Mathematical Methods for $B^0\bar{B}^0$ Oscillation Analyses

H.-G. Moser

Max-Planck-Institut für Physik, Werner-Heisenberg-Institut, Föhringer Ring 6, Munich, Germany

A. Roussarie

Service de Physique des Particules, DAPNIA, CE-Saclay, 91191 Gif-sur-Yvette cedex, France

The measurement of the $B_s^0\bar{B}_s^0$ mixing frequency Δm_s requires the search for a periodic pattern in the time distribution of the data. Using Fourier analysis the consequences of vertex and boost resolution, mistag and statistical fluctuations are treated analytically and a general expression to estimate the significance of a $B^0\bar{B}^0$ mixing analysis is derived. With the help of Fourier analysis the behaviour of a classical maximum likelihood analysis in time space is studied, too. It can be shown that a naive maximum likelihood fit fails in general to give correct confidence levels. This is especially important if limits are calculated. Alternative methods, based on the likelihood, which give correct limits are discussed. A new method, the amplitude fit, is introduced which combines the advantages of a Fourier analysis with the power and simplicity of a maximum likelihood fit.

Preprint submitted to Elsevier Preprint *21 June 1996*

1 Introduction

$B^0_d \bar{B}^0_d$ mixing is by now a well established phenomena, the time dependence has been measured by several experiments [1]. However the probably very fast oscillations of the B^0_s have not been seen to date, only lower limits on the oscillation frequency exist [2–8]. It turns out that the usual way to analyse the data by looking directly at the time distribution of the B^0 decays is very limited. Effects of resolution and statistics can only be estimated very indirectly, even the way how to establish correct limits is still under discussion [9]. On the other hand, the method of Fourier analysis offers several advantages and the problems mentioned above can be handled easily [10].

In this paper it is shown how Fourier analysis can be used as a tool to understand the analysis of $B^0 \bar{B}^0$ oscillations. In section 2 a short summary of the phenomenology of $B^0 \bar{B}^0$ is given. Then, in section 3, the basic formulas of the Fourier transformation are explained. In the following sections Fourier analysis is used to understand the effects of resolution (section 4) and of statistical fluctuations (section 5). This knowledge can be used to obtain general formulas for the mixing analysis, which are given in section 6.

All analyses published to date are based on maximum likelihood fits of the oscillation frequency in time space. As discussed in section 6.1 special attention is given to derive a method to calculate lower limits for the oscillation frequency. Various options of the likelihood method are discussed in section 7, using Fourier analysis to understand their properties. It can be shown that limits obtained from such fits are not necessarily correct, unless special precautions are made. A new method, the amplitude fit, which combines the advantages of maximum likelihood fits and Fourier analysis is introduced in section 8.

The paper finishes with a comparison of the different methods and the conclusion.

2 Phenomenology

The phenomenology of $B^0 \bar{B}^0$ mixing has been discussed in many papers [11]. In short the probability to observe a $\bar{B}^0$ if a B^0 was produced is:

$$P(t)^{mix} = \frac{1}{2} \exp(-\Gamma t)[1 - \cos(\omega t)], \qquad (1)$$

and a B^0:

$$P(t)^{unmix} = \frac{1}{2}\exp(-\Gamma t)[1+\cos(\omega t)], \quad (2)$$

Γ is the B^0 decay width, ω the oscillation frequency, which equals the mass difference Δm.

The oscillation frequency of the B_d^0 has been measured to be $\omega \approx 0.5$ ps^{-1}. For the B_s^0 frequencies in the range of about $6-55$ ps^{-1} are expected in the Standard Model [12].

Experimentally the $B^0/\bar{B}^0$ state has to be tagged at production and decay time as a function of the decay time. The tagging can be done using leptons, jet charge, reconstructed D_s, fragmentation kaons, etc.. Usually samples of 'like' sign events, enriched with $B^0 \to \bar{B}^0$ transitions and 'unlike' sign events, enriched with $B^0 \to B^0$ events, are obtained. The samples are not pure, because of mistag (wrong charge determination) and backgrounds. E.g., if f_s is the fraction of B_s^0 events in the sample, η the mistag probability, and assuming the background to be due to other B which don't mix, the following is observed:

$$P^{like}(t) = \left[f_s\frac{1}{2}[(1-\eta)(1-\cos(\omega t))+\eta(1+\cos(\omega t))]+(1-f_s)\eta\right]\Gamma\exp(-\Gamma t), \quad (3)$$

$$P^{unlike}(t) = \left[f_s\frac{1}{2}[(1-\eta)(1+\cos(\omega t))+\eta(1-\cos(\omega t))]+(1-f_s)(1-\eta)\right]\Gamma\exp(-\Gamma t), \quad (4)$$

The analysis function could be the difference of the 'like' and 'unlike' samples

$$P^{unlike}(t) - P^{like}(t) = [f_s(1-2\eta)\cos(\omega t)+(1-f_s)(1-2\eta)]\Gamma\exp(-\Gamma t). \quad (5)$$

In reality other kinds of background with different decay constants have to be added and the experimental resolution of the proper time t has to be taken into account.

3 Fourier Analysis

A very powerful mathematical method to analyse periodic signals is the Fourier analysis: The Fourier transformation (FT) of a function f(t) is:

$$FT[f(t)](\nu) = g(\nu) = \frac{2}{\sqrt{\pi}}\int_0^{+\infty} f(t)\exp(-i\nu t)\,dt. \quad (6)$$

The FT of the functions needed in the analysis of $B\bar{B}$ oscillations are:

$f(t)$	REAL$[g(\nu)]$
$\cos(\omega t)$	$\sqrt{\frac{2}{\pi}}\delta(\nu - \lvert\omega\rvert)$
$\Gamma\exp(-\Gamma t)$	$\sqrt{\frac{2}{\pi}}\frac{\Gamma^2}{\Gamma^2+\nu^2}$
$\Gamma\exp(-\Gamma t)\cos(\omega t)$	$\frac{1}{\sqrt{2\pi}}\left(\frac{\Gamma^2}{\Gamma^2+(\omega-\nu)^2}+\frac{\Gamma^2}{\Gamma^2+(\omega+\nu)^2}\right)$
$\frac{1}{\sqrt{2\pi\sigma}}\exp\left(-\frac{t^2}{2\sigma^2}\right)$	$\frac{1}{\sqrt{2\pi}}\exp\left(-\frac{\nu^2\sigma^2}{2}\right)$.

(7)

The real part of the Fourier transformation of a cosine shows a peak at the oscillation frequency $\nu = \omega$.
For practical calculations the Fast Fourier Transformation algorithm (FFT) is used, which can be applied to discrete (=binned) distributions $f_k = f(t_k)$, $k = 0, N-1$, where N has to be a power of 2, over the time interval T, $t_k = kT/N$:

$$FFT(f_k)_j = g_j = \frac{1}{N}\sum_{k=0}^{N-1} f_k \exp\left(-2\pi i\frac{jk}{N}\right). \tag{8}$$

g_j is essentially given by $g(\nu_j)$ with $\nu_j = 2\pi j/T$, $j = 0, N-1$.
In many cases the FT or FFT can be calculated analytically. Important systematic effects can be studied easily, e.g. the effect of resolution or statistical fluctuations. Therefore the analysis can be optimised without need for lengthy MC studies. Errors and confidence levels can be calculated in a straightforward way.

4 Resolution effects: convolution theorem

The effect of the lifetime resolution results in the convolution of the time distribution $P(t)$ with a resolution function $g(t-t',\sigma)$:

$$g(t-t',\sigma) \;=\; \frac{1}{\sqrt{2\pi}\sigma}\exp\left(-\frac{(t-t')^2}{2\sigma^2}\right), \tag{9}$$

$$P(t) \;\rightarrow\; (P\otimes g)(t) \;=\; \int_{-\infty}^{+\infty} P(t')g(t-t',\sigma)dt'. \tag{10}$$

In a first step the resolution is assumed to be constant with time, which is the case for the decay-length resolution. Then the convolution theorem can

be applied which states that the FT of the convolution of two function equals the product of the FT of these functions:

$$FT[(P \otimes g)(t)](\nu) \;=\; \sqrt{2\pi}\; FT[P(t)](\nu) \times FT[g(t)](\nu). \qquad (11)$$

The Fourier transformation of a Gaussian is:

$$FT[g(t,\sigma_l)](\nu) \;=\; \frac{1}{\sqrt{2\pi}} \exp\left(-\frac{\sigma_l^2}{2}\nu^2\right). \qquad (12)$$

Hence the amplitude of high frequencies is reduced by a damping factor.

The boost resolution σ_p adds a time-dependent term to the resolution function, $\sigma = \sqrt{\sigma_l^2 + \sigma_p^2 t^2}$. The calculation becomes more complicated, first the effect of the boost resolution alone is considered:

$$g(t-t',\sigma_p) \;=\; \frac{1}{\sqrt{2\pi}\sigma_p t'} \exp\left(-\frac{(t-t')^2}{2(\sigma_p t')^2}\right), \qquad (13)$$

then:

$$FT\left(\int_{-\infty}^{+\infty} P(t')g(t-t',\sigma_p)dt'\right) \;=\; \int_{-\infty}^{+\infty} P(t) \exp\left(-\frac{(\nu\sigma_p t)^2}{2}\right) e^{-i\nu t} dt. \qquad (14)$$

In the case where $P(t) = \cos(\omega t)$ the FT can be solved in a good approximation:

$$FT(\nu) \;\approx\; \frac{1}{\sigma_p \nu} \exp\left(-\frac{(\nu-\omega)^2}{2(\sigma_p\nu)^2}\right). \qquad (15)$$

Due to the boost resolution the peak amplitude is reduced by $1/(\sigma_p\omega)$, the width of the peak increases like $\sigma_p\omega$. In case of an exponentially damped cosine: $P(t) = \Gamma\exp(-\Gamma t)\cos(\omega t)$, (15) has to be convoluted with the Breit-Wigner $\sqrt{\frac{2}{\pi}}\Gamma^2/(\Gamma^2+\nu^2)$. If $\sigma_p\omega < \Gamma$ essentially the Breit-Wigner is obtained, for $\sigma_p\omega > \Gamma$ the resolution becomes important and the result approaches (15). The peak amplitude is reduced by a factor:

$$r(\omega,\sigma_p) \;\approx\; \sqrt{\pi} Y \exp(Y^2)\mathrm{ERFC}(Y), \qquad (16)$$

with $Y = \frac{1}{\sqrt{2}}\frac{\Gamma}{\sigma_p\omega}$, $\mathrm{ERFC}(x) = \frac{2}{\sqrt{\pi}}\int_x^\infty e^{-t^2}dt$.

Putting together the boost and decay length resolution, the damping factor (12) applies in addition to the terms of the boost resolution giving a global damping factor:

$$D(\omega, \sigma_l, \sigma_p) = \exp\left(-\frac{\sigma_l^2}{2}\omega^2\right) r(\omega, \sigma_p). \tag{17}$$

A comparison of the numerically calculated FFT and the analytical calculation is shown in Figure 1.

