Information content in $B \to VV$ decays and the angular moments method

Amol Dighe*
The Abdus Salam International Centre for Theoretical Physics, 34100 Trieste, Italy

Śaunak Sen†
Department of Statistics, Stanford University, Stanford, California 94305
(Received 21 October 1998; published 16 February 1999)

The time-dependent angular distributions of decays of neutral B mesons into two vector mesons contain information about the lifetimes, mass differences, strong and weak phases, form factors, and CP violating quantities. A statistical analysis of the information content is performed by giving the ''information'' a quantitative meaning. It is shown that for some parameters of interest, the information content in time and angular measurements combined may be orders of magnitude more than the information from time measurements alone and hence angular measurements are highly recommended. The method of angular moments is compared with the (maximum) likelihood method to find that it works almost as well in the region of interest for the one-angle distribution. For the complete three-angle distribution, an estimate of possible statistical errors expected on the observables of interest is obtained. It indicates that the three-angle distribution, unraveled by the method of angular moments, would be able to nail down many quantities of interest and will help in pointing unambiguously to new physics. [S0556-2821(99)00307-0]

PACS number(s): 13.25.Hw, 11.30.Er

I. INTRODUCTION

Among the available methods for studying CP violation, the decay modes of B mesons into two vector mesons, both of which decay into two particles each, are very promising mainly because of the larger number of observables at one's disposal through the angular distributions of the decays [1]. A disadvantage of having a large number of observables may be the difficulty in separating them from one another because of the correlations between them. The method of angular moments [2,3] helps in extracting the observables from the angular distributions by using judiciously chosen weighting functions. From the time evolutions of these observables, it is then possible to extract information about the lifetimes, mass differences, strong and weak phases, form factors, and CP violating quantities.

Here we will concentrate on decays of the type $B \to V_1(\to X_1 Y_1) V_2(\to X_2 Y_2)$, where B is a neutral B meson, V_1 and V_2 are vector mesons and X_1, X_2, Y_1, Y_2 are the four final state particles. We shall illustrate the technique by using the particular decay $B_s \to J/\psi(\to l^+ l^-)\phi(\to K^+ K^-)$. The other decay modes of the form $B \to VV$ might have different angular distributions, and the method of angular moments will need corresponding different weighting functions (which can always be found [3]), but the observables in all these decay modes are the same. In addition, $B_s \to J/\psi\phi$ decay holds the promise of being able to measure the lifetimes of B_s^H and B_s^L separately and, if this lifetime difference is sizable (as estimated in [4]), the prospect of measuring CP-violating quantities even without tagging [5]. By quantifying the information content in the data we can judge the relative importance of the measurement of various possible quantities. The approach we have used to analyze the information content may be used in modes of decay other than the one we have considered here.

Generally speaking, in any experiment the amount of information obtained depends on

(i) what quantities are recorded

(ii) what numerical summaries of the recorded data were used

(iii) the number of data points

(iv) the parameter values governing the outcome of the experiment

Since only the first two are under the direct control of the experimentalist, we will address those two issues in this paper. We argue that the expected information *per observation* available in the $B_s \to J/\psi\phi$ decay is substantially more when both time and angular information are used instead of using the time information alone. Moreover, we show that the method of angular moments, when used to summarize and estimate the parameters, is computationally easy to implement and efficient (in the statistical sense) in extracting information from the data.

In Sec. II, we give the angular distribution and the time evolutions of the observables for the decay $B_s \to J/\psi\phi$. The definition of information in the data about a parameter value that we will use is standard in the statistical literature and will be described briefly in the Appendix. Section III outlines why the angular information may be useful and then follows up with an analytic investigation of the additional information in the transversity angle over and above the time information. In Sec. IV, we discuss the efficiency of the method of angular moments by comparing it with the maximum likelihood method in the case of the transversity angle distribution. In Sec. V we carry out a simulation study of the method of angular moments for extracting the relevant parameters from the three angle distribution. Section VI concludes.

*Email address: amol@ictp.trieste.it
†Email address: sen@stat.stanford.edu

 ©1999 The American Physical Society

TABLE I. Time evolution of the decay $B_s \to J/\psi(\to l^+l^-)\phi(\to K^+K^-)$ of an initially (i.e. at $t=0$) pure B_s meson.

Observable	Time evolution
$\lvert A_0(t)\rvert^2$	$\lvert A_0(0)\rvert^2[e^{-\Gamma_L t}-e^{-\overline{\Gamma}t}\sin(\Delta m t)\delta\phi]$
$\lvert A_\parallel(t)\rvert^2$	$\lvert A_\parallel(0)\rvert^2[e^{-\Gamma_L t}-e^{-\overline{\Gamma}t}\sin(\Delta m t)\delta\phi]$
$\lvert A_\perp(t)\rvert^2$	$\lvert A_\perp(0)\rvert^2[e^{-\Gamma_H t}+e^{-\overline{\Gamma}t}\sin(\Delta m t)\delta\phi]$
$\mathrm{Re}[A_0^*(t)A_\parallel(t)]$	$\lvert A_0(0)\rvert\lvert A_\parallel(0)\rvert\cos(\delta_2-\delta_1)[e^{-\Gamma_L t}-e^{-\overline{\Gamma}t}\sin(\Delta m t)\delta\phi]$
$\mathrm{Im}[A_\parallel^*(t)A_\perp(t)]$	$\lvert A_\parallel(0)\rvert\lvert A_\perp(0)\rvert[e^{-\overline{\Gamma}t}\sin(\delta_1-\Delta m t)+\frac{1}{2}(e^{-\Gamma_H t}-e^{-\Gamma_L t})\cos(\delta_1)\delta\phi]$
$\mathrm{Im}[A_0^*(t)A_\perp(t)]$	$\lvert A_0(0)\rvert\lvert A_\perp(0)\rvert[e^{-\overline{\Gamma}t}\sin(\delta_2-\Delta m t)+\frac{1}{2}(e^{-\Gamma_H t}-e^{-\Gamma_L t})\cos(\delta_2)\delta\phi]$

II. ANGULAR DISTRIBUTIONS AND TIME EVOLUTIONS OF OBSERVABLES

The most general decay amplitude for $B\to VV$ takes the form [6,7]

$$A(B_q(t)\to V_1V_2)=\frac{A_0(t)}{x}\epsilon_{V_1}^{*L}\epsilon_{V_2}^{*L}-A_\parallel(t)\epsilon_{V_1}^{*T}\cdot\epsilon_{V_2}^{*T}/\sqrt{2} - iA_\perp(t)\epsilon_{V_1}^{*}\times\epsilon_{V_2}^{*}\cdot\hat{\mathbf{p}}_{V_2}/\sqrt{2}, \tag{1}$$

where $x\equiv p_{V_1}\cdot p_{V_2}/(m_{V_1}m_{V_2})$ and $\hat{\mathbf{p}}_{V_2}$ is the unit vector along the direction of motion of V_2 in the rest frame of V_1. Here the time dependences originate from B_q-$\overline{B_q}$ mixing. In our notation only a B_q meson is present at $t=0$.

