cdaqs2:~cdaq/gen/Documents/analysis/Hpair_asyms.ps

G_{en} Analysis
Helicity Pair Asymmetries

M. Zeier, May 5, 2000

1 Introduction

During the G_{en} experiment in 1998 the helicity flip frequency was 1 Hz. The helicity pattern was pairwise randomized, i.e. it consisted of a sequence of $+-$ and $-+$ helicity pairs. Scalers were read out every potential helicity transition. In this report asymmetries for each helicity pair are calculated for protons and a particular set of cuts. The asymmetries are then statistically analyzed and a spectral analysis is done to search for hidden oscillations.

2 Calculation of Helicity Pair Asymmetries

- The ntuples were produced in a replay in February 1999. The analyzer did not use the gamma flash procedure yet to determine timing corrections. The coincidence time window was ± 20 ns. Runs 21816 - 22080 (all q^2=0.5) were replayed. In the following the ntuples were processed to deliver proton yields for each helicity packet.
- The following basic cuts were applied:
  $evtype = 3$ Only coincidence events
  $0.85 \leq w \leq 1.025$ Cut in Invariant Mass
  $-45 \leq Ypos \leq 45$ Cut in the Horizontal Position in the neutron detector
- Proton Identification Cut:
  $((hitpat1 > 0)\ OR\ (hitpat2 > 0))\ AND\ (hitpat3 > 0)$

requires that one of the two paddle planes and the first bar plane have fired.

- Four different coincidence time cuts were applied:
  Cut 0: $-2 \leq \mathit{cointime} \leq 2$
  Cut 1: $-2 \leq \mathit{cointime} \leq 0$
  Cut 2: $0 \leq \mathit{cointime} \leq 2$
  Cut 3: $2 \leq \mathit{cointime} \leq 5$

- Asymmetries A_i and statistical errors ΔA_i were calculated for each helicity pair from the resulting proton yields, taking dead time, charge and target polarization into account. A detailed description can be found in appendix A.

- For the calculation of helicity pair asymmetries a rigorous filtering of the helicity packets was done in order to make the calculation more stable and exclude beam-off periods. Specifically packets with pathological deadtimes (>50%) and charge values that deviated from the mean by more than 30% were ignored. The deadtime requirement eliminates most of the helicity packets which are affected by the scaler problem of the earlier runs. (A larger part of the runs between 21816 and 21856 show periods where the scaler channel of the coincidence trigger was not counting, which leads to a wrong dead time calculation.) To see how much the "bad" helicity packets affect the overall asymmetry, the total asymmetry was calculated for the full and filtered data. Table 1 shows the comparison of the results. The difference in asymmetry is several time smaller than the statistical error.

- 140 runs were analyzed, resulting in 93082 helicity pair asymmetries, 52806 (40267) for positive (negative) target polarization. As an example Fig. 1 shows asymmetries with statistical errors for run 21955 and Cut 0.

3 Statistical Analysis

The distribution of A_i's was compared to expectations from statistics in order to detect single runs or run periods that might show an abnormal behaviour. Some basic statistical quantities were calculated and statistical tests were performed for each run and coincidence time cut. Formulas used for the calculation of mean and moments are given in Appendix B. The results are displayed in fig. 3, 4, 5, 6, and the following is a discussion of these figures.

Top right The weighted mean $\bar{A}$ of the Asymmetries A_i, $\min(\mathrm{A}_i)\cdot 10^{-1}$ and $\max(\mathrm{A}_i)\cdot 10^{-1}$ are plotted vs. run number. Negative (positive) $\bar{A}$ corresponds to positive (negative) target polarization. The monotonous increase (decrease) of the maximum (minimum) during run periods reflects the decay of the target polarization. The measured asymmetry is divided by an increasingly small number (see eq. 1), which blows up the spread of the A_i's. This is in particular well visible for the lower statistics of Cut 3. The huge spike in Minimum and Maximum for run 21881 is due to the fact that the target polarization was very low for that run (0.03). Runs with very low target polarization show up after a change in the sign of target polarization, when data was taken while the target was still polarizing up.