5 Statistical fluctuations

Statistical fluctuations of the sample generate noise in the Fourier transformation. The measurement is successful only if an observed signal is significantly larger than the noise. In order to calculate the noise, the *rms* of the Fourier transformation is needed, given by the power spectrum. It can easily be calculated using the Wiener-Kinchin theorem. The detailed calculations can be found in the Appendix 12.2. Here the results are summarised:
For a discrete Fourier transformation like in (8): $g_j = FFT(f_i)_j$, where the f_i have statistical fluctuations due to the finite number of events in bin i, n_i, $\sigma(f_i) = \sqrt{n_i}$, the real part of the Fourier transformation has a sigma of

$$\sigma[REAL(g_j)] = \frac{1}{N}\sqrt{\frac{n}{2}}, \tag{18}$$

where N is the number of bins and $n = \sum_i n_i$ the total number of events in the sample. This noise is equal for all bins j. In general:

- Given a distribution of any shape the noise spectrum due to statistical fluctuations is flat (= white noise) and the rms amplitude of the real part of the FT is $\frac{\sqrt{n}}{\sqrt{2}N}$.
- The signal is proportional to $\frac{n}{N}$. Hence the signal/noise ratio behaves like $\frac{1}{\sqrt{n}}$ and does not depend on the binning (in the following the normalisation constant $1/N$ will be dropped).
- The noise fluctuations in the frequency spectrum are correlated. If the basic time spectrum has a time constant $\tau = 1/\Gamma$ the correlation has a Breit-Wigner shape $\frac{\Gamma^2}{\Gamma^2+\nu^2}$.

Figure 2 shows the FT of a typical oscillation signal. The Wiener-Kinchin theorem describes exactly the distribution of the noise in the FT.

6 Significance of an experiment

The expected FT of a mixing signal as given in (5) can be written as (real part):

$$FT(\nu) \approx \frac{n f_s (1-2\eta)}{2} \left(\frac{\Gamma^2}{\Gamma^2 + (\nu-\omega)^2} + \frac{\Gamma^2}{\Gamma^2 + (\nu+\omega)^2} \right) + n(1-f_s)(1-2\eta)\frac{\Gamma^2}{\Gamma^2+\nu^2}. \quad (19)$$

It can be assumed that the shape of the last term (from the non-mixing B) can be subtracted in the final analysis. This ignores non-b background which is assumed to be already subtracted. The second term of the Breit-Wigner can also be neglected if $\omega > \Gamma$. In addition the damping due to the resolution must be taken into account. Altogether a peak in the FT at $\nu_i = \omega$ is observed with the size

$$FT(\omega) \;=\; \frac{n f_s (1-2\eta)}{2} D(\omega, \sigma_l, \sigma_p), \quad (20)$$

where $D(\omega, \sigma_l, \sigma_p)$ denotes the damping due to the resolution, as defined in formula (17). For a real experiment with a finite data sample, (20) describes the expected peak size of a signal. On top of that the fluctuations due to the white noise have to be added. For them all events, including the background, count, while resolution and mistag do not matter. The rms of the FT is, independent of the frequency:

$$\sigma(FT) = \sqrt{\frac{n}{2}}. \quad (21)$$

Therefore the expected signal to noise ratio is:

$$S/N = \sqrt{\frac{n}{2}} f_s (1-2\eta) D(\omega, \sigma_l, \sigma_p). \quad (22)$$

This formula can be used to calculate the typical sensitivity of an experiment. Some examples are shown in Figure 3 using experimental parameters of Table 1. The example a) shows the (quite unrealistic) case of a single Gaussian vertex resolution with negligible boost resolution. For simplicity these parameters are used as reference in the Monte Carlo studies of this paper. A more realistic resolution is used in example b) including tails and boost resolution. These parameters correspond to a dilepton analysis like [2]. The example c) corresponds to an exclusive analysis with reconstructed D_s-lepton events and

a jet-charge tag [4]. Despite the low statistics and the slightly inferior mistag such an analysis is still competitive because of the high purity, and, due to the better resolution, becomes superior at high frequencies.

In conclusion, using Fourier analysis the analysis of B_s^0 oscillations resembles the search for a resonance. Instead of searching for a peak in a mass spectrum, a peak in the frequency spectrum is searched for. The noise corresponds to fluctuations of the background. In this picture it becomes clear that the physical observable is not Δm_s, but the Fourier amplitude at a certain frequency Δm_s, like in a resonance search, where the observable is the number of events in a certain mass range.

6.1 Obtaining limits

In a real experiment one might observe a significant signal, where the significance can be estimated by comparing the observed peak size with the typical noise given by (21). It should be mentioned that the whole frequency spectrum has to be taken into account. The fluctuations per frequency interval 2Γ are independent, therefore the probability to observe a certain fluctuation increases with the size of the frequency range !

If no signal is observed a limit can be given by comparing the expected signal size given by (20) with the FT of the data, taking into account the fluctuations and as well the systematic uncertainties of the expectation.

The FT of an experiment corresponding to a true oscillation frequency ω gives data points $d(\nu_i)$:

$$\overline{d(\nu_i)} \approx 0, \tag{23}$$

if $|\omega - \nu_i| > 2\Gamma$, and a Breit-Wigner with a peak size

$$\overline{d(\nu_i)} = a(\omega), \tag{24}$$

at $\nu = \omega$. There $a(\omega) = a(\omega, f_s, \eta, \sigma, ..) = FT(\omega)$ describes the expected peak amplitude. As shown in equation (20) it depends on B_s fraction, mistag, resolution etc. In particular, because of the resolution damping, $a(\infty) = 0$. The statistical fluctuations of the FT are Gaussian with zero mean and a frequency independent rms given by (21):

$$\sigma_d = \sqrt{\frac{n}{2}}. \tag{25}$$

Systematic errors give an uncertainty on a. For example in the case of f_s:

$$\sigma_s \;=\; \frac{\partial a}{\partial f_s} \times \sigma(f_s) \;=\; \frac{a}{f_s} \times \sigma(f_s).$$

where the last term of the equation is derived from (20). While the statistical error on the Fourier amplitude is constant, the systematic error vanishes at large frequencies.

To assess the significance of a B_s mixing signal the probability that the expected signal $a(\nu_i)$ fluctuates to the observed data point $d(\nu_i)$ has to be known. The error on the difference $a(\nu_i) - d(\nu_i)$ is $\sqrt{\sigma_s^2 + \sigma_d^2}$. A signal at the oscillation frequency ν_i can be excluded at 95% C.L. if this difference is larger than $1.645 \times \sqrt{\sigma_d^2 + \sigma_s^2}$. Actually the $d(\nu_i)$ spectrum gives a probability for each frequency to show oscillations or not. The highest frequency ν_l where $a(\nu_l)$ can still be excluded can be regarded as the lower limit of the experiment [*]. Therefore the lower limit ν_l on the oscillation frequency is obtained by the smallest ν_l which satisfies:

$$a(\nu_l) \;=\; d(\nu_l) \;+\; 1.645\, \sqrt{\sigma_d^2 + \sigma_s^2}. \qquad (26)$$

If B_s do not mix, the measured amplitude averaged over a large number of experiments is $\overline{d(\nu_i)} = 0$. So the average limit (or sensitivity) of a particular analysis is at a frequency such that:

$$a(\nu_l) \;=\; 1.645\, \sqrt{\sigma_d^2 + \sigma_s^2}. \qquad (27)$$

Of course the limit can be a bit worse, if the $d(\nu_i)$ fluctuate to positive values, or better ('lucky'), if they fluctuate to negative values. The latter case poses a problem. It might happen that a limit is set in a low sensitivity region: if $d(\nu)$ is very negative, this ν value can be excluded although $a(\nu)$ is close to 0. In practise one can apply a method like in [14] to obtain reasonable limits in this case. Anyway, large deviations which are not compatible with the noise fluctuations indicate either a signal, or - especially if they are negative - systematic problems, e.g. due to an incorrect background subtraction. There is also the possibility that disconnected regions exist at frequencies above ν_l which are excluded as well.

[*] Here the meaning of 95% C.L. is that each value below the limit is excluded at, at least, 95% C.L. This means that at most 5% of the experiments with any frequency below the limit would give an amplitude for this frequency less than the measured value.

7 Maximum likelihood methods

7.1 Description

So far only maximum likelihood fits have been used to search for $B^0\bar{B}^0$ oscillations. In a likelihood analysis the rates $N_i^{like,unlike}$ of like and unlike sign events in time bins i have to be compared to $P_i^{like,unlike}(\nu)$, the time distribution probabilities expected for a mixing of given frequency ν (3, 4). The log-likelihood of the observed distributions is defined, for any frequency ν, as:

$$-\ln[L(\nu)] = \sum N_i^{like} \ln[P_i^{like}(\nu)] + \sum N_i^{unlike} \ln[P_i^{unlike}(\nu)]. \qquad (28)$$

This log-likelihood cannot be absolutely normalised. To make quantitative arguments on its values, they have to be compared to a given reference value. Log-likelihood differences are computed. Two methods have been used. In the first one, the log-likelihood is referenced to its value for an infinite oscillation frequency:

$$\Delta\mathcal{L}^{\infty}(\nu) = -\ln[L(\nu)] + \ln[L(\infty)]. \qquad (29)$$

The other possibility is to calculate the difference with the log-likelihood value at the frequency ν_{min} which minimises $-\ln(L)$:

$$\Delta\mathcal{L}^{Min}(\nu) = -\ln[L(\nu)] + \ln[L(\nu_{min})]. \qquad (30)$$

In the following subsections these two solutions will be compared with respect to there ability to give limits.

7.2 Likelihood referenced to infinity

Like in reference [2], the log-likelihood is referenced to its value for an infinite oscillation frequency:

$$\Delta\mathcal{L}^{\infty}(\nu) = -\ln[L(\nu)] + \ln[L(\infty)]. \qquad (31)$$

According to the central limit theorem of likelihood theory, for large sample of events, the log-likelihood difference between two hypotheses (with mixing of frequency ν and without mixing, $\nu = \infty$) is χ^2 distributed:

$$\Delta\mathcal{L}^{\infty}(\nu) = \frac{1}{2}\chi^2. \qquad (32)$$

The statistical fluctuations of $\Delta\mathcal{L}^{\infty}$ are therefore Gaussian. The value $\Delta\mathcal{L}^{\infty}_{data}(\nu)$, obtained for a given sample, can be compared to the average value, $\overline{\Delta\mathcal{L}^{\infty}_{Mix}(\nu)}$, expected for a real mixing at frequency ν. It can also be compared to the average value, $\overline{\Delta\mathcal{L}^{\infty}_{Nomix}(\nu)}$, expected for no mixing. If $\sigma(\Delta\mathcal{L}^{\infty}(\nu))$ is the variance of $\Delta\mathcal{L}^{\infty}$, any given value of ν can be excluded if the data value satisfies:

$$\Delta\mathcal{L}^{\infty}_{data}(\nu) > \overline{\Delta\mathcal{L}^{\infty}_{Mix}(\nu)} + 1.645 \cdot \sigma(\Delta\mathcal{L}^{\infty}(\nu)). \tag{33}$$

The global features of such log-likelihood functions can be understood in terms of Fourier analysis. Since a correct treatment would result in rather complicated formulas, following approximations are used, which keep the calculations simple without changing the conclusions:

- The log-likelihood is calculated for a binned distribution using the difference $\delta_i = N_i^{unlike} - N_i^{like}$ of like and unlike sign rates in bin i (at time t_i). This has to be compared to the expected distribution $\delta_i^0 = n \cdot (P_i^{unlike} - P_i^{like})$ (n is the total event rate). In this case the log-likelihood equals an χ^2-function: $-\ln(L) = \frac{1}{2}\chi^2$. This approximation is valid for large event samples.
- The FT of the time dependence of the signal function δ^{sig} is approximated by an exact Breit-Wigner with a global damping factor $F(\nu)$ instead of a distorted Breit-Wigner:

  $$FT(\delta^{sig})(\nu') \approx \frac{C}{2} F(\nu) \frac{\Gamma^2}{\Gamma^2 + (\nu' - \nu)^2}$$

  which implies

  $$\delta_i^0(\nu) = \delta_i^{sig}(\nu) + b_i = C \ F(\nu) \exp(-\Gamma t_i) \cos(\nu t_i) + b_i, \tag{34}$$

  where b_i stands for all background terms which do not show oscillatory behaviour and where $F(\nu) = f_s \cdot (1 - 2\eta) \cdot D(\nu, \sigma_l, \sigma_p)$, absorbs all damping terms due to mistag and resolution ($D(\nu, \sigma_l, \sigma_p)$, (17)) and C is a normalisation constant. This approximation holds if the vertex resolution dominates over the boost resolution.
- Signal and background follow the same exponential behaviour. Their ratio is constant as a function of time. Otherwise weights in the FT would be needed for exact calculations. This technical difficulty has prevented until now the FT method to be effectively used to give mixing results.