For angles, we will use the same conventions as in Ref. [7]; i.e., ϕ moves in the x direction in the J/ψ rest frame, the z axis is perpendicular to the decay plane of $\phi\to K^+K^-$, and $p_y(K^+)\geqslant 0$. The coordinates (θ,φ) describe the decay direction of l^+ in the J/ψ rest frame and ψ is the angle made by $\vec{p}(K^+)$ with the x axis in the ϕ rest frame. With this convention,

$$\mathbf{x}=\mathbf{p}_\phi,\quad \mathbf{y}=\frac{\mathbf{p}_{K^+}-\mathbf{p}_\phi(\mathbf{p}_\phi\cdot\mathbf{p}_{K^+})}{\lvert\mathbf{p}_{K^+}-\mathbf{p}_\phi(\mathbf{p}_\phi\cdot\mathbf{p}_{K^+})\rvert},\quad \mathbf{z}=\mathbf{x}\times\mathbf{y},$$
$$\sin\theta\cos\varphi=\mathbf{p}_{l^+}\cdot\mathbf{x},\quad \sin\theta\sin\varphi=\mathbf{p}_{l^+}\cdot\mathbf{y},\quad \cos\theta=\mathbf{p}_{l^+}\cdot\mathbf{z}. \tag{2}$$

Here boldface characters represent *unit* 3-vectors and everything is measured in the rest frame of J/ψ. Also

$$\cos\psi=-\mathbf{p}'_{K^+}\cdot\mathbf{p}'_{J/\psi}, \tag{3}$$

where the primed quantities are *unit vectors* measured in the rest frame of ϕ.

With this convention, the three angle distribution is given by [3,7]

$$\frac{d^3\Gamma[B_s(t)\to J/\psi(\to l^+l^-)\phi(\to K^+K^-)]}{d\cos\theta d\varphi d\cos\psi}\propto\frac{9}{32\pi}\Bigg[2\lvert A_0(t)\rvert^2\cos^2\psi(1-\sin^2\theta\cos^2\varphi)+\sin^2\psi\{\lvert A_\parallel(t)\rvert^2(1-\sin^2\theta\sin^2\varphi) +\lvert A_\perp(t)\rvert^2\sin^2\theta-\mathrm{Im}[A_\parallel^*(t)A_\perp(t)]\sin 2\theta\sin\varphi\} +\frac{1}{\sqrt{2}}\sin 2\psi\{\mathrm{Re}[A_0^*(t)A_\parallel(t)]\sin^2\theta\sin 2\varphi+\mathrm{Im}[A_0^*(t)A_\perp(t)]\sin 2\theta\cos\varphi\}\Bigg]. \tag{4}$$

The time evolutions of the coefficients of the angular terms are given in Table I. Here Γ_L and Γ_H are the widths of the light and heavy B_s mass eigenstates, B_s^L and B_s^H, respectively, and Δm is the mass difference between them. $\overline{\Gamma}$ is the average of Γ_L and Γ_H. Here $\delta_1\equiv\mathrm{Arg}[A_\parallel^*(0)A_\perp(0)]$ and $\delta_2\equiv\mathrm{Arg}[A_0^*(0)A_\perp(0)]$ are the strong phases, and $\delta\phi\approx 2\lambda^2\eta$ is related to an angle of a (squashed) unitarity triangle [8], which is very small in the standard model (≈ 0.03). We will denote the values of $A_X(0)$ (where $X\in\{0,\parallel,\perp\}$) simply as A_X in the rest of the paper.

The ''transversity angle'' θ separates the CP-even and CP-odd decays. If we integrate over the remaining two angles and include the time dependence explicitly, the angular distribution in Eq. (4) becomes

$$\frac{d^2\Gamma}{d\cos\theta dt}\propto(\lvert A_0\rvert^2+\lvert A_\parallel\rvert^2)(1+\cos^2\theta)e^{-\Gamma_L t} +\lvert A_\perp\rvert^2\sin^2\theta e^{-\Gamma_H t} \tag{5}$$

or, in the form of a normalized probability distribution,

$$p(u,t|\beta,\Gamma_H,\Gamma_L)=\frac{3}{8}\beta\Gamma_L(1+u^2)e^{-\Gamma_L t}+\frac{3}{4}(1-\beta)\Gamma_H(1-u^2)e^{-\Gamma_H t}, \quad (6)$$

where $u \equiv \cos\theta$ and

$$\beta \equiv \left(1+\frac{1}{2}\frac{\Gamma_L}{\Gamma_H}\frac{|A_\perp|^2}{|A_0|^2+|A_\parallel|^2}\right)^{-1}. \quad (7)$$

The corresponding value of β in the $B_d \to J/\psi K^*$ mode is measured [9,10] to be 0.93 ± 0.03; so with conservative estimates for the breaking of flavor SU(3) symmetry, the value of β is expected to lie between 0.8 and 1.0.

III. INFORMATION IN THE TRANSVERSITY ANGLE DISTRIBUTION

By considering the case when the time (t) and transversity angle (θ) measurements are available, we will argue, in this section, that additional information on θ is substantial and worth the extra effort put into the angular measurements. In Sec. III A we will explain what makes the estimation of $\Gamma_L - \Gamma_H$ hard and why gathering the transversity angle data in addition to time is attractive. In Sec. III B we will analyze the information content analytically and determine the numerical magnitude of the information gain.

A. Why collect angular information?

If the objective is to estimate $\Gamma_L - \Gamma_H$, which is the difference between the reciprocal of mean lifetimes, one may ask, how is it possible that the angular data are useful?

Let us first consider the estimation of the parameters when we have only time information available. In that case the distribution of the lifetime (t) is given by

$$p(t|\beta,\Gamma_H,\Gamma_L)=\beta\Gamma_L e^{-\Gamma_L t}+(1-\beta)\Gamma_H e^{-\Gamma_H t}, \quad (8)$$

which is a *mixture* of the two lifetimes. With probability β we observe the lifetime of a particle with mean lifetime $1/\Gamma_L$ and with probability $(1-\beta)$ we observe the lifetime of a particle with lifetime $1/\Gamma_H$. The expectation of the observed (mixed) lifetime is $\beta/\Gamma_L+(1-\beta)/\Gamma_H$. Since

$$\left(\frac{\beta}{\Gamma_L}+\frac{(1-\beta)}{\Gamma_H}\right)[\beta\Gamma_L+(1-\beta)\Gamma_H] = 1-\beta(1-\beta)\left(\sqrt{\frac{\Gamma_H}{\Gamma_L}}-\sqrt{\frac{\Gamma_L}{\Gamma_H}}\right)^2, \quad (9)$$

the derived parameter $\beta\Gamma_L+(1-\beta)\Gamma_H$ (which we will later call θ_1) is, to the first order (when the Γ's are close to each other), the reciprocal of the expected mean observed lifetime. Thus the estimation of $\beta\Gamma_L+(1-\beta)\Gamma_H$ which is a ''mean'' parameter is not hard even if we cannot ''guess'' the decay type.

If we knew what type of decay each time measurement was coming from, then we could estimate Γ_L and Γ_H separately by using the reciprocal of the sample mean lifetimes of the two kinds of decays. We can then construct an estimate for $\Gamma_L - \Gamma_H$. Given enough observations of both types, we can get good estimates for the difference of the lifetimes. However, the identity of the decay type is not known in real data and statistical procedures have to, at least indirectly, guess it as well as possible from the available data. When the two component lifetimes are widely different, then the observed lifetime is a good clue as to the identity of the decay type. However, when the lifetimes are close to each other, the clues in the time signature alone are not decisive.

If only the transversity angle, θ, is measured, then the density of the data $u=\cos(\theta)$ is given by

$$p(u|\beta)=\frac{3}{8}\beta(1+u^2)+\frac{3}{4}(1-\beta)(1-u^2). \quad (10)$$

The distribution of the angles is very dependent on the type of the decay; therefore, by observing the angle alone one can have a fair idea as to what kind of decay has been observed. This is why angular information is useful, even though it is not directly about lifetimes.