Bottom right The variance and higher order moments of the asymmetries are plotted vs. run number to identify runs with highly abnormal distributions. The piecewise increase in variance reflects again the decay of the target polarization and spikes in the variance represent runs with very low target polarization, as explained before.
The interpretation of the values for Squewness and Kurtosis is a delicate business (see Appendix B). Significant deviation of skewness from 0 is observed for early runs and Cut 0 (partially Cut 1 too). In all these runs outliers (see below) are responsible for this behaviour. This is typically just one value per run and its elimination leads to a reduction of the skewness by up to one order of magnitude. For Cut 3 skewness and kurtosis look best, which is due to the low statistics for this cuts. The distribution is so broad that potential outliers can "hide" in the statistical fluctuations.

Top left For each A_i the deviation from the mean in units of the statistical error was calculated: $\frac{A_i - \bar{A}}{\Delta A_i}$
The resulting values were histogrammed with bin size 0.5 (see fig. 2). One expects a normal distribution with $\sigma = 1$, assuming that the statistical error reflects the standard deviation (see discussion in appendix A). In order to test this a χ^2 test was done. The reduced χ^2 plotted vs. run number is in average significantly larger than 1.

Bottom left In a next step the histogram was fitted with a normal distribution leaving σ as the only free parameter (see fig. 2). The plot shows σ and reduced χ^2 vs. run number. Reduced χ^2 is in average 1, whereas σ shows a deviation from 1 which is significant and consistent over the whole run set but varies for the different coincidence time cuts (see table 2). The origin of this bias is not understood. Its cut dependence is correlated with the statistical accuracy of the corresponding cut. For Cut 3 the effect has largely disappeared in the statistical noise.

3.1 Outliers

As outliers we define asymmetry values further away than 4 standard deviations from the mean. Number of outliers for each cut is listed in table 2. The number of outliers decreases with poorer statistics. For Cut 0 (Cut 3) the fraction is larger (smaller) than what is expected for a Gaussian distribution. For Cut 0 a larger part (7) of outliers occurs in the early runs, which is related to the scaler problem mentioned above. Another class of outliers in later runs is related to other weird but also very rare effects which are recognized but not understood. For the other cuts these systematic effects increasingly disappear in the statistical fluctuations. Cut 3 however has too few outliers, which is due to the fact, that the value of the measured asymmetry is limited to ± 1. The statistics for this cut is so low that σ is typically 0.3. Therefore values further away from the mean than 4σ can't occur.
The conclusion here is that the number of outliers is within expectations, taking all the nonstatistical effects into account.

4 Spectral Analysis

A spectral analysis was performed on the helicity pair asymmetries in order to search for possible hidden frequencies beyond statistical noise. A few basic remarks about discrete sampling and spectral analysis follow. For more details refer to *W.H. Press et al: Numerical Recipes* from where the codes for calculation were taken and references therein.

- For Fourier Transforms of discretly sampled data a limit in frequency is given by the sampling rate Δ. The critical frequency is called Nyquist frequency and given as
$$f_c = \frac{1}{2\Delta}$$
In our case Δ is 2 seconds and thus $f_c = 0.25$. The accessible frequency range for a spectral analysis of the helicity pair asymmetries is 0 - 0.25 Hz.
A band-width limited function with $f<f_c$ is completely determined by sampling with the corresponding Δ. If the function is not band-width limited to f_c an effect called *aliasing* occurs. The frequency content beyond f_c is falsely translated into the range $0<f<f_c$. The only way to control *aliasing* is to know the band-width of the function and to adjust the sampling frequency accordingly. In our case frequency contributions from beyond f_c can not be excluded and there is nothing you can do about it.

- Discrete Fourier Transform maps N points in the time domain $A(t_k)$ to N points in the frequency domain $\breve{A}(f_k)$ equally spaced from $-f_c$ to f_c. For real data the following relation holds: $\breve{A}(-f_k) = [\breve{A}(f_k)]^*$. Thus we can restrict ourselves to the frequency region $[0..f_c]$ with (N/2 + 1) frequencies. For the spectral analysis we would like to know the power of each frequency:
$$P(f_k) \propto \mid \breve{A}(f_k) \mid^2 \quad ; \quad f_k = \frac{2k}{N} f_c \quad ; \quad k = 0, 1, ..., \frac{N}{2}$$

- The fact that the Fourier Transform is applied to data sampled over a finite time interval leads to an effect called frequency leakage. This can be largely suppressed by a technique called "Data Windowing". The data is multiplied with a window function, in our case with the Parzen window.