Using the data distribution δ_i the likelihood difference with infinity can be formulated:

$$\Delta\mathcal{L}^{\infty}_{data}(\nu) = \sum_i \frac{1}{2} \left(\frac{\delta_i - \delta_i^0(\nu)}{\sigma(\delta_i^0)} \right)^2 - \sum_i \frac{1}{2} \left(\frac{\delta_i - \delta_i^0(\infty)}{\sigma(\delta_i^0)} \right)^2. \tag{35}$$

Using δ_i^0 of equation (34) and the fact that at high frequency the oscillations are washed out due to the resolution and $F(\infty) = 0$, this gives:

$$\Delta\mathcal{L}^{\infty}_{data}(\nu) = -\sum_i \frac{(\delta_i - b_i) C F(\nu) \cos(\nu t_i) \exp(-\Gamma t_i)}{\sigma^2(\delta_i^0)} + \sum_i \frac{1}{2} \frac{[C F(\nu) \cos(\nu t_i) \exp(-\Gamma t_i)]^2}{\sigma^2(\delta_i^0)}.$$

$\sigma^2(\delta_i^0)$ is given by the total number of events in each bin $C f_s \exp(-\Gamma t_i) + b(t) = C \exp(-\Gamma t_i)$, assuming that the background follows the same exponential behaviour as the signal:

$$\Delta\mathcal{L}^{\infty}_{data}(\nu) \approx -F(\nu) \left[\sum_i (\delta_i - b_i) \cos(\nu t_i)\right] + \frac{1}{4} N C F(\nu)^2. \qquad (36)$$

The term in square brackets can be interpreted as Fourier transformation of the data minus the expected background.

The same expression can be derived directly from the Fourier transformation. Using the notation given in (23) - (25) [†]:

$$\Delta\mathcal{L}^{\infty}_{data}(\nu) = \sum_i \frac{1}{2} \left[\frac{d(\nu_i) - a^{\nu}(\nu_i)}{\sigma_d}\right]^2 - \sum_i \frac{1}{2} \left[\frac{d(\nu_i) - a^{\infty}(\nu_i)}{\sigma_d}\right]^2. \qquad (37)$$

$a^{\nu}(\nu_i)$ is the expected Fourier amplitude at ν_i if the real frequency is ν:

$$\begin{aligned} a^{\nu}(\nu_i) &= 0 \qquad |\nu_i - \nu| > 2\Gamma \\ &= a(\nu) \;\; \nu_i = \nu, \end{aligned} \qquad (38)$$

$a(\nu) = FT[\delta_0 - b(t)](\nu)$ is identical with (20). Because $a^{\infty}(\nu_i) = 0$, only the terms with $\nu_i = \nu$ remain and the results is:

$$\Delta\mathcal{L}^{\infty}_{data}(\nu) \;\; = \;\; -\frac{d(\nu) a(\nu)}{\sigma_d^2} + \frac{1}{2} \frac{a(\nu)^2}{\sigma_d^2}, \qquad (39)$$

[†] This form is strictly valid only if the correlations between different frequency bins can be ignored. This is the case if the bin width is larger than 2Γ. Otherwise the correlations have to be taken into account and the expression becomes:

$$\Delta\mathcal{L}^{\infty}_{data}(\nu) = [d(\nu_i) - a^{\nu}(\nu_i)]^T (V_{ij})^{-1} [d(\nu_j) - a^{\nu}(\nu_j)] - [d(\nu_i) - a^{\infty}(\nu_i)]^T (V_{ij})^{-1} [(d(\nu_j) - a^{\infty}(\nu_j)],$$

where V_{jj} is the covariance matrix given by $V_{ij} = \sigma_d^2 C_{ij}$ as defined in (76). In the approximations used here, it turns out that the off-peak and off-diagonal elements cancel exactly, resulting also in expression (39).

which is identical with (36). This allows to calculate the expected value of $\Delta\mathcal{L}^{\infty}$, averaged over a large number of experiments. If there is a signal at frequency ω, $\overline{d(\omega)} = a(\omega)$ and:

$$\overline{\Delta\mathcal{L}^{\infty}_{Mix}(\omega)} = -\frac{1}{2}a^2(\omega)/\sigma_d^2, \tag{40}$$

otherwise, if no mixing exists, $\overline{d(\nu)} = 0$ and:

$$\overline{\Delta\mathcal{L}^{\infty}_{Nomix}(\nu)} = +\frac{1}{2}a^2(\nu)/\sigma_d^2 \tag{41}$$

Of course the likelihood reflects the noise, it has fluctuations with a variance:

$$\sigma(\Delta\mathcal{L}^{\infty}(\nu)) \;=\; \frac{a(\nu)}{\sigma_d}. \tag{42}$$

Two interesting equations are therefore satisfied by log-likelihood with reference to infinity averages:

$$\overline{\Delta\mathcal{L}^{\infty}_{Mix}(\omega)} \;=\; -\overline{\Delta\mathcal{L}^{\infty}_{Nomix}(\omega)}, \tag{43}$$

$$\sigma(\Delta\mathcal{L}^{\infty}(\nu)) \;=\; \sqrt{2\cdot\overline{\Delta\mathcal{L}^{\infty}_{Nomix}(\nu)}}. \tag{44}$$

An example of such a likelihood is shown in Fig (4). If the systematic errors σ_s are included, the variance becomes:

$$\sigma(\Delta\mathcal{L}^{\infty}(\nu)) \;=\; \frac{a(\nu)}{\sigma_d^2}\sqrt{\sigma_d^2+\sigma_s^2}. \tag{45}$$

Using the above expressions of the log-likelihood in terms of Fourier amplitudes, the limit condition given in (33) can be converted to

$$-d(\nu) + a(\nu) = 1.645\sqrt{\sigma_d^2+\sigma_a^2}, \tag{46}$$

like in (26). Therefore the limits obtained with this method are identical with the Fourier limits.
Analytical formulae for computing the expected likelihood distributions (average and rms) from the time distributions are given in the Appendix (12.1).

This is the classical use of log-likelihood [13]. In this method, the frequency ν_{min}, which minimises $-\ln(L)$ is searched for. The log-likelihood difference

$$\Delta\mathcal{L}^{Min}(\nu) = -\ln[L(\nu)] + \ln[L(\nu_{min})], \tag{47}$$

will be referred as likelihood referenced to the minimum. In principle the frequency ν_{min} is the estimator of the true oscillation frequency and all frequencies ν such that $\Delta\mathcal{L}^{Min}(\nu) > \Delta\mathcal{L}_{cut}$ can be excluded at a certain confidence level. Theoretical values for $\Delta\mathcal{L}_{cut}$ are 1.92 for a double sided 95% C.L. (excluding values below and above a certain region around ν_{min}) and 1.34 for a single sided 95% C.L. (only a lower limit is given, values above ν_{min} are not constrained). This method is used in [5] [‡]. However, here this simple rule cannot be applied. If the real oscillation frequency is beyond the sensitivity of the experiment (e.g. $\omega_{true} = \infty$) very likely a minimum at some intermediate frequency will be found which can be estimated as follows: The contour of the -1σ range of the average likelihood is given by (see also Figure 4):

$$-\ln(L)^{Inf}_{-1\sigma} = \frac{1}{2}\frac{a^2}{\sigma_d^2} - \frac{a}{\sigma_d}. \tag{48}$$

This contour has a minimum at a value ν_{min} such that $a(\nu_{min}) = \sigma_d$. In the absence of a significant signal it is therefore very likely that the minimum is found in the region where $a(\nu) \approx \sigma_d$. This is not surprising. In this region the statistical fluctuations due to the noise σ_d have about the same size as the expected signal $a(\nu)$. Hence the frequency ν_{min} is not a good and unbiased estimator of the true frequency, and the simple rules for $\Delta\mathcal{L}_{cut}$ are not justified (this is discussed in [13]).

The problem can be bypassed by calibrating the likelihood difference using Monte Carlo [2]: A large number of Monte-Carlo experiments with a true frequency ω are generated. For each MC experiment ν_{min} is determined and $\Delta\mathcal{L}^{\min}(\omega)$ is calculated. A value $\Delta\mathcal{L}_{cut}$ can be determined such that $\Delta\mathcal{L}^{\min}(\omega) < \Delta\mathcal{L}_{cut}$ in 95% of the samples. This will result in a condition for double sided limits. In order to obtain $\Delta\mathcal{L}_{cut}$ for single sided lower limits the condition $\omega < \nu_{min}$ is imposed. The result of this calibration is given for a particular choice of parameters in Fig. 6, both for the double sided definition as described above and for the single sided definition. To date only the double

[‡] In this paper the 95% C.L. limit is defined by a difference of 1.92. This is not strictly correct, as the limit has to be understood as a lower limit, so actually it exceeds 95% C.L.

sided definition has been used, so only this will be discussed further. The log-likelihood cut value turns out to be always larger than the theoretical 1.92 and depends on ω: Only in the low frequency range, where a significant minimum is observed, the 95% likelihood difference is consistent with 1.92. At higher frequencies, where signal and noise becomes comparable, this value increases considerably. For the various Aleph analyses [2–4,8], where this calibration has been performed, values ranging from 1.9 to 3.5 and large variations with ω are observed. Although this method gives statistically correct limits (in the sense that it fails in less then 5%), it has some disadvantages:

- A lower limit set by the double-sided definition corresponds actually to a confidence level of 97.5%. The sensitivity of the experiment is therefore decreased. In Figure 5 the limits obtained by the two likelihood methods in a large number of Monte-Carlo experiments are shown. Plotted are average limits obtained at each true frequency, and the value such that 95% of the limits are below this value (If the limits are statistically correct this value must always be lower than the true frequency !). For large frequencies the limits obtained by the likelihood referenced to the minimum method are lower than the limits obtained by the likelihood referenced to infinity.
- It needs complicated and CPU intensive studies to calibrate the likelihood.
- When several analyses are combined by adding the individual log-likelihoods a calibration of this summed log-likelihood is necessary. However, it appears impossible to perform this calibration as it would require adding Monte Carlo generations from all these analyses (especially if they come from different experiments). This fatal obstacle does not hold for the first method (reference to infinity), because only little information (average and rms) has to be given to characterise completely the expected log-likelihood distribution of each analysis.

Because of these problems the likelihood referenced to the minimum is not recommended to be used in an analysis where a lower limit is derived.

8 Amplitude fit

A variation of the likelihood method is to fit for the amplitude A of the oscillation, as function of ν, first used in [8].