As an illustrative example, consider the case when the two lifetimes are equally likely, i.e. $\beta=0.5$ (see Fig. 1). If $\Gamma_H/\Gamma_L=1$ and only time measurements are available, then we have no way of ''guessing'' what kind of decay we have observed. This is reflected in the fact that the *a priori* probability (the probability before making the measurement) that the decay is of the first type is the same as the *a posteriori* probability (the probability after making the measurement) and equal to 0.5. On the other hand, if the decay widths are dramatically different, say $\Gamma_H/\Gamma_L=20$, then the time measurement provides a good clue: if the time observation is very large, then it is more likely that the decaying particle is the one with the smaller decay width, and if the time measurement is very small, then it is more likely that the decaying particle has the larger decay width.

If only angular measurements are available, then when $\boldsymbol{u}=\cos(\theta)$ is very small or very large, there is a higher chance that the particle measured was of the first type because it has an angular distribution of $\frac{3}{8}(1+u^2)$ which implies more probability for large values of $|u|$. Note that the power to discriminate between the two kinds of decay by observing the transversity angle is not affected by the ratio of the decay widths. When we have *both* time and angular information, we will be able to benefit from the information contained in both which will be at least as much as the information in the angle alone.

The heuristic ideas above are graphically presented in Fig. 1. The x axis was chosen to be the percentile of the observed data (time or angle, as the case may be) so that we can plot the different scenarios on the same scale. Additionally the plot has the desirable property that all points along the x axis occur with equal probability for all four scenarios considered. (This is because the percentile of the observed data is

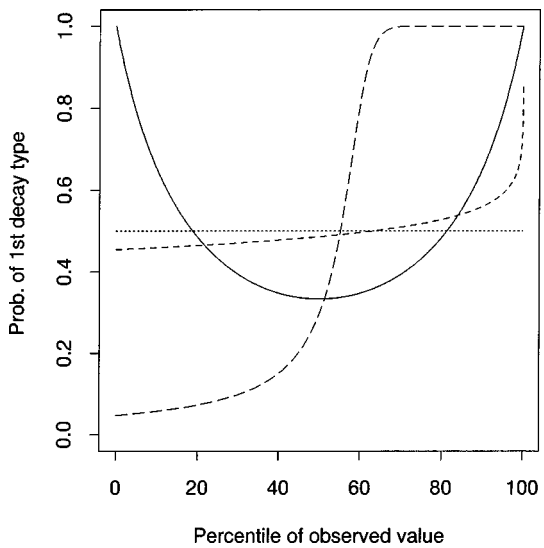

FIG. 1. The ability to guess the decay type in four scenarios: Plots of the *posterior* probability that a decay is of the first type (with mean lifetime $1/\Gamma_L$) given only angular information, $u=\cos(\theta)$ (solid line), only time data, $\Gamma_H/\Gamma_L=1$ (dotted line), only time data, $\Gamma_H/\Gamma_L=1.2$ (thin dashed line), only time data, $\Gamma_H/\Gamma_L=20$ (thick dashed line).

just 100 times the *probability integral transform*[1] of the data point.) Thus we can visually look at the four curves and compare how much they deviate *away in either direction* from the line $y=0.5$ to get an idea as to how well the data predicts the kind of decay.

The curve corresponding to $\Gamma_H/\Gamma_L=1.2$ when only time is measured is *closer* to the line $y=0.5$ than the curve corresponding to when only angular measurements are taken. This implies that when the decay widths are close (for example when the ratio is 1.2), the information in the angular data alone is greater than that in the time data alone. Only in extreme cases, such as when the ratio of the decay widths is 20, can we predict well on the basis of time alone.

B. Theoretical investigation of information content

In this subsection, we will consider the problem of extracting information from a mixture of two distributions with density of observations of the form

$$p(x|\lambda_1,\lambda_2,\beta)=\beta g_1(x|\lambda_1)+(1-\beta)g_2(x|\lambda_2). \quad (11)$$

The data with or without angular information have this form (x denotes the data from a single observation and may be a vector). According to the equation above, the data come from the distribution $g_1(x|\lambda_1)$ with probability β and from the distribution $g_2(x|\lambda_2)$ with probability $(1-\beta)$. This setup is more general than the one we have, but it enables us to analyze the phenomenon with greater clarity and it is also applicable to data-collection scenarios other than $B\to VV$. For the $B\to VV$ case, $\lambda_1=\Gamma_L$ and $\lambda_2=\Gamma_H$.

When only time information is available,

$$x\equiv t, \quad g_1(x|\lambda_1)\equiv\lambda_1\exp(-\lambda_1 t),$$

$$g_2(x|\lambda_2)\equiv\lambda_2\exp(-\lambda_2 t). \quad (12)$$

When both time and angular information are available,

$$x\equiv(t,u),$$

$$g_1(x|\lambda_1)\equiv\frac{3}{8}(1+u^2)\lambda_1\exp(-\lambda_1 t), \quad (13)$$

$$g_2(x|\lambda_1)\equiv\frac{3}{4}(1-u^2)\lambda_2\exp(-\lambda_2 t). \quad (14)$$

When only time information is collected, the functions $g_1(\cdot)$ and $g_2(\cdot)$ are the same. When both time and angular information are recorded, they will be different.

The expected information matrix (see the Appendix) for the parameters $\lambda=(\lambda_1,\lambda_2,\beta)$ can be found to be

$$I(\lambda)=\begin{pmatrix} \beta^2\int A^2 d\mu & \beta(1-\beta)\int AB d\mu & \beta\int AC d\mu \\ \beta(1-\beta)\int AB d\mu & (1-\beta)^2\int B^2 d\mu & (1-\beta)\int BC d\mu \\ \beta\int AC d\mu & (1-\beta)\int BC d\mu & \int C^2 d\mu \end{pmatrix} \quad (15)$$

$$=\int [v,v]d\mu \quad (16)$$

[1]If X is a continuous random variable with density function $f(x)$, then the function $F(x)=\int_{-\infty}^{x}f(t)dt=P(X\leqslant x)$ defines the probability integral transform. Then the random variable $F(X)$ is uniformly distributed over the interval [0,1].

where $A = g_1'(x|\lambda_1)$, $B = g_2'(x|\lambda_2)$, $C = g_1(x|\lambda_1) - g_2(x|\lambda_2)$,

$$v = \begin{pmatrix} \beta A \\ (1-\beta)B \\ C \end{pmatrix},$$

$[\cdot,\cdot]$ denotes outer product and $\int d\mu$ denotes integration with respect to the measure $dx/p(x|\lambda_1,\lambda_2,\beta)$.

If we are interested in the difference of the two λ's, then we may be interested in the following derived parametrization of the problem:

$$\theta = \begin{pmatrix} \theta_1 \\ \theta_2 \\ \theta_3 \end{pmatrix} = \begin{pmatrix} \beta\lambda_1 + (1-\beta)\lambda_2 \\ \lambda_1 - \lambda_2 \\ \beta \end{pmatrix}.$$

In that case the information matrix for θ will be

$$I(\theta) = \int [w,w]\, d\mu \tag{17}$$

$$= \begin{pmatrix} \int [\beta A + (1-\beta)B]^2 d\mu & * & * \\ * & [\ 1-\beta\ ]^2 \int \beta(1-\beta)(A-B)^2 d\mu & * \\ * & * & \int \{(\lambda_2-\lambda_1)[\beta A + (1-\beta)B] + C\}^2 d\mu \end{pmatrix}, \tag{18}$$

where

$$w = \begin{pmatrix} \beta A + (1-\beta)B \\ \beta(1-\beta)(A-B) \\ (\lambda_2-\lambda_1)(\beta A + (1-\beta)B) + C \end{pmatrix}.$$

The entries marked with an asterisk (*) are important; they are omitted for the sake of brevity since the qualitative features of the information matrix are clear without them. Careful inspection of the entries in the above matrix in Eq. (18) reveals some qualitative features of the dependence of the information content in the data on the parameter values.