- The standard deviation for the estimated power of each frequency is, independent of the number of data points, always 100%. To reduce that, the data set can be divided into segments. The power estimation is done for each segment seperately and finally averaged. For K segments the standard deviation reduces by a factor $\sqrt{K}$. Overlapping of the segments by half doubles the number of segments and leads to an even stronger reduction of the standard deviation by $\simeq \sqrt{\frac{9}{11}K}$. Reduction of the standard deviation goes at the expense of loss in frequency resolution.

Fig. 7 and 8 show the result of the spectral analysis for positive and negative target polarization and for each coincidence time cut. The data was divided into overlapping segments in order to reduce the standard deviation. There is no distinct signal present which shows up for all coincidence time cuts consistently, neither for the positive nor the negative target polarization. Repeating the analysis for different segmentations and frequency resolution does not change the picture.

To give an idea of the sensitivity of this method, artificially a sine curve was superimposed to the data with frequency 0.1 and amplitude 0.05. This would roughly cause the broadening of the distribution of helicity pair asymmetries as observed in the statistical analysis (see table 2). There is no way to recognize this effect by eye from the time series directly (see fig 9), but the power spectrum reveals a clear peak at 0.1 Hz at least for coincidence time Cut 0 (see fig. 10). However one should keep in mind that this example here is the ideal case of a harmonic function of sharp frequency with no discontinuities. Real life frequency jitters or phase jumps could reduce the detectability of the effect.

5 Conclusion

From the statistical analysis of helicity pair asymmetries we see no abnormal behaviour of single runs or single run periods, except for the period of early runs, which has a scaler problem. Even if the influence on the whole data set

is neglible, it is recommended for runs 21818 to 21856 not to use the coincidence scaler values (e$*$B trigger and pretrigger) to calculate the deadtime, but rather use the HMS Single scalers.
The number of single outliers (more than 4 σ away from the mean) is in agreement with expectations from statistics, if one takes nonstatistical effects into account.
The distribution of helicity pair asymmetries is perfectly Gaussian for all coincidence time cuts. However the standard deviation of the distribution is by a few percent larger than what one expects. The absolut value of this bias is in the same order as the overall asymmetry ($\simeq$0.05).
A spectral analysis excludes the possibility of a monoharmonic contamination of the asymmetries of this size to a certain degree and reveals no further insight into the problem.