Using the notations and approximations of (34), the data distribution, binned in time, is δ_i. The values for each bin have statistical fluctuations given by a rms of $\sigma_i = \sqrt{n_i} = \sqrt{C \exp(-\Gamma t_i)}$. The average data function expected for full mixing with frequency ω is:

$$\begin{aligned}\overline{\delta_i} &\approx C f_s(1-2\eta) D(\omega,\sigma_l,\sigma_p)\exp(-\Gamma t_i)\cos(\omega t_i) + b_i \\ &= C\ F(\omega)\exp(-\Gamma t_i)\cos(\omega t_i) + b_i,\end{aligned}$$

C is a normalisation constant, $F(\nu)$ absorbs the damping terms due to sample purity, mistag, and resolution as given in 34). The amplitude fit function is

$$\delta^0(A,\nu)_i \;=\; C\ F(\omega)A(\nu)\exp(-\Gamma t_i)\cos(\omega t_i) + b_i \tag{49}$$

If the amplitude parameter A is one, the nominal oscillation is observed, else it should be 0. The likelihood for an amplitude parameter A at frequency ν is:

$$-\ln[L(A,\nu)] \;=\; \sum_i \frac{1}{2}\frac{[\delta^0(A,\nu)_i - \delta_i]^2}{\sigma_i^2}, \tag{50}$$

This likelihood is minimised with respect to A. Because $F(\nu)$, which enters in $\delta(A,\nu)_i$, does not depend on t_i, one obtains:

$$0 = \sum_i [\delta^0(A,\nu)_i - \delta_i]\cos(\nu t_i). \tag{51}$$

This is the real part of a Fourier transformation:

$$FT[\delta^0(t)](\nu) - FT[\delta(t)](\nu) = 0, \tag{52}$$

which results in:

$$A(\nu) \;=\; \frac{F(\omega)}{F(\nu)}\frac{\Gamma^2}{\Gamma^2 + (\nu-\omega)^2}. \tag{53}$$

$A(\nu)$ has the same Breit-Wigner shape as in the Fourier analysis. If $|\nu-\omega| > 2\Gamma$, A is about 0. If $\nu = \omega$, $A = 1$. Since $FT(\delta)$ has a noise of $\sqrt{n/2}$, the statistical error on A is:

$$\sigma(A) \;=\; \sqrt{\frac{2}{n}}\frac{1}{F(\nu)}. \tag{54}$$

Because A is essentially a normalised Fourier amplitude, (37) can be used to calculate the behaviour of an amplitude fit:

$$-\ln[L(A,\nu)] = \sum_i \frac{1}{2}\left[\frac{d(\nu_i) - Aa^{\nu}(\nu_i)}{\sigma_d}\right]^2 - \sum_i \frac{1}{2}\left[\frac{d(\nu_i) - Aa^{\infty}(\nu_i)}{\sigma_d}\right]^2. \tag{55}$$

with the definitions of $a^{\nu}(\nu_i)$, etc., like above. Again:

$$-\ln[L(A,\nu)] \;=\; -A\frac{d(\nu)a(\nu)}{\sigma_d^2}+\frac{1}{2}A^2\frac{a(\nu)^2}{\sigma_d^2}. \tag{56}$$

Minimising $-\ln[L(A)]$ results in:

$$A \;=\; \frac{d(\nu)}{a(\nu)}. \tag{57}$$

Since $d(\nu)$ fluctuates by σ_d, A varies by:

$$\sigma(A)=\frac{\sigma_d}{a(\nu)}. \tag{58}$$

Of course the systematic errors have to be added. This can be done in the classical way, varying the respective input parameters. The correct treatment of systematic errors in the A fit is discussed in the Appendix (86).

In order to obtain a limit, one can simply exclude all values of ν where $A=1$ can be ruled out at the desired confidence level. This can be calculated using the error on A from the fit. The limit condition is:

$$A(\nu) \;+\; 1.645\cdot\sigma[A(\nu)]=1, \tag{59}$$

$$\frac{d(\nu)}{a(\nu)}+1.645\frac{\sigma_d}{a(\nu)} \;=\; 1, \tag{60}$$

which is again identical with (26). Again the smallest ν which satisfies this condition gives the lower limit. As one obtains a series of measurements of A, different experiments can easily be combined by averaging the respective A values taking into account the errors.
This can be demonstrated by making a joint fit of two independent experiments. Two measurements (labelled m = 1 , 2) with individual likelihoods:

$$-\ln[L(A_m,\nu)]_m \;=\; -A_m\frac{d(\nu)_m a(\nu)_m}{\sigma_{dm}^2}+\frac{1}{2}A_m^2\frac{a(\nu)_m^2}{\sigma_{dm}^2}. \tag{61}$$

with individual solutions $A(\nu)_m=\frac{d(\nu)_m}{a(\nu)_m}$, $\sigma(A)_m=\frac{\sigma_{dm}}{a(\nu)_m}$. The same average amplitude is obtained by the joint fit of the two likelihoods and the weighted average of the two solutions:

$$\sigma(\bar{A})=\left(\sigma(A)_1{}^{-2}+\sigma(A)_2{}^{-2}\right)^{-\frac{1}{2}},$$

$$\bar{A} = \left(\frac{A(\nu)_1}{(\sigma(A)_1)^2} + \frac{A(\nu)_2}{(\sigma(A)_2)^2} \right) \cdot \sigma(\bar{A})^2.$$

The results for $A(\nu)$ can be treated like independent measurements of the same physical quantity and therefore can be combined. Of course common systematic errors have to be treated properly, this requires an individual breakdown of the systematic errors of the individual measurements.

9 Comparison of the methods

A general comparison of the oscillation analysis methods (Fourier transformation, likelihood referenced to infinity, and the amplitude fit) has been performed using a simple Monte Carlo. The Monte Carlo parameters are shown in Table 1a) (with perfect boost resolution). In Figures 7 and 8 two examples are shown. In figure 7 the generated frequency is $\Delta m_s = 5.5$ ps^{-1}, the signal is clearly visible in all analysis. In figure 8 the generated frequency was set to infinity. Here the lower limits obtained by the three methods agree. It was verified with high statistics that the three methods give identical results, and that the 95% C.L. limits are correct, that is that the generated frequency is lower than the limit in less than 5% of the experiments. This is of course expected from the calculations presented in the sections above. The formulae relating the different methods are summarised in the table 2 and 3.

10 Conclusions

Fourier analysis is a powerful tool to analyse data for $B_s^0 \bar{B}_s^0$ oscillations. It offers the possibility to calculate the effects of the experimental resolution and statistical fluctuations. This allows a prediction of the expected signal and the statistical noise. Hence the significance of a possible signal or the confidence level of the absence of such a signal can be calculated correctly. Since it is a linear transformation of the data, systematic errors on input parameters propagate in a straightforward manner and systematic effects can be controlled easily.

Furthermore, using Fourier analysis, the properties of alternative analysis methods, like likelihood fits, can be studied. It can be shown that a similar way to do the analysis is the amplitude fit: for each oscillation frequency an hypothetical amplitude is fitted. This method has the same properties as a Fourier analysis, however, it has the advantage that a 'classical' likelihood fit can be used. On the other hand, a maximum likelihood fit with the frequency as free parameter will generally not result in correct limits.

It could be shown that Fourier analysis, likelihood referenced to infinity and amplitude fit are mathematically equivalent methods. To derive this correspondence some approximations were used, this was mainly done to keep the calculations simple, the main features are conserved in the exact formalism. All properties given in this paper have been verified by extensive Monte Carlo studies. Amongst them:

- The damping of the oscillations due to vertex and boost resolution given in (17).
- The noise theorems derived in section 5.
- The features of the likelihood referenced to infinity (43), (44).
- The correspondence between the amplitude fit and the likelihood referenced to infinity (table 3).

It should be stressed that the Fourier transformation of the background subtracted data plotted together with the expected peak size, or equivalent a plot of the A values, contains several important informations:

- The correctness of the background subtraction: the Fourier spectrum should be flat (within the fluctuations allowed by statistics).
- The typical sensitivity given by the frequency where the expectation (or $A = 1$) intersects the 1.645σ band around 0.
- The actual limit, where the expectation (or $A = 1$) intersects the 1.645σ band around the data
- The 'luck' of the experiment, given by the deviation of the data from 0 at the limit.

Furthermore the A-fit offers a possibility to combine the results of independent analyses by simply averaging the different A-spectra taking into account the errors. In the case of likelihood methods, a similar combination is only possible when using reference to infinity. In this case individual likelihood functions must be added . The log-likelihood distributions expected for each analysis, simply characterised by average and rms, can be easily combined and used to derive a limit. Such a combination is not feasible when using likelihood differences from minimum. This would need a calibration of the total likelihood which appears to be not feasible.

11 Acknowledgements

We thank Sandrine Emery, Roger Forty, Marie Claude Lemaire and Jacques Lefrancois for sharing part of the tests needed by these studies, for stimulating discussions and useful suggestions.

12 Appendix

12.1 Analytic formulae for computing the expected likelihood distribution

The log-likelihood difference, referenced to infinity, is given for any mixing frequency $\nu = \Delta m_s$, by

$$\Delta\mathcal{L}^{\infty}(\nu) = -\sum_{i=1} n_i \ln\left[\frac{P_i(\nu,\alpha^0)}{P_i(\infty,\alpha^0)}\right],$$

where n_i is the measured rate in time bin i and P_i is the expected time distribution probability (given in section 7). The summation is made for the like and unlike sign distributions, the indexes are omitted for simplicity. α^0 denotes the set of parameters (fractions f, mistag η, and other parameters) on which the charge correlation depends.

A Monte Carlo generating data according to P_i has been used to verify that the distribution of $\Delta\mathcal{L}^{\infty}$ is Gaussian, as expected from first principles (see section 7.2). Therefore the log-likelihood difference distribution is completely determined by its average and its sigma.

This average and sigma originating from statistical fluctuations and systematic uncertainties can be calculated analytically:

For samples generated with a mixing frequency ν, the average log-likelihood difference is given by:

$$\Delta\mathcal{L}^{\infty}_{Mix}(\nu) = -n\sum_{i=1} P_i(\nu,\alpha^0)\ln\left[\frac{P_i(\nu,\alpha^0)}{P_i(\infty,\alpha^0)}\right],$$

where n_i has been replaced by its average $n\cdot P_i$ to obtain the expected average time distribution. Similarly, for samples generated with $\nu = \infty$, the average log-likelihood difference is given by:

$$\Delta\mathcal{L}^{\infty}_{Nomix}(\nu) = -n\sum_{i=1} P_i(\infty,\alpha^0)\ln\left[\frac{P_i(\nu,\alpha^0)}{P_i(\infty,\alpha^0)}\right].$$

and the statistical rms of $\Delta\mathcal{L}^{\infty}(\nu)$ by:

$$\sigma^{stat}[\Delta\mathcal{L}^{\infty}(\nu)] = \sqrt{n\sum_{i=1} P_i(\nu,\alpha^0)\ln\left[\frac{P_i(\nu,\alpha^0)}{P_i(\infty,\alpha^0)}\right]^2},$$

The following relations for the average log-likelihood difference and the statistical rms can be demonstrated:

$$\Delta\mathcal{L}^{\infty}_{Mix}(\nu) = -\Delta\mathcal{L}^{\infty}_{Nomix}(\nu).$$

$$\sigma^{stat}[\Delta\mathcal{L}^{\infty}(\nu)] = \sqrt{2 \cdot \Delta\mathcal{L}^{\infty}_{Nomix}(\nu)}.$$

The systematic rms of $\Delta\mathcal{L}^{\infty}(\nu)$ originates from the fact that the best estimate of the parameters α^0 which is used, may not correspond to the true value of the data. For instance the true value of each parameter, say α_l, may differ from the estimated value α^0_l, by a systematic spread, assumed Gaussian of rms σ_{α_l}. The systematic rms of $\Delta\mathcal{L}^{\infty}(\nu)$ which results from this parameter uncertainty is:

$$\sigma^{syst}_l[\Delta\mathcal{L}^{\infty}(\nu)] = n\sigma_{\alpha_l} \sum_{i=1} \frac{\partial P_i(\nu,\alpha^0)}{\partial\alpha_l} \ln\left[\frac{P_i(\nu,\alpha^0)}{P_i(\infty,\alpha^0)}\right].$$

The total systematic error is given by adding quadratically the effects of all parameters α_l:

$$\sigma^{syst}[\Delta\mathcal{L}^{\infty}(\nu)] = \sqrt{\sum_l \sigma^{syst}_l[\Delta\mathcal{L}^{\infty}(\nu)]^2}.$$

The total error is:

$$\sigma^{tot} = \sqrt{\sigma^2_{stat} + \sigma^2_{syst}},$$

the $\Delta\mathcal{L}^{\infty}$ distribution being Gaussian, the 95 % confidence level is obtained as:

$$\Delta\mathcal{L}^{\infty}_{95} = \Delta\mathcal{L}^{\infty}_{Mix}(\nu) + 1.645 \times \sigma^{tot}[\Delta\mathcal{L}^{\infty}(\nu)].$$

It has been verified extensively that these analytical computations give the same results as the Monte Carlo. They allow a much faster study of the errors, more ν points, and the separation of the systematic contributions of all parameters.