The information on $\theta_2 = \lambda_1 - \lambda_2$ is low when $A \simeq B$ or if $\beta(1-\beta) \simeq 0$. If $g_1 \equiv g_2$ and $\lambda_1 \simeq \lambda_2$, then $A \simeq B$. This is what happens when we have only time data and the two Γ's are close to each other. If $\lambda_1 \simeq \lambda_2$, but $g_1 \not\equiv g_2$, then this problem does not occur. In fact, if g_1 and g_2 are very different functions, then even if $\lambda_1 \simeq \lambda_2$, we can recover information about $\lambda_1 - \lambda_2$ from the data. When both time *and* angle data are collected, the component densities as in Eqs. (13) and (14) are well separated and there is a lot more information on $\Gamma_H - \Gamma_L$ than what would have been with time data alone. If most of the observations are from one component of the mixture, the information on $\lambda_1 - \lambda_2$ is small, since $\beta(1-\beta) \simeq 0$.

The estimation of $\beta\Gamma_L + (1-\beta)\Gamma_H$, on the other hand, is not affected much by the separation of densities since it is a ''mean'' parameter, as shown in Sec. III A.

As mentioned in the Appendix, the inverse of the expected information matrix also gives the approximate variance matrix of the maximum likelihood estimates in large samples. If $I(\theta)$ is the expected information matrix from a single sample, $nI(\theta)$ is the expected information matrix based on a sample of size n. Hence the diagonal elements of the matrix

$$\frac{V(\theta)}{n} = \frac{1}{n}[I(\theta)]^{-1} \tag{19}$$

will give us the approximate variance, $V(\hat{\theta}_i)$, of the maximum likelihood estimates, $\hat{\theta}_i$ $(i=1,2,3)$, in samples of size n when n is large.

Let $\hat{\theta}_i(t)$ denote the maximum likelihood estimate of θ_i given time data alone and let $\hat{\theta}_i(t,u)$ denote the estimate using both time and angular information. By calculating the inverses of the corresponding expected information matrices, one can calculate the ratios

$$\frac{V(\hat{\theta}_i(t))}{V(\hat{\theta}_i(t,u))},$$

for $i=1,2,3$. Figures 2–4 show the plots of these ratios for various values of β and Γ_H/Γ_L.

Whereas most of the information about θ_1 is indeed in the time measurements as expected, it can be seen that the variance of the parameters $\lambda_1 - \lambda_2$ and β is orders of magnitude higher if we have only time information than if we had both time and angle information. The physical region of interest lies around $0.8 < \beta$ and $0.8 < \Gamma_H/\Gamma_L$, where the disparity in

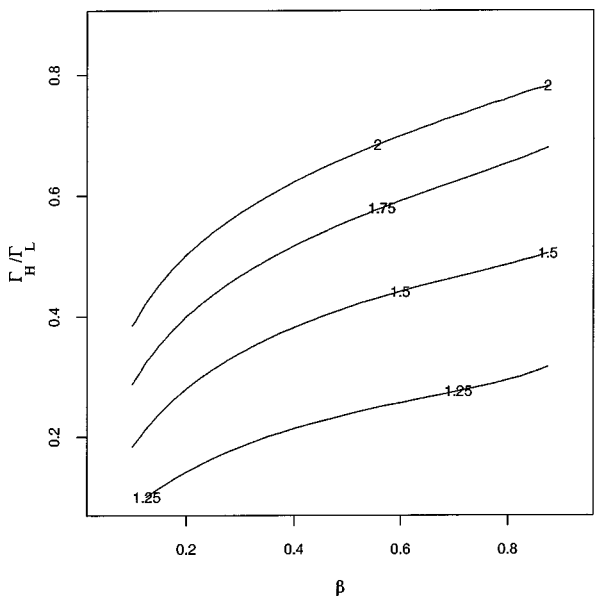

FIG. 2. The ratio of the variances, $V(\hat{\theta}_i(t))/V(\hat{\theta}_i(u,t))$, of the estimates of $\theta_1=\beta\Gamma_L+(1-\beta)\Gamma_H$.

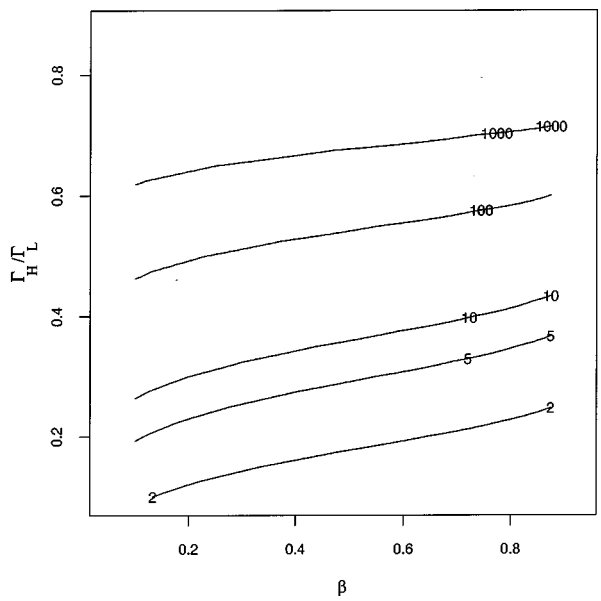

FIG. 4. The ratio of the variances, $V(\hat{\theta}_i(t))/V(\hat{\theta}_i(u,t))$, of the estimates of $\theta_3=\beta$.

the two values of variances is striking. The width of confidence intervals is proportional to the standard error which is the square root of the variance of the estimator. Since the variance of maximum likelihood estimates and the moment estimates are inversely proportional to the number of data points, the ratio of the number of data points needed to have confidence intervals of a given length, with time data alone instead of time and angle data, is equal to the levels of the contours in Figs. 2–4. Looking at the upper right hand corner of the plot in Fig. 3, we can see that with the inclusion of angular information, the sample sizes required for estimation of $\Gamma_L-\Gamma_H$ to a desired level of accuracy will be smaller by a factor of at least 10 times or even more than 100 times compared to those required with the time information alone.

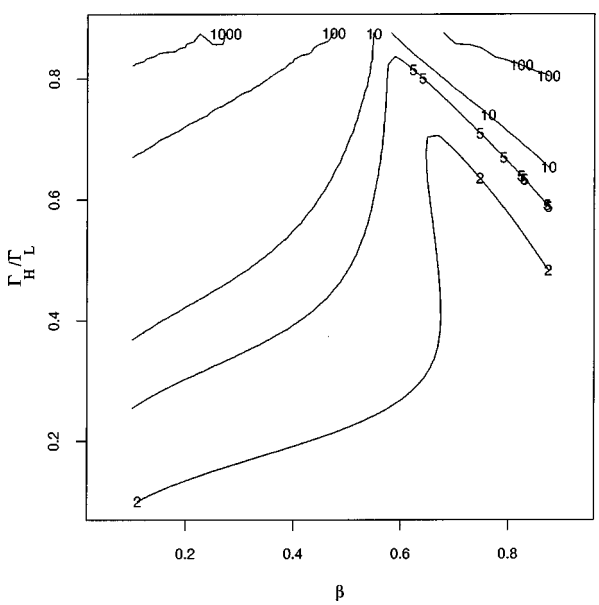

FIG. 3. The ratio of the variances, $V(\hat{\theta}_i(t))/V(\hat{\theta}_i(u,t))$, of the estimates of $\theta_2=\Gamma_L-\Gamma_H$.

Analysis of the data using the angular information available is, therefore, highly recommended.