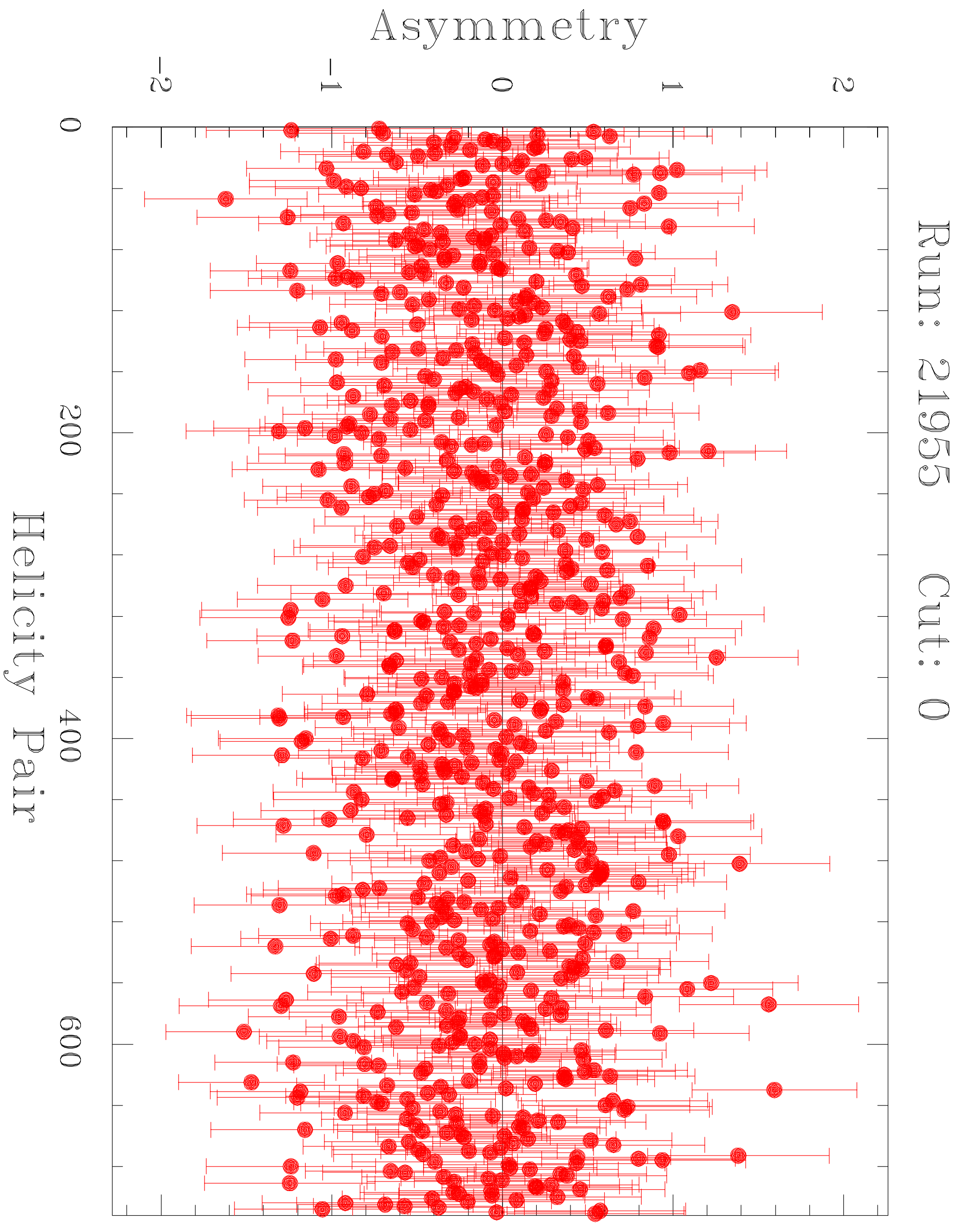


Figure 1: Proton asymmetries vs. helicity pair for run 21955 and Cut 0.

Cut	T_p	Total Asymmetries $\cdot 10^{-2}$			
		Filtered	All	Stat. Error	Filtered - All
0	+	-7.18	-7.13	0.22	-0.05
1	+	-7.15	-7.10	0.29	-0.05
2	+	-7.23	-7.18	0.35	-0.05
3	+	-5.15	-5.19	0.61	0.04
0	-	6.20	6.16	0.27	0.04
1	-	7.58	7.56	0.35	0.02
2	-	4.16	4.10	0.42	0.06
3	-	2.17	2.17	0.73	<0.01

Table 1: Comparison of total asymmetries for the filtered (see text) and unfiltered data sets for different coincidence time cuts and target polarization (T_p)

	σ (fitted)	Outliers ($> 4\sigma$)	Fraction
Cut 0	1.07	20	$2.1 \cdot 10^{-4}$
Cut 1	1.04	9	$9.7 \cdot 10^{-5}$
Cut 2	1.03	4	$4.3 \cdot 10^{-5}$
Cut 3	1.01	0	0
Expected (Gaussian)	1.00	$\simeq 6$	$6.3 \cdot 10^{-5}$

Table 2: For each coincidence time cut the following values are listed characterizing the distribution of helicity pair asymmetries: Average fitted standard deviation σ in units of the statistical error, number of outliers ($> 4\sigma$ away from mean) and their fraction with respect to the total number of helicity pairs. The last row shows corresponding values as expected from a normal distribution.