12.2.1 Wiener-Kinchin theorem

As mentioned in section 5 the determination of the noise in the Fourier transformation needs the calculation of the power spectrum. The power spectrum can be obtained easily applying the Wiener-Kinchin theorem. According to the Wiener-Kinchin theorem the rms power spectrum of a function is identical to the Fourier transformation of its average autocorrelation function A:

$$\overline{PW_j} \;=\; \overline{|FT(f)_j|^2} \;=\; \frac{1}{N}\sum_{k=0}^{N-1}\overline{A(f)_k}\exp\left(-2\pi i\frac{jk}{N}\right), \tag{62}$$

The autocorrelation function $A(f)_k$ is:

$$A(f)_k \;=\; \frac{1}{N}\sum_{j=0}^{N-1} f_j f_{j+k}. \tag{63}$$

The mean autocorrelation function of a binned distribution f_k can be calculated: The value f_k in each bin has the mean f_k^0 and the error σ_k, with a Gaussian probability distribution:

$$p(f_k, f_k^0) \;=\; \frac{1}{\sqrt{2\pi}\sigma_k}\exp\left(-\frac{1}{2}\frac{(f_k - f_k^0)^2}{\sigma_k^2}\right). \tag{64}$$

The fluctuations in the time bins are uncorrelated. The mean autocorrelation is:

$$\begin{aligned}\bar{A}(f)_{k,k\neq 0} &= \frac{1}{N}\sum_{j=0}^{N-1}\left[\int_{-\infty}^{+\infty}\int_{-\infty}^{+\infty}\left(f_j\; p(f_j, f_j^0)\; f_{j+k}\; p(f_{j+k}, f_{j+k}^0)\right)\; df_j\; df_{j+k}\right] \\ &= \frac{1}{N}\sum_{j=0}^{N-1}\left[f_j^0 f_{j+k}^0\right] = S_k,\end{aligned} \tag{65}$$

where S is defined by the respective summation. For $k = 0$ the expression is slightly different, as the same bin has to be averaged:

$$\begin{aligned}\bar{A}(f)_{k=0} &= \frac{1}{N}\sum_{j=0}^{N-1}\left[\int_{-\infty}^{+\infty}\left(f_j^2 p(f_j, f_j^0)\right) df_j\right] \\ &= \frac{1}{N}\sum_{j=0}^{N-1}\left[(f_j^0)^2 + \sigma_j^2\right] = S_0 + \frac{1}{N}\sum_{j=0}^{N-1}\sigma_j^2.\end{aligned} \tag{66}$$

For all k the expression is:

$$\bar{A}(f)_k \;=\; S_k + \delta_{0k}\frac{1}{N}\sum_{j=0}^{N-1}\sigma_j^2, \tag{67}$$

($\delta_{0k} = 1$ for $k = 0$, 0 otherwise). The FT of the power spectrum is:

$$FT[\bar{A}(f)]_j = FT(S)_j + \frac{1}{N^2}\sum_{j=0}^{N-1}\sigma_j^2. \tag{68}$$

This is the total power spectrum of the signal plus the noise contribution. Since $FT(S)_j = |FT(f)_j|^2$, the noise contribution alone is:

$$(rms_j^{noise})^2 \;=\; \frac{1}{N^2}\sum_{j=0}^{N-1}\sigma_j^2. \tag{69}$$

If the error in each bin is given by $\sigma_j \;=\; \sqrt{n_j}$, the population in each bin, the sum becomes $\sum \sigma_j^2 = n$, the total number of events:

$$rms_j^{noise} \;=\; \frac{\sqrt{n}}{N}. \tag{70}$$

The noise spectrum is flat, the noise is the same at all frequencies. In general, uncorrelated fluctuations in the time spectrum give rise to white noise in the frequency spectrum. This rms noise applies to the absolute value of the FT. For our purpose only the noise of the real part is of interest. The noise of the real and imaginary part is (mostly) uncorrelated and related by

$$\sigma_{\rm real} = \sigma_{\rm imag} = \frac{1}{N}\sqrt{\frac{n}{2}}.$$

The exact calculation, obtained using error propagation gives for the rms of the real part of a Fourier transformation of a binned distribution:

$$\sigma[\mathrm{REAL}(FT(\nu))] = \frac{1}{N}\sqrt{\frac{n}{2} + \frac{N}{2}\mathrm{REAL}[FT(n_i, 2\nu)]}, \tag{71}$$

$\mathrm{REAL}(FT(n_i, 2\nu))$ is the real part of the Fourier transformation of the event distribution at double the frequency. If the events follow an exponential time dependence $n(t) \;=\; n\Gamma\exp(-\Gamma t)$, the second term equals $n\frac{\Gamma^2}{\Gamma^2+(2\nu)^2}$ and becomes negligible for $\nu > \Gamma$. Hence for reasonable large frequencies the $\sqrt{\frac{n}{2}}$ law holds.

For completeness it should be mentioned that this has to be modified if event likelihood fits are used. This can be demonstrated by introducing the event Fourier transform $g(q_i,\nu) = \frac{1}{n}\sum_{i=1}^{n} q_i \cos(\nu t_i)$ of a sample of n events, $q_i = \pm 1$ whether the event is like or unlike sign. In this case the noise becomes:

$$\sigma[q(\nu)] = \sqrt{\frac{1}{2n} + \frac{1}{2n}g(|q_i|,2\nu))^2 - \frac{1}{n}q(q_i,\nu)^2}, \tag{72}$$

while the signal follows:

$$g[\Gamma\cos(\omega t)\exp(-\Gamma t)] = \frac{1}{2}\left(\frac{\Gamma^2}{\Gamma^2+(\omega-\nu)^2} + \frac{\Gamma^2}{\Gamma^2+(\omega+\nu)^2}\right). \tag{73}$$

Again there is a contribution due to the event distribution, which becomes negligible for $\nu > \Gamma$. However there is an additional noise reduction close to the signal frequency ω. This is due to the fact that in an event-likelihood the total number of events is fixed and has no error. Using the usual definitions (20), (17) for the damping due to purity, mistag and resolution:

$$\epsilon = f_s(1-2\eta)D(\omega,\sigma_l,\sigma_p),$$

the signal/noise at the signal frequency ω becomes:

$$S/N = \epsilon\sqrt{\frac{n}{2}}\left(\sqrt{1-\frac{\epsilon^2}{2}}\right)^{-1}.$$

Since ϵ is small in most of the analyses, the signal/noise is again $S/N = \epsilon\sqrt{\frac{n}{2}}$. Only for high purity, low mistag analyses with a very good resolution, ϵ can approach one, and the correction needs to be applied.

12.2.2 Correlations of the noise

In general the noise fluctuations in adjacent frequency bins of the FT are correlated; this correlation can also be calculated using the Wiener-Kinchin theorem. What is needed is the mean autocorrelation function of the deviations from the mean Fourier spectrum $\Delta g_k = g_k - g_k^0 = FT(f)_k - FT(f^0)_k$:

$$\overline{A(\Delta g)_k} = \frac{1}{N}\sum_{j=0}^{N-1}\overline{\Delta g_j \Delta g_{j+k}} = FT^{-1}(\overline{|FT(\Delta g)|^2})_k = FT^{-1}(\overline{|f-f^0|^2})_k. \tag{74}$$

Since $\sigma(f_i) = \sqrt{f_i}$, $\overline{|f_i - f_i^0|^2} = \sigma(f_i)^2 = f_i^0$:

$$\overline{A(\Delta g)_k} \;=\; FT^{-1}(f^0)_k. \tag{75}$$

Hence, if the basic time distribution is an exponential with a decay time $\tau = 1/\Gamma$, the mean autocorrelation of the FT is a Breit-Wigner $\frac{\Gamma^2}{\Gamma^2+\nu^2}$. The noise spectrum is then a series of Breit-Wigner bumps. As a consequence the noise fluctuations in two frequency bins ν_i and ν_j are uncorrelated only for $|\nu_i - \nu_j| > 2\Gamma$. The correlation coefficient is given by [§]:

$$C_{ij} \approx \frac{\Gamma^2}{\Gamma^2 + (\nu_i - \nu_j)^2} \tag{76}$$

12.3 Systematic errors in the amplitude fit

In the formulas for the likelihood background contributions have been neglected up to now. In general the oscillation signal sits on top of background, from other physics sources but also from the exponential component of the B_s^0 signal. For a correct treatment of systematic errors this has to be taken into account. Therefore (23) - (25) has to be modified:

$$\begin{aligned} \overline{d(\nu_i)} &= d_0(\nu_i) && |\omega - \nu_i| > 2\Gamma \\ &= a(\omega) + d_0(\nu_i) && \omega = \nu_i \\ \sigma[d(\nu_i)] &= \sigma_d && \text{all i,} \end{aligned} \tag{77}$$

where $d_0(\nu)$ stands for all background terms. The systematic error due to a parameter p with error $\sigma(p)$ may affect signal and background:

$$\sigma[a(\omega)]_{sys} = \frac{\partial a(\omega)}{\partial p}\sigma(p), \qquad \sigma[d_0(\nu)]_{sys} = \frac{\partial d_0(\nu)}{\partial p}\sigma(p) \tag{78}$$

Both errors are of course fully correlated. The expression for a limit is now:

$$a(\nu) - [d(\nu) - d_0(\nu)] > 1.645\sqrt{\sigma_d^2 + \left[\left(\frac{\partial a(\nu)}{\partial p} + \frac{\partial d_0(\nu)}{\partial p}\right)\sigma(p)\right]^2}. \tag{79}$$

[§] Again this is an approximation for $\nu > \Gamma$, the exact formula is

$$C_{ij} = \frac{\frac{\Gamma^2}{\Gamma^2+(\nu_i-\nu_j)^2} + \frac{\Gamma^2}{\Gamma^2+(\nu_i+\nu_j)^2}}{\sqrt{1+\frac{\Gamma^2}{\Gamma^2+(2\nu_i)^2}}\sqrt{1+\frac{\Gamma^2}{\Gamma^2+(2\nu_j)^2}}}$$