IV. METHOD OF ANGULAR MOMENTS FOR EXTRACTING INFORMATION

Because of the optimality properties that the maximum likelihood estimates enjoy in large samples, the method of maximum likelihood is widely used. The likelihood function, indeed, contains all the information available on the data, and is the most efficient method for summarizing the information on the data when the form of the probabilistic model for the data is known [11]. However, there are some practical limitations to the likelihood method. When the number of parameters is large, exploring the likelihood surface is problematic. Finding the maximum is also difficult. In addition, if proper care is not taken, misleading results may be obtained (see, for example, [12], chapter on ''Non-Linear Statistical Methods''), and when there are random errors in the measurement process, the likelihood function may not be computable (see Sec. IV C).

We therefore propose the method of angular moments, which sacrifices some information (as compared to the likelihood function), but can give consistent and reliable estimates of the parameters in a clear way.

In the following section, we will show that in the case of one angle distribution at least, the angular moments method is almost as efficient as the maximum likelihood method. In Secs. IV B and IV C we will discuss the effects of imperfections in the measurement process on both the method of angular moments and the likelihood method.

A. Efficiency of the method of angular moments

The method of angular moments is described in [3]. It involves finding a set of weighting functions $w_i(\underline{u})$ such that, given an angular distribution $\Sigma_i b_i f_i(\underline{u})$, where $\underline{u}$ is the vector of angular variables,

$$E(w_i(\underline{u})) = b_i, \tag{20}$$

where E stands for the expectation value. Such a set of weighting functions always exists [3] if the angular distribution is of the form mentioned above. An estimate of b_i is then

$$\hat{b}_i = \frac{1}{n} \sum_{j=1}^{n} w_i(\underline{u}_j). \tag{21}$$

The estimate is unbiased, i.e. $E(\hat{b}_i) = b_i$, and its standard error is equal to $\sigma_i / \sqrt{n}$, where

$$\sigma_i^2 = \int [w_i(\underline{u}) - b_i]^2 f(\underline{u}) d\underline{u} = \left(\int [w_i(\underline{u})]^2 f(\underline{u}) d\underline{u} \right) - \left(\int w_i(\underline{u}) f(\underline{u}) d\underline{u} \right)^2. \tag{22}$$

The information in a sample of size n, on b_i, can be measured by the inverse of the variance of $\hat{b}_i$ and is equal to n/σ_i^2. The information *per observation* is then $1/\sigma_i^2$. To compare the angular moments method with the likelihood method we will compare the information per observation with that of the likelihood method, as defined in the previous section.

Let us take the example of the transversity angle distribution without any time information. The density is given by Eq. (10). We shall see that in this case the method of angular moments performs almost as well as the maximum likelihood estimate in the region of interest to us.

The information content in the maximal likelihood method is as given in Eq. (A1). For the method of angular moments, the density of $u = \cos(\theta)$ is

$$p(u) = (3/8)[\beta(3u^2 - 1) + 2(1 - u^2)].$$

The weighting function for β may be chosen to be $w(u) = 5u^2 - 1$, so that $E(w) = \beta$ and $E(w^2) = (24/7)\beta + (8/7)(1 - \beta)$. The ratio

$$I(\beta|u)_{\mathrm{AM}} / I(\beta|u)_{\mathrm{ML}}$$

(where AM represents the method of angular moments and ML represents the maximal likelihood method) is plotted in Fig. 5.

The plot shows that for $\beta > 0.3$, the ratio of variances is more than 0.9. The expected value of β $(0.8 < \beta < 1.0)$ is well within this range. Thus, in the physical region of interest, the method of angular moments seems to perform almost as well as the maximal likelihood fit.

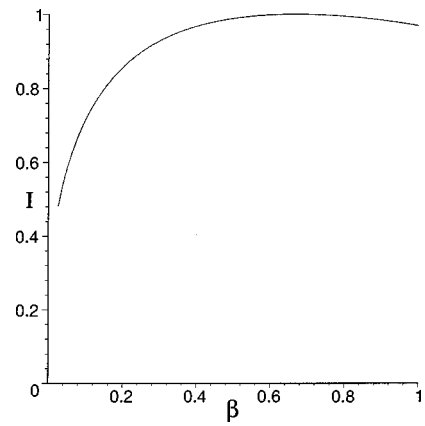

FIG. 5. The ratio of information content about β extracted through the angular moments method and the maximal likelihood method. The X axis is the actual value of β.

When we move to a higher number of parameters, the maximum likelihood method will try to maximize the multidimensional likelihood function and the complexity of the method increases rapidly with the number of dimensions. The dimension of the parameter space does not affect the implementation of angular moment method. Therefore it is useful, at the least, as a method for providing good initial estimates. Additionally, if it is as efficient compared to the full likelihood method as the one-angle case suggests, it could render the maximum likelihood method unnecessary.

B. Effect of measurement discretization

So far, we have assumed that the data are measured to the maximum precision available. In practice, that is not the case and the data are usually reported as the midpoint of the interval in which the measurement actually fell. For example, if we are measuring a random variable X to a precision h, that means that all observations falling in the interval $(x^* - h/2, x^* + h/2]$ are reported as x^*. The resulting discretization of measurements can lead to a systematic bias in measurements, because instead of recording the random variable X, we are recording the *derived* random variable X^* with the probability distribution given by

$$\mathrm{P}(X^* = x^*) = \int_{x^* - h/2}^{x^* + h/2} f(x) dx.$$

Let us compare the difference the means of the true and derived random variable, i.e., $E(X)$ and $E(X^*)$. It suffices to compare terms of the form

$$\int_{x^* - h/2}^{x^* + h/2} x f(x) dx \quad \text{and} \quad x^* \int_{x^* - h/2}^{x^* + h/2} f(x) dx.$$

Now,

$$\int_{x^*-h/2}^{x^*+h/2} t f(t)\,dt - x^* \int_{x^*-h/2}^{x^*+h/2} f(t)\,dt$$

$$= \int_{x^*-h/2}^{x^*+h/2} (t-x^*) f(t)\,dt$$

$$= \int_{x^*-h/2}^{x^*+h/2} (t-x^*)[f(x^*) + (t-x^*) f'(x^*) + O((t-x^*)^2)]\,dt$$

$$\simeq \int_{-h/2}^{+h/2} [t f(x^*) + t^2 f'(x^*)]\,dt = f'(x^*)\frac{h^3}{6}.$$

Thus, the error in discretization is of the order of the cube of the length of the interval length of discretization if the density is ''well behaved.'' If the (effective) support of the distribution of X is $[a,b]$, and the precision is h, then using a crude bound we would get

$$|E(X) - E(X^*)| \leq \frac{(b-a)}{h}\frac{h^3}{6}\left(\sup_{x\in[a,b]} |f'(x)|\right)$$

$$= (b-a)\frac{h^2}{6}\left(\sup_{x\in[a,b]} |f'(x)|\right). \qquad (23)$$

The bound has obvious modifications when the discretization is done over intervals of varying length.

Thus the bias due to discretization is of the order of the square of the bin widths. If the bin widths are sufficiently small, neither the moment method nor the likelihood method will be affected significantly.

C. Effect of random error in measurements

Another source of error in measurements comes from errors in the measuring instruments. Suppose the true variable we want to measure is X, but instead, as a result of random error, we measure

$$Y = X + E,$$

where the error distribution has density $f_E(\cdot)$ and is independent of X. It is reasonable to assume that the random error has mean 0. Suppose its variance is σ_E^2. Then, $E(Y) = E(X)$, but $V(Y) = V(X) + \sigma_E^2$. In other words, the mean of our measurements is unchanged by the random error, but the variance is increased. The implication is that the method of angular moments is unaffected by random error as far as the estimation goes. However, the standard errors of the estimates will be increased.

