opinion, have sufficient statistical merit and that have not been disproved by more reliable data. However, we promote to the Summary Tables only those states that we feel are well established. This judgment is, of course, somewhat subjective and no precise criteria can be given. For more detailed discussions, see the minireviews in the Full Listings.

4.2. *Averages and fits*: We divide this discussion on obtaining averages and errors into three sections: (1) treatment of errors; (2) unconstrained averaging; (3) constrained fits.

4.2.1. *Treatment of errors:* In what follows, the "error" δx means that the range $x \pm \delta x$ is intended to be a 68.3% confidence interval about the central value x. We treat this error as if it were Gaussian. Thus when the error *is* Gaussian, δx is the usual one standard deviation (1σ). Many experimenters now give statistical and systematic errors separately, in which case we usually quote both errors, with the statistical error first. For averages and fits, we then add the the two errors in quadrature and use this combined error for δx.

When experimenters quote asymmetric errors $(\delta x)^+$ and $(\delta x)^-$ for a measurement x, the error that we use for that measurement in making an average or a fit with other measurements is a continuous function of these three quantities. When the resultant average or fit $\overline{x}$ is less than $x-(\delta x)^-$, we use $(\delta x)^-$; when it is greater than $x+(\delta x)^+$, we use $(\delta x)^+$. In between, the error we use is a linear function of x. Since the errors we use are functions of the result, we iterate to get the final result. Asymmetric output errors are determined from the input errors assuming a linear relation between the input and output quantities.

In fitting or averaging, we usually do not include correlations between different measurements, but we try to select data in such a way as to reduce correlations. Correlated errors are, however, treated explicitly when there are a number of results of the form $A_i \pm \sigma_i \pm \Delta$ that have identical systematic errors Δ. In this case, one can first average the $A_i \pm \sigma_i$ and then combine the resulting statistical error with Δ. One obtains, however, the same result by averaging $A_i \pm (\sigma_i^2 + \Delta_i^2)^{1/2}$, where $\Delta_i = \sigma_i \Delta [\sum (1/\sigma_j^2)]^{1/2}$. This procedure has the advantage that, with the modified systematic errors Δ_i, each measurement may be treated as independent and averaged in the usual way with other data. Therefore, when appropriate, we adopt this procedure. We tabulate Δ and invoke an automated procedure that computes Δ_i before averaging and we include a note saying that there are common systematic errors.

Another common case of correlated errors occurs when experimenters measure two quantities and then quote the two and their difference, *e.g.*, m_1, m_2, and $\Delta = m_2 - m_1$. We cannot enter all of m_1, m_2 and Δ into a constrained fit because they are not independent. In some cases, it is a good approximation to ignore the quantity with the largest error and put the other two into the fit. However, in some cases correlations are such that the errors on m_1, m_2 and Δ are comparable and none of the three values can be ignored. In this case, we put all three values into the fit and invoke an automated procedure to increase the errors prior to fitting such that the three quantities can be treated as independent measurements in the constrained fit. We include a note saying that this has been done.

4.2.2. *Unconstrained averaging:* To average data, we use a standard weighted least-squares procedure and in some cases, discussed below, increase the errors with a "scale factor." We begin by assuming that measurements of a given quantity are uncorrelated, and calculate a weighted average and error as

$$\overline{x} \pm \delta\overline{x} = \frac{\sum_i w_i \, x_i}{\sum_i w_i} \pm \left(\sum_i w_i\right)^{-1/2} , \qquad (1)$$

where

$$w_i = 1/(\delta x_i)^2 .$$

Here x_i and δx_i are the value and error reported by the ith experiment, and the sums run over the N experiments. We then calculate $\chi^2 = \sum w_i(\overline{x} - x_i)^2$ and compare it with $N-1$, which is the expectation value of χ^2 if the measurements are from a Gaussian distribution.

If $\chi^2/(N-1)$ is less than or equal to 1, and there are no known problems with the data, we accept the results.

If $\chi^2/(N-1)$ is very large, we may choose not to use the average at all. Alternatively, we may quote the calculated average, but then make an educated guess of the error, a conservative estimate designed to take into account known problems with the data.

Finally, if $\chi^2/(N-1)$ is greater than 1, but not greatly so, we still average the data, but then also do the following:

(a) We increase our quoted error, $\delta\overline{x}$ in Eq. (1), by a scale factor S defined as

$$S = \left[\chi^2/(N-1)\right]^{1/2} . \qquad (2)$$

Our reasoning is as follows. The large value of the χ^2 is likely to be due to underestimation of errors in at least one of the experiments. Not knowing which of the errors are underestimated, we assume they are all underestimated by the same factor S. If we scale up all the input errors by this factor, the χ^2 becomes $N-1$, and of course the output error $\delta\overline{x}$ scales up by the same factor. See Ref. 3.

When combining data with widely varying errors, we modify this procedure slightly. We evaluate S using only the experiments with smaller errors. Our cutoff or ceiling on δx_i is arbitrarily chosen to be

$$\delta_0 = 3N^{1/2}\, \delta\overline{x} ,$$

where $\delta\overline{x}$ is the unscaled error of the mean of all the experiments. Our reasoning is that although the low-precision experiments have little influence on the values $\overline{x}$ and $\delta\overline{x}$, they can make significant contributions to the χ^2, and the contribution of the high-precision experiments thus tends to be obscured. Note that if each experiment has the same error δx_i, then $\delta\overline{x}$ is $\delta x_i/N^{1/2}$, so each δx_i is well below the cutoff. (More often, however, we simply exclude measurements with relatively large errors from averages and fits: new, precise data chase out old, imprecise data.)

Our scaling procedure has the property that if there are two values with comparable errors separated by much more than their stated errors (with or without a number of other values of lower accuracy), the scaled-up error $\delta\,\overline{x}$ is