This can be converted into a condition for the A-fit using:

$$A(\nu) = \frac{d(\nu) - d_0(\nu)}{a(\nu)}, \qquad \sigma[A(\nu)] \;=\; \frac{\sigma_d}{a(\nu)}, \tag{80}$$

and therefore:

$$1 - A(\nu) > 1.645\sqrt{\sigma[A(\nu)]^2 + \left[\left(\frac{1}{a(\nu)}\frac{\partial a(\nu)}{\partial p} + \frac{1}{a(\nu)}\frac{\partial d_0(\nu)}{\partial p}\right)\sigma(p)\right]^2}, \tag{81}$$

which can be formulated like:

$$1 - A(\nu) > 1.645\sqrt{\sigma[A(\nu)]^2 + \sigma[A(\nu)]^2_{sys}}, \tag{82}$$

with the definition:

$$\sigma[A(\nu)]_{sys} = \left(-\frac{1}{a(\nu)}\frac{\partial d_0(\nu)}{\partial p} - \frac{1}{a(\nu)}\frac{\partial a(\nu)}{\partial p}\right)\sigma(p). \tag{83}$$

However, if the parameter p varies in the fit, A changes by:

$$\Delta A(\nu) = \left(-\frac{1}{a(\nu)}\frac{\partial d_0(\nu)}{\partial p} - \frac{A(\nu)}{a(\nu)}\frac{\partial a(\nu)}{\partial p}\right)\sigma(p), \tag{84}$$

which is not identical with the definition of $\sigma(A)_{sys}$ (83)! In general the statistical error of A changes also with p:

$$\Delta\sigma[A(\nu)] = -\sigma[A(\nu)]\frac{1}{a(\nu)}\frac{\partial a(\nu)}{\partial p}\sigma(p), \tag{85}$$

which can be used to obtain:

$$\sigma[A(\nu)]_{sys} = \Delta A(\nu) + (1 - A)\frac{\Delta\sigma(A)}{\sigma[A(\nu)]}. \tag{86}$$

With this expression the limits obtained form the A fit are identical with those from the Fourier analysis. Formula (86) takes into account, that varying a parameter p changes both A and the error on A.

References

[1] D. Buskulic *et al.* (ALEPH Collaboration), Phys. Lett. **B313** (1993) 498
D. Buskulic *et al.* (ALEPH Collaboration), Phys. Lett. **B322** (1994) 441
R. Akers *et al.* (OPAL Collaboration), Phys. Lett. **B327** (1994) 411
R. Akers *et al.* (OPAL Collaboration), Phys. Lett. **B336** (1994) 485
P. Abreau *et al.* (DELPHI Collaboration), Phys. Lett. **B338** (1994) 409

[2] D. Buskulic *et al.* (ALEPH Collaboration), Phys. Lett. **B322** (1994) 441

[3] D. Buskulic *et al.* (ALEPH Collaboration), Phys. Lett. **B356** (1995) 409

[4] D. Buskulic *et al.* (ALEPH Collaboration), CERN PPE/96-030 (1996), to be published in Phys. Lett. **B**

[5] R. Akers *et al.* (OPAL Collaboration), Z. Phys. **C66** (1995) 555

[6] OPAL Collaboration,'A study of B Meson Oscillations Using Inclusive Lepton Events', submitted to the International Symposium on Lepton-Photon Interactions, Beijing, China (1995).

[7] DELPHI Collaboration, 'Improved Measurement of the oscillation frequency of B^0 mesons', Contributed paper to the International Europhysics Conference on High Energy Physics, Brussels, Belgium (1995), reference eps0568

[8] ALEPH Collaboration, 'Time dependent B_s^0 mixing from lepton-kaon correlations', Contributed paper to the International Europhysics Conference on High Energy Physics, Brussels, Belgium (1995), reference eps0410

[9] O. Schneider, talk given at the EPS Conference, Brussels, Belgium, July 1995.

[10] H.-G. Moser, Nucl. Inst. & Meth. **A295** (1990) 435

[11] A. Pais and S.B. Treiman, Phys. Rev. **D12** (1975) 2744
L.B. Okun, V.I. Zakharov and B.M. Pontecorvo, Nuovo Cim. Lett **13** (1975) 218
J. Ellis, M.K. Gaillard and D.V. Nanopoulos, Nucl. Phys. **B109** (1976) 213
A.J. Buras, W. Slominski and H. Steger, Nucl. Phys. **B245** (1984) 369

[12] A. Ali and D. London, CERN-TH 7398/94 (1994)
S. Herrlich and U. Nierste, PSI-PR-95-13 (1995)
S. Narison, HEP-PH-9503234 (1995)

[13] Review of Particle Properties, Phys. Rev. **D50** (1994) 1275

[14] Review of Particle Properties, Phys. Rev. **D50** (1994) 1278

parameter	value (a)	value (b)	value (c)
n	10000	10000	200
f_s	0.12	0.12	0.65
η	0.18	0.18	0.25
σ_l	0.15 ps	0.15 ps (50%), 0.45 ps (50%)	0.12 ps (80%), 0.45 ps (20%)
σ_p	0	0.1 (65%), 0.25 (35%)	0.08 (80%), 0.25 (20%)

Table 1
Parameters of typical B_s^0 oscillation experiments at LEP. The number of events corresponds to a sample of 3.3×10^6 Z^0. Example a) correspond to a dilepton analysis with a single Gaussian boost resolution. Example b) shows a more realistic resolution function, adding tails and the boost resolution. In example c) the scenario of an exclusive analysis is given (e.g. D_s-lepton with a jet-charge tag). The statistics is much smaller, this is compensated by higher purity and improved resolution.

quantity	FT	Likelihood(∞)	A-fit
data	$d(\nu)$	$\Delta\mathcal{L}^{\infty}(\nu)_d = \frac{-d(\nu)a(\nu)}{\sigma_d^2} + \frac{1}{2}\frac{a(\nu)^2}{\sigma_d^2}$	$A_d = \frac{d(\nu)}{a(\nu)}$
expectation (signal at ν)	$d_{Mix}(\nu) = a(\nu) = \frac{n}{2} f_s(1-2\eta)D(\omega,\sigma_l,\sigma_p)$	$\Delta\mathcal{L}^{\infty}_{Mix}(\nu) = -\frac{1}{2}\frac{a(\nu)^2}{\sigma_d^2}$	$A_{mix}(\nu) = 1$
expectation (no signal)	$d_{Nomix}(\nu) = 0$	$\Delta\mathcal{L}^{\infty}_{Nomix}(\nu) = +\frac{1}{2}\frac{a(\nu)^2}{\sigma_d^2}$	$A_{Nomix}(\nu) = 0$
variance	$\sigma_d = \sqrt{n/2}$	$\sigma[\Delta\mathcal{L}^{\infty}(\nu)] = \frac{a(\nu)}{\sigma_d}$	$\sigma(A(\nu)) = \frac{\sigma_d}{a(\nu)}$
limit condition	$a(\nu) > d(\nu) + c\sigma_d$	$\Delta\mathcal{L}^{\infty}(\nu)_d > \Delta\mathcal{L}^{\infty}_{Mix}(\nu) + c\sigma[\Delta\mathcal{L}^{\infty}(\nu)]$	$A(\nu) < 1 - c\sigma[A(\nu)]$

Table 2
Corresponding quantities in the Fourier analysis, likelihood with reference to infinity and amplitude fit. The Fourier quantities are defined by the analysis parameters, f_s (signal fraction), η (mistag), n (total event number), σ_l and σ_p (vertex and boost resolution), c defines the confidence level, e.g. 1.645 for 95%.

Likelihood	A-fit
data likelihood	$\Delta\mathcal{L}^{\infty}(\nu) = [\frac{1}{2} - A(\nu)]\frac{1}{\sigma[A(\nu)]^2}$
average (signal at ν)	$\Delta\mathcal{L}^{\infty}(\nu)_{mix} = -\frac{1}{2}\frac{1}{\sigma[A(\nu)]^2}$
average (no signal)	$\Delta\mathcal{L}^{\infty}(\nu)_{nomix} = \frac{1}{2}\frac{1}{\sigma[A(\nu)]^2}$
variance	$\sigma[\Delta\mathcal{L}^{\infty}(\nu)] = \frac{1}{\sigma[A(\nu)]}$

Table 3
Correspondence between the amplitude fit and the likelihood referenced to infinity. This relations allow to calculate the likelihood function from the A-fit and vice versa.

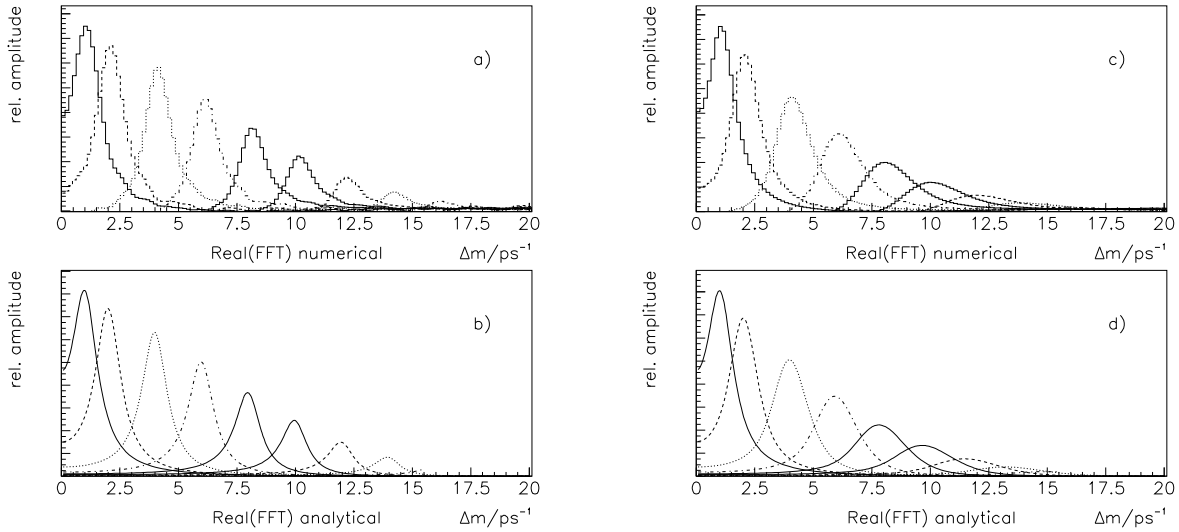

Fig. 1. FFT of oscillations (real part) with a vertex resolution of 0.15 ps (a) and the approximate analytical calculation (b). The peaks correspond to ω = 1,2,4,6,8,12,14,16 ps^{-1}. The vertical scales are arbitrary. The same calculations with an additional boost resolution of 10% are shown in (c,d).