The effect of random error on the likelihood method is more serious, because it relies on the *exact* mathematical form of the density of the observed measurements. For example, when we are measuring time and the transversity angle, the density of the data will no longer be Eq. (6) but

$$p(u,t|\beta,\Gamma_H,\Gamma_L) = \int\int \left(\frac{3}{8}\beta\Gamma_L(1+u_*^2)e^{-\Gamma_L t_*} + \frac{3}{4}(1-\beta)\Gamma_H(1-u_*^2)e^{-\Gamma_H t_*}\right) \times f_E(u-u_*, t-t_*)\,du_*\,dt_*. \qquad (24)$$

In general, the likelihood function in the presence of random noise is in the form of an integral with respect to the noise terms which may be analytically intractable unless the distribution of the noise term is known and is in a simple form.

If the noise term is believed to be significant, then it may be better to use the method of moments because it is robust to the presence of additional random noise.

V. THREE-ANGLE DISTRIBUTION

Given the enormous additional amount of information available in the angular data θ as compared to the time data alone, we expect that the information embedded in the measurements of the two additional physical angles ψ and φ would be useful in reducing the uncertainty on the parameters which can in principle be measured by the time and transversity angle data. Moreover, the additional terms available for measurement in the three angle case (see Table I) allow us the access to additional parameters. The quantities $A_\parallel/A_0$, $A_\perp/A_0$ (both magnitudes and phases), Δm and the CP-violating parameter $\delta\phi$ need the measurement of these two extra angles. The CP asymmetry

$$(e^{-\Gamma_H t} - e^{-\Gamma_L t})\cos(\delta_i)\,\delta\phi \qquad (25)$$

can be measured even without tagging (without knowing whether the initial particle was a B_s or $\overline{B_s}$) as long as we have this information. Using all the angular data, therefore, is highly recommended. Here, we perform some Monte Carlo simulations to estimate how well the above parameters will be known in the next few years.

At the end of Collider Detector at Fermilab (CDF) run II (expected integrated luminosity $\approx 2\ \mathrm{fb}^{-1}$), we should have around 9000 fully reconstructed $B_s \to J/\psi(\to l^+l^-)\phi(\to K^+K^-)$ events [13], whereas this number is expected to increase by a factor of at least 15 (just due to the integrated luminosity improvement) with TeV33. Sets of 10 000 and 100 000 events were generated, with the accuracy in the measurements of time and angles taken at $\Delta t = 0.1/\Gamma_L$ and $\Delta\theta = 0.005$. The method of angular moments and time moments[2] was used to recalculate all the input parameters (with only the data, without any external information) and histograms were plotted for the recalculated parameters. The simulations use the following set of parameters:

[2]The nth time moment of a quantity $Q(t)$ is defined as $T^{(n)} \equiv \int_0^\infty dt\, t^n Q(t)$. The zeroth time moment is just the time integrated quantity.

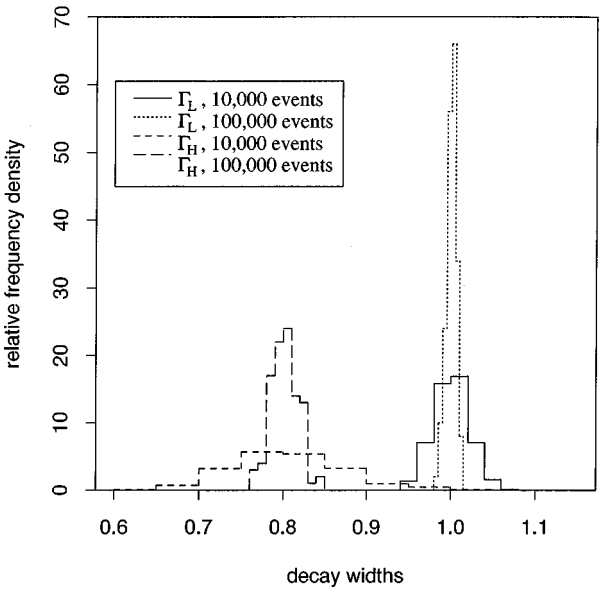

FIG. 6. Determination of Γ_H and Γ_L. The X axis has been normalized to $\Gamma_L(\text{actual}) = 1.0$.

$$\frac{\Gamma_H}{\Gamma_L} = 0.8, \quad \left|\frac{A_\parallel}{A_0}\right| = 0.55, \quad \left|\frac{A_\perp}{A_0}\right| = 0.45,$$

$$\frac{\delta m}{\overline{\Gamma}} = 10.0, \quad \delta_1 = 0.5, \quad \delta_2 = 2.5.$$

This choice of parameters is consistent with the corresponding ones reported in [10] for $B \to J/\psi K^*$ and flavor $SU(3)$ (except for the lifetime difference). It is seen that varying these parameters does not change the essential conclusions.

Figures 6–9 show the results of these simulations. The Y axis has been normalized to get the ''relative frequency density,'' such that the area under each histogram is equal. The following observations should be noted.

(i) As can be seen from Fig. 6, the values of Γ_H and Γ_L are well separated in the first stage (10 000 events) itself. With 100 000 events, the difference $\Gamma_L - \Gamma_H$ can be determined to nearly $\pm 0.05\Gamma_L$ to more than 95% confidence level. By virtue of the central limit theorem, the method of moment estimates is approximately Gaussian in large samples. The visual appearance of the histograms is consistent with this theoretical property. The width of the Gaussian distribution $\Gamma_L - \Gamma_H$ is then (approximately) inversely proportional to $\beta(1-\beta)$ (see Sec. III) and has a weak dependence on the actual value of $\Gamma_L - \Gamma_H$ as long as $\Gamma_L - \Gamma_H$ is small, which is the case here. So the above quantitative inferences from this histogram should stay valid even with a smaller value of $\Gamma_L - \Gamma_H$. Determination of $1 - \Gamma_H/\Gamma_L$ to 0.05 is thus within reach. Even the small lifetime difference predicted recently in [14] may be probed with this.

(ii) The accuracy in the measurements of $|A_\parallel/A_0|$, $|A_\perp/A_0|$ is as indicated in Figs. 7 and 8 respectively. The predictions of form factor models [15] can thus be directly tested here.

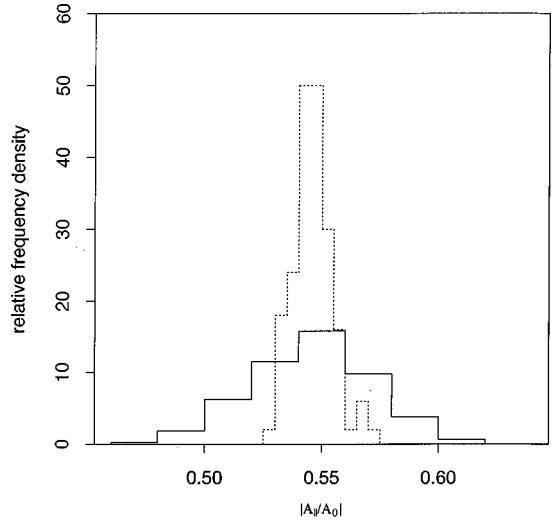

FIG. 7. Determination of $|A_\parallel/A_0|$. The solid line is for 10 000 events and the dashed line is for 100 000 events.

(iii) The signs of $\cos(\delta_1)$ and $\cos(\delta_2)$ are important in order to resolve a discrete ambiguity in the Cabibbo-Kobayashi-Maskawa (CKM) angle β, as pointed out recently [16]. In fact, if $\delta\phi$ is small (≈ 0.03) as predicted by the standard model, these signs may be obtained without any time measurements as follows. With $\delta\phi$ neglected, the time integrated angular moments of the ''Im'' terms in Table I give

$$-|A_X||A_\perp|\cos(\delta_i + \kappa)\overline{\Gamma}/\sin\kappa, \tag{26}$$

where $\kappa = \tan^{-1}(\overline{\Gamma}/\Delta m)$ and $X \in \{0,\parallel\}$. Since $\sin\kappa$ is positive, the sign of these moments immediately gives the sign of $\cos(\delta_i + \kappa)$, and given an upper limit (of ≈ 0.1) on the value of κ, will give the sign of $\cos(\delta_i)$ as long as the value of this moment (26) is not close to zero. Thus, just the sign of the angular moments of the ''Im'' terms in Table I would be sufficient to resolve a discrete ambiguity in β. The relevant

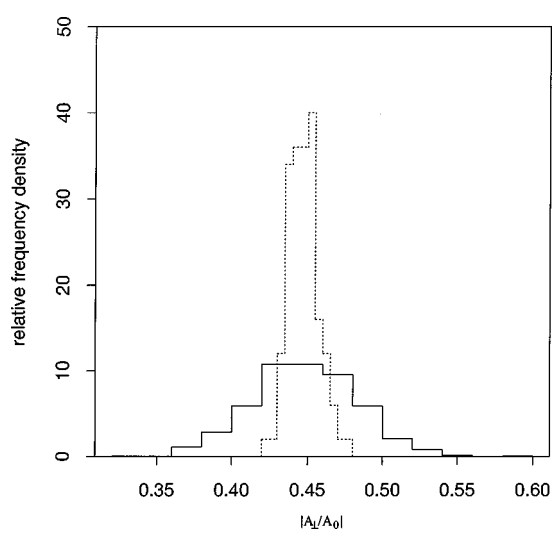

FIG. 8. Determination of $|A_\perp/A_0|$. The solid line is for 10 000 events and the dashed line is for 100 000 events.

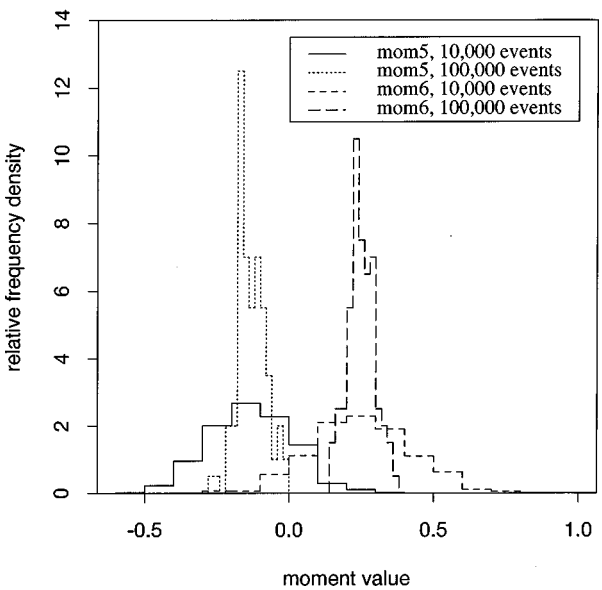

FIG. 9. The moments of the ''Im'' terms in Eq. (26). mom5 is the value of $-|A_{\parallel}||A_{\perp}|\cos(\delta_1+\kappa)\overline{\Gamma}/\sin\kappa$ and mom6 is the value of $-|A_0||A_{\perp}|\cos(\delta_2+\kappa)\overline{\Gamma}/\sin\kappa$.

angular moments (time integrated) are shown in Fig. 9. The widths of these moment histograms depend only weakly on the actual parameter values and the plot can be used as a guide to estimate the errors on the values of these moments for any other parameter values.

(iv) When $\Gamma_H \approx \Gamma_L$,

$$\int_0^{\infty} dt(e^{-\Gamma_H t}-e^{-\Gamma_L t}) \approx (\Gamma_L-\Gamma_H)/\overline{\Gamma}^2. \tag{27}$$

The ability to measure $\Gamma_L-\Gamma_H$, combined with the measurement of the time-integrated CP asymmetry in Eq. (25) (even without tagging) would give a lower bound on $\delta\phi$. A high value of $\delta\phi$ would be a clear signal of physics beyond the standard model. In the next generation of experiments (TeV33 or LHC), accurate values of δ_i $(i=1,2)$ will be obtained and $\delta\phi$ can be pinpointed.

Feasibility studies for the measurement of $\Delta m/\Gamma$ and the asymmetries in this decay mode using the angular moments method and weighting functions have been made in [17] for the CMS detector, which claim that with $L \approx 10 fb^{-1}$, reasonable sensitivity on the oscillations will be obtained at $\Delta m/\Gamma < 40$. The angular moments method has also been used for the analysis of $B^0 \to D^{*-}\rho^+$ and $B^+ \to \overline{D}^{*0}\rho^+$ [18] and the error estimation (Tables III and IV in [18]) indicates that the angular moments method is almost as efficient as the best fit method in estimating the observables, even with the three angle distribution.

VI. SUMMARY AND CONCLUSIONS

Using a ''reasonable'' method for quantifying the information in the data, we have shown that the information content in the data may increase by orders of magnitude in the region of interest if angular information is added to the time information. This is true even if the quantity to be measured, e.g. the lifetime difference between B_s^L and B_s^H, has no direct angular dependence. We have also isolated ''averaged'' quantities for which this increase of information is small, which means that their measurements are not helped much by the angular data.

The actual use of the angular data involves the choice of a statistical method to summarize the data. The standard maximum likelihood method is theoretically the ''best'' when the number of data points is very large. However, when the number of parameters to be estimated is large, the numerical maximization of the likelihood may be difficult, and if proper care is not taken, misleading results may be obtained. Also, if there are random errors in the measurement process, then the likelihood function would be an integral that may not be mathematically known or, if known, not evaluable in a closed form.

The method of angular moments is very straightforward to implement, and the connections to the parameters to be determined are more transparent. It is consistent in the statistical sense that, with infinite data, it will nail the parameters down. Unlike the maximum likelihood method, it is robust under random errors of measurement. In the one angle case at least, as we have shown explicitly, it is almost as efficient as the maximum likelihood method in the region of interest. Both methods are subject to discretization errors which will be small if the interval of discretization is small. We therefore recommend the use of the method of angular moments for extracting information, at least for the initial estimates. If necessary, they can be refined with the likelihood method. Even if the maximum likelihood method is used, optimization routines require consistent starting values which can be provided by the method of angular moments.

We have used the angular moments method on simulated sets of data to estimate the accuracy to which it may determine the quantities of interest. In the case of the decay $B_s \to J/\psi(\to l^+l^-)\phi(\to K^+K^-)$, we find that in the first stage of experiments (CDF II), this method should be sufficient to give reliable values of $\Gamma_L-\Gamma_H$, $|A_{\parallel}/A_0|$ and $|A_{\perp}/A_0|$. This, combined with the untagged CP asymmetry measured through the same decay, would give a lower bound for $\delta\phi$, which is expected to be very small in the standard model. The signs of $\cos(\delta_1)$ and $\cos(\delta_2)$, which are useful in resolving a discrete ambiguity in the CKM angle β, can be determined in the next stage (TeV33), along with more accurate determination of $\delta\phi$, which may point unambiguously to new physics.

ACKNOWLEDGMENTS

We would like to thank the University of Chicago Physical Sciences Division where this work was commenced, and was supported in part by the U.S. Department of Energy under Contract No. DE FG02 90ER40560. A.D. would like to thank I. Dunietz, R. Fleischer, H. Lipkin, and J. Rosner for previous collaborations on angular distributions and for helpful comments on this manuscript, S. Pappas for experimental insights, and K. Uryu for help in numerical computations. S.S. would like to thank J. Liu for supporting this work

through National Science Foundation grants DMS-9101311 DMS-9501570 and P. McCullagh and X.-L. Meng for helpful discussions.