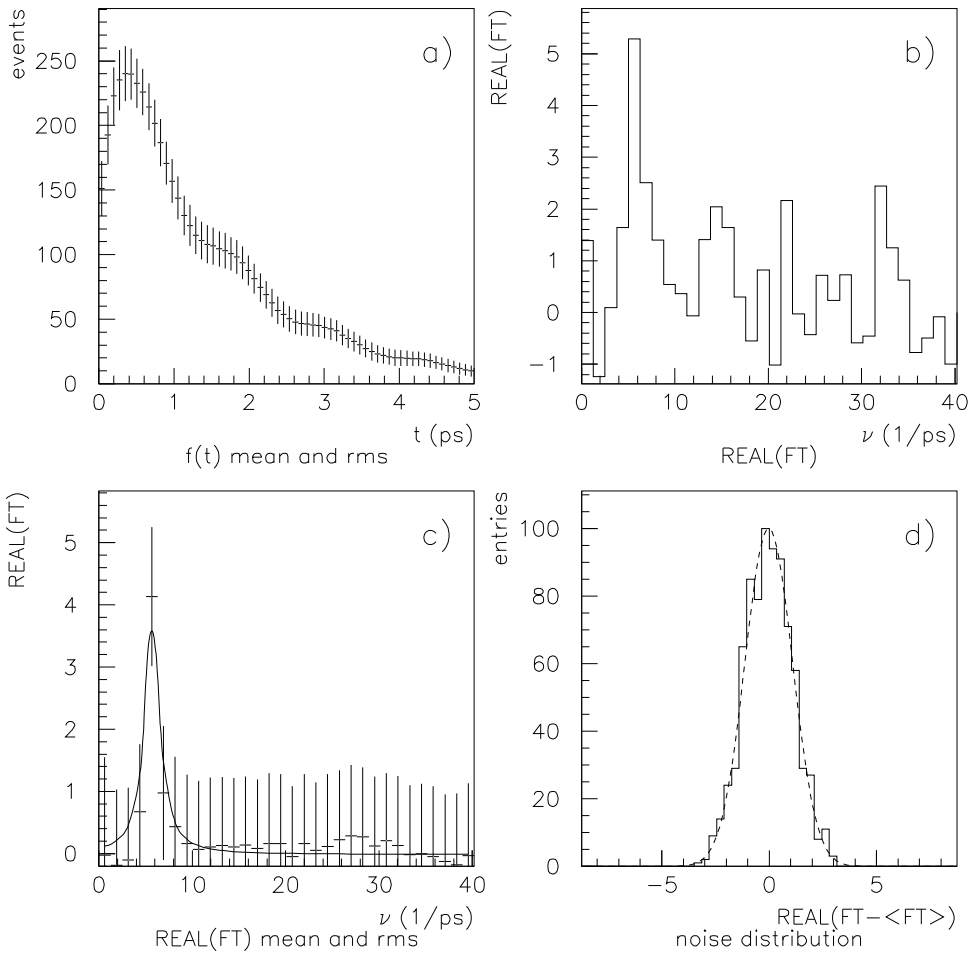

Fig. 2. a) Time distribution of an oscillation signal with $\Delta m = 6$ ps^{-1} (average of 800 experiments, *mean* $\pm$ *rms*). b) Fourier transformation of a single experiment (corrected for the underlying exponential background). c) Fourier transformation (*mean* $\pm$ *rms*) averaged over the same 800 experiments. d) Distribution of the noise. Plotted is the distribution of the Fourier amplitude around its average value. The dashed line is the expectation from the Wiener Kinchin theorem.

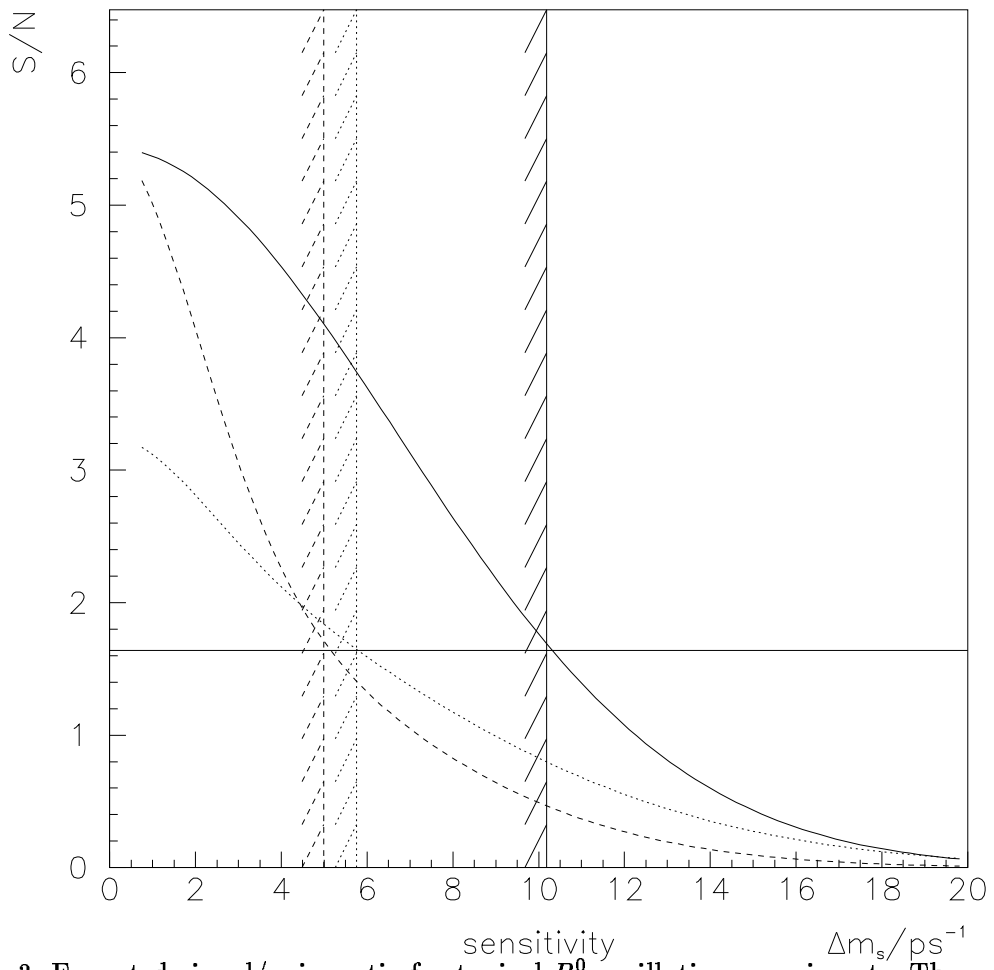

Fig. 3. Expected signal/noise ratio for typical B_s^0 oscillation experiments. The parameters of the various analyses are shown in Table 1. The solid lines assumes a vertex resolution of 0.15 ps^{-1} and a perfect boost reconstruction. Mistag, purity and statistics corresponds to a dilepton analysis (example a) in Table 1). The dashed line represents a more realistic case adding tails to the resolution function (example b)). The dotted line corresponds to an exclusive analysis using D_s-lepton events, with lower statistics, higher purity and improved resolution (example c)). Although the signal/noise ratio is inferior at low frequencies, the analysis improves at high frequencies due to the better resolution.
The vertical lines show the intersection with the $S/N = 1.645$ line, the condition for a 95% lower limit.

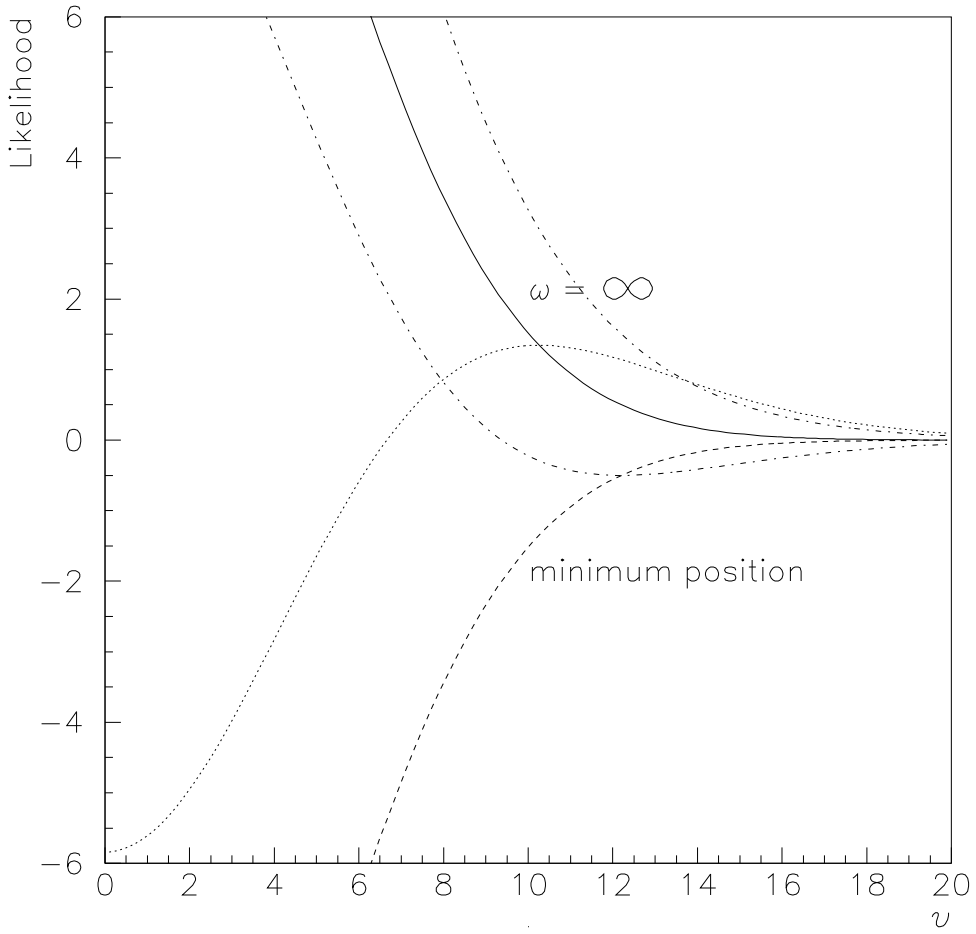

Fig. 4. Typical log-likelihood function versus the oscillation frequency: Plotted is the average likelihood for $\omega = \infty$ (solid line), the average minimum in case of oscillations (dashed line) and the $\pm 1\sigma$ contours around the average likelihood (dashed dotted lines). The 95%C.L. upper limit of the log-likelihood in case of mixing is shown as dotted line.

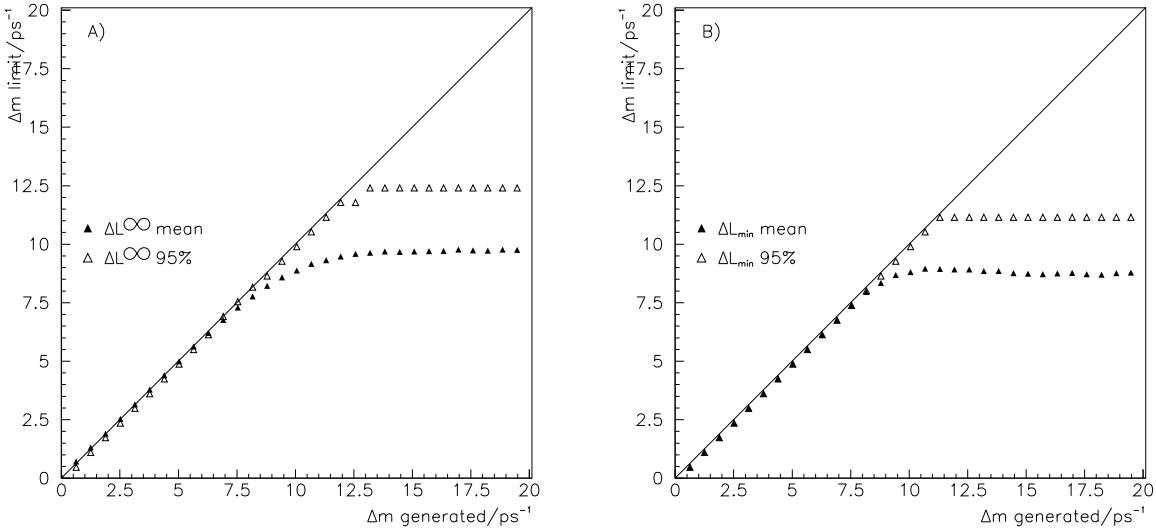

Fig. 5. Frequency limits obtained from Monte Carlo experiments. A large number of Monte Carlo experiment (each corresponding to 10000 events with $f_s = 0.12$, $\eta = 0.18$, and 0.15 ps resolution) are generated with the same input frequency. Two different methods are used to obtain from each experiment a lower limit on the frequency:
Figure a) obtaining the limit from the likelihood referenced to infinity. Figure b) using the calibrated likelihood referenced to the minimum (double sided definition). For each frequency two values are displayed: The full dots show the average of all limits. The empty dots indicate the value which exceeds 95% of the limits. If the limits are correct the empty dots must always be equal or less than the generated value. At high input values where the oscillations are washed out by the resolution the limit stays at a constant level, indicating the sensitivity of the experiment.