APPENDIX: QUANTIFYING INFORMATION IN AN EXPERIMENT

The notion of ''information'' that we have used in this paper is derived from statistical theory. A good reference is [11]. Suppose an experiment is performed to determine the value of the parameter α. The (average) information in the experiment to discriminate between different possible values of the parameter when the true value of the parameter is α_0 is measured by

$$I(\alpha_0,X) = -\int \ddot{l}(\alpha_0)p(x|\alpha_0)dx, \tag{A1}$$

often called the *expected* Fisher information. Here X is the random variable denoting the data used from the experiment, $p(x|\alpha)$ is the probability of X given the parameter value α, and $l(\alpha)=\log[p(x|\alpha)]$ is the logarithmic likelihood function. Note the dependence of the expected information in the experiment on the true value of the parameter, α_0, and on the data used, X. Both α and X may be vector valued. This measure of information possesses the additivity property; i.e., if X_1 and X_2 denote data from two independent experiments about the same parameters, then

$$\mathrm{I}(\alpha_0,(X_1,X_2)) = \mathrm{I}(\alpha_0,X_1) + \mathrm{I}(\alpha_0,X_2).$$

In particular this implies that if n independent and identically distributed data points, $X_1,X_2,\ldots,X_n$, are collected from an experiment, the expected information in the whole experiment is n times the expected information in one observation:

$$\mathrm{I}(\alpha_0,(X_1,X_2,\ldots,X_n)) = n\mathrm{I}(\alpha_0,X_1).$$

Additionally, it can be shown that for any estimator of α, say $\hat{\alpha}_n$, based on a sample $X_1,X_2,\ldots,X_n$ (the Cramér-Rao inequality),

$$V(\hat{\alpha}_n) \geqslant \frac{1}{\mathrm{I}(\alpha_0,(X_1,X_2,\ldots,X_n))} = \frac{n}{\mathrm{I}(\alpha_0,X_1)}, \tag{A2}$$

where V is the variance and $\hat{\alpha}_n$ is based on a sample of size n. When the sample size is large and certain regularity conditions hold, the lower bound in the variance is achieved by the maximum likelihood estimate. It is in this sense that the maximum likelihood estimate is the ''best.''

It can also be shown that when we have independent and identically distributed data points, for large samples,

$$l(\alpha) \simeq l(\hat{\alpha}) + (\alpha-\hat{\alpha})\dot{l}(\hat{\alpha}) + (\alpha-\hat{\alpha})^2\ddot{l}(\hat{\alpha}) \tag{A3}$$

where $\hat{\alpha}$ denotes the maximum likelihood estimate of α. By construction, $\dot{l}(\hat{\alpha})=0$ and hence

$$l(\alpha) \simeq l(\hat{\alpha}) + (\alpha-\hat{\alpha})^2\ddot{l}(\hat{\alpha}). \tag{A4}$$

In other words, the logarithmic likelihood surface is approximately quadratic and its shape can be described by the position of the maximum ($\hat{\alpha}$) and the curvature of the logarithmic likelihood in the neighborhood of the maximum [$\ddot{l}(\hat{\alpha})$]. The latter describes how fast the logarithmic likelihood falls off; the larger the value of $\ddot{l}(\hat{\alpha})$, the steeper the fall and stronger is the evidence in favor of values near the maximum. For this reason, $\ddot{l}(\hat{\alpha})$ is also used as a measure of information, but since it varies from sample to sample, it is called the *observed* Fisher information. Its average value is the *expected* Fisher information mentioned above.

While we have defined the information for a scalar parameter, the general idea can be extended to vector-valued parameters. When there are two or more parameters, the appropriate measure of expected information is the expected information matrix which is the expected value of the Hessian of the logarithmic likelihood as in Eq. (A1). See [11] for details and additional references.

[1] G. Valencia, Phys. Rev. D **39**, 3339 (1989); G. Kramer and W. F. Palmer, *ibid.* **45**, 193 (1992); Phys. Lett. B **279**, 181 (1992); Phys. Rev. D **46**, 2969 (1992); **46**, 3197 (1992); G. Kramer, W. F. Palmer and H. Simma, Nucl. Phys. **B428**, 77 (1994); G. Kramer, T. Mannel and W. F. Palmer, Z. Phys. C **55**, 497 (1992).

[2] I. Dunietz, H. R. Quinn, A. Snyder, W. Toki and H. J. Lipkin, Phys. Rev. D **43**, 2193 (1991).

[3] A. S. Dighe, I. Dunietz and R. Fleischer, Report No. CERN-TH/98-85, IC/98/25, hep-ph/9804253.

[4] J. S. Hagelin, Nucl. Phys. **B193**, 123 (1981); E. Franco, M. Lusignoli and A. Pugliese, *ibid.* **B194**, 403 (1982); L. L. Chau, W.-Y. Keung and M. D. Tran, Phys. Rev. D **27**, 2145 (1983); L. L. Chau, Phys. Rep. **95**, 1 (1983); A. J. Buras, W. Slominski and H. Steger, Nucl. Phys. **B245**, 369 (1984); M. B. Voloshin, N. G. Uraltsev, V. A. Khoze and M. A. Shifman, Yad. Fiz. **46**, 181 (1987) [Sov. J. Nucl. Phys. **46**, 112 (1987)]; A. Datta, E. A. Paschos and U. Türke, Phys. Lett. B **196**, 382 (1987); A. Datta, E. A. Paschos and Y. L. Wu, Nucl. Phys. **B311**, 35 (1988); M. Lusignoli, Z. Phys. C **41**, 645 (1989); R. Aleksan, A. Le Yaouanc, L. Oliver, O. Pène and Y.-C. Raynal, Phys. Lett. B **316**, 567 (1993); I. Bigi, B. Blok, M. Shifman, N. Uraltsev and A. Vainshtein, in *B Decays*, 2nd ed., edited by S. Stone (World Scientific, Singapore, 1994), p. 132 and references therein; M. Beneke, G. Buchalla and I. Dunietz, Phys. Rev. D **54**, 4419 (1996).

[5] R. Fleischer and I. Dunietz, Phys. Lett. B **387**, 361 (1996); Phys. Rev. D **55**, 259 (1997).

[6] J. L. Rosner, Phys. Rev. D **42**, 3732 (1990).
[7] A. S. Dighe, I. Dunietz, H. J. Lipkin and J. L. Rosner, Phys. Lett. B **369**, 144 (1996).
[8] R. Aleksan, B. Kayser and D. London, Phys. Rev. Lett. **73**, 18 (1994).
[9] CDF Collaboration, F. Abe *et al.*, Phys. Rev. Lett. **75**, 3068 (1995).
[10] CLEO Collaboration, Report No. CLEO Conf. 96-24, ICHEP96 PA05-074.
[11] D. R. Cox and D. V. Hinkley, *Theoretical Statistics* (Chapman and Hall, London, 1974).
[12] R. A. Thisted, *Elements of Statistical Computing* (Chapman and Hall, New York, 1988).
[13] The CDF II Detector Technical Design Report, edited by the CDF Collaboration and D. Amidei, Report No. FERMILAB-Pub-96/390-E, 1996.
[14] M. Beneke *et al.*, hep-ph/9808385.
[15] M. Bauer, B. Stech and M. Wirbel, Z. Phys. C **29**, 637 (1985); **34**, 103 (1987); J. M. Soares, Phys. Rev. D **53**, 241 (1996); H. Y. Cheng, Z. Phys. C **69**, 647 (1996).
[16] A. S. Dighe, I. Dunietz and R. Fleischer, Phys. Lett. B **433**, 147 (1998).
[17] P. Galumian (in preparation).
[18] CLEO Collaboration, Report No. CLEO Conf. 98-23, ICHEP98 852.