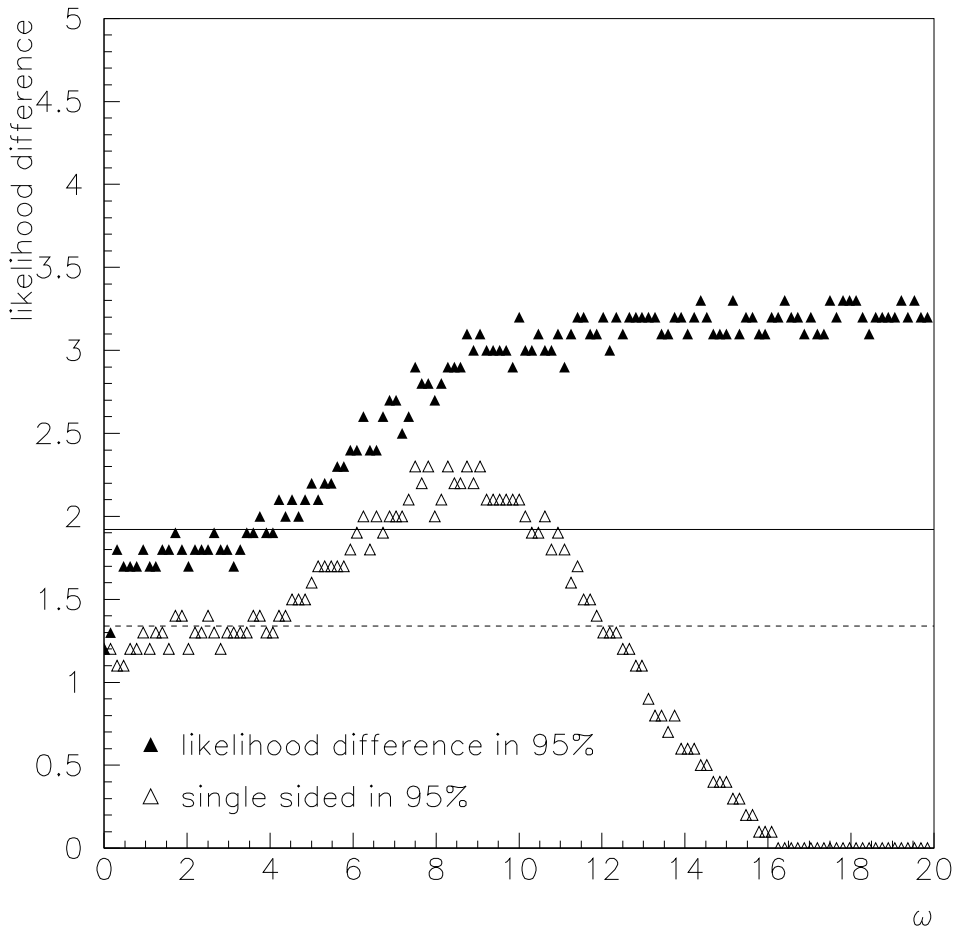

Fig. 6. Calibration of the log-likelihood difference cut, referenced to the minimum as function of the generated frequency ω (95% of the Monte-Carlo experiments have a log-likelihood difference below the cut). Solid triangles correspond to the double-sided definition (see text) and empty triangles to the single-sided one. Only at low frequencies, where the log-likelihood minimum is deep and narrow, the cuts are consistent with the theoretical values (1.92 and 1.34 respectively).

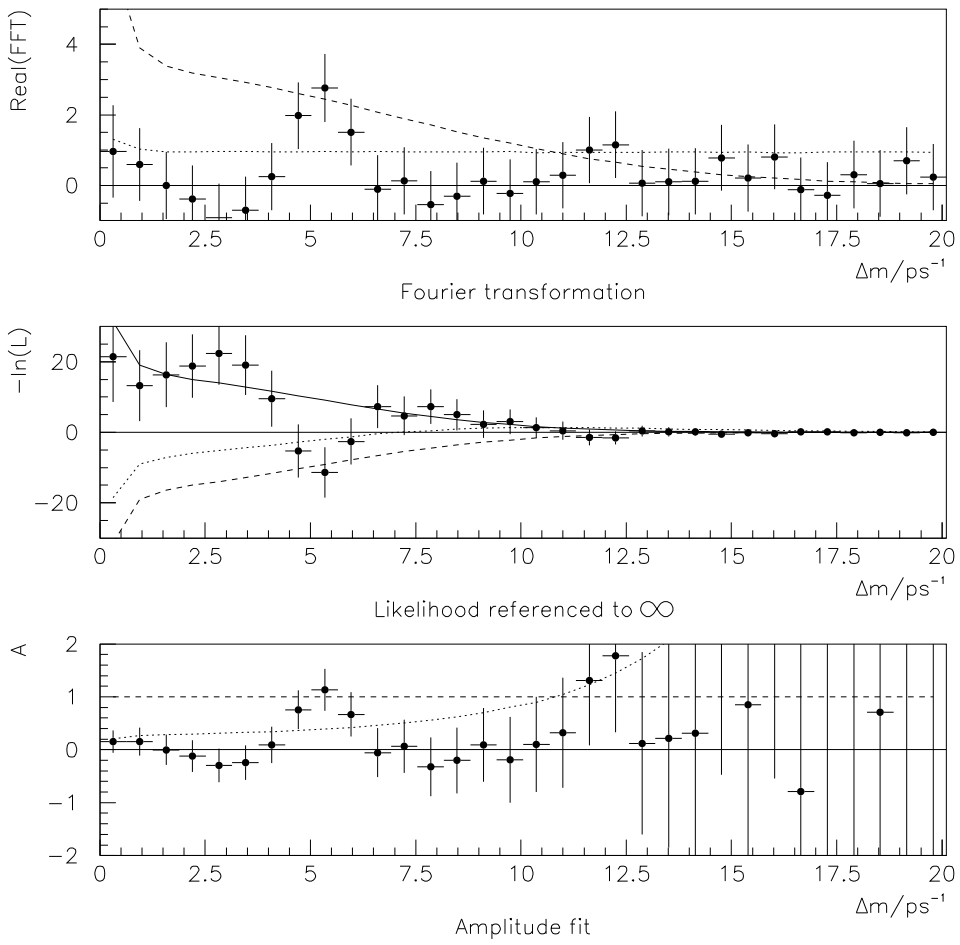

Fig. 7. Analysis of oscillations in Monte Carlo experiments using (from top to bottom) Fourier transformation, likelihood referenced at infinity and amplitude fit. The points represent a given Monte Carlo sample, the error bars indicate the 95% C.L. errors [$\sigma \times 1.645$]. The solid line the average expectation for no oscillations ($\Delta m = \infty$), the dashed line the average peak or minimum position for oscillations at this value of Δm. The dotted line shows the 95% C.L. level contour around the expectation for no oscillations.
Each Monte Carlo experiment generates 10000 events with $f_s = 0.12$, $\eta = 0.18$ and 0.15 ps resolution. The generated frequency is 5.5 ps^{-1}.
All methods show a clear signal at the generated frequency.

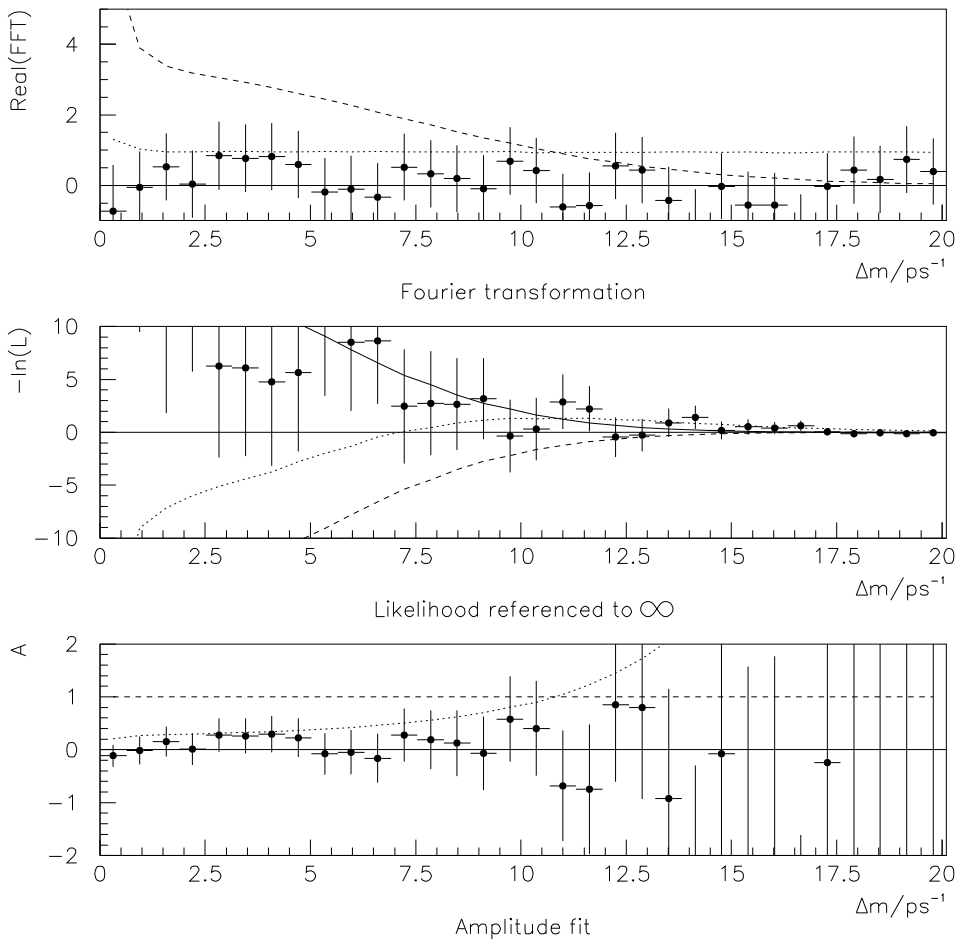

Fig. 8. Analysis of oscillations in Monte Carlo experiments using (from top to bottom) Fourier transformation, likelihood referenced at infinity and amplitude fit. The points represent a given Monte Carlo sample, the error bars indicate the 95% C.L. errors [$\sigma \times 1.645$]. The solid line the average expectation for no oscillations ($\Delta m = \infty$), the dashed line the average peak or minimum position for oscillations at this value of Δm. The dotted line shows the 95% C.L. level contour around the expectation for no oscillations.
Each Monte Carlo experiment generates 10000 events with $f_s = 0.12$, $\eta = 0.18$ and 0.15 ps resolution with a generated frequency $\Delta m = \infty$. All methods are in best agreement with the expectation for no oscillations and give a limit at 10 ps^{-1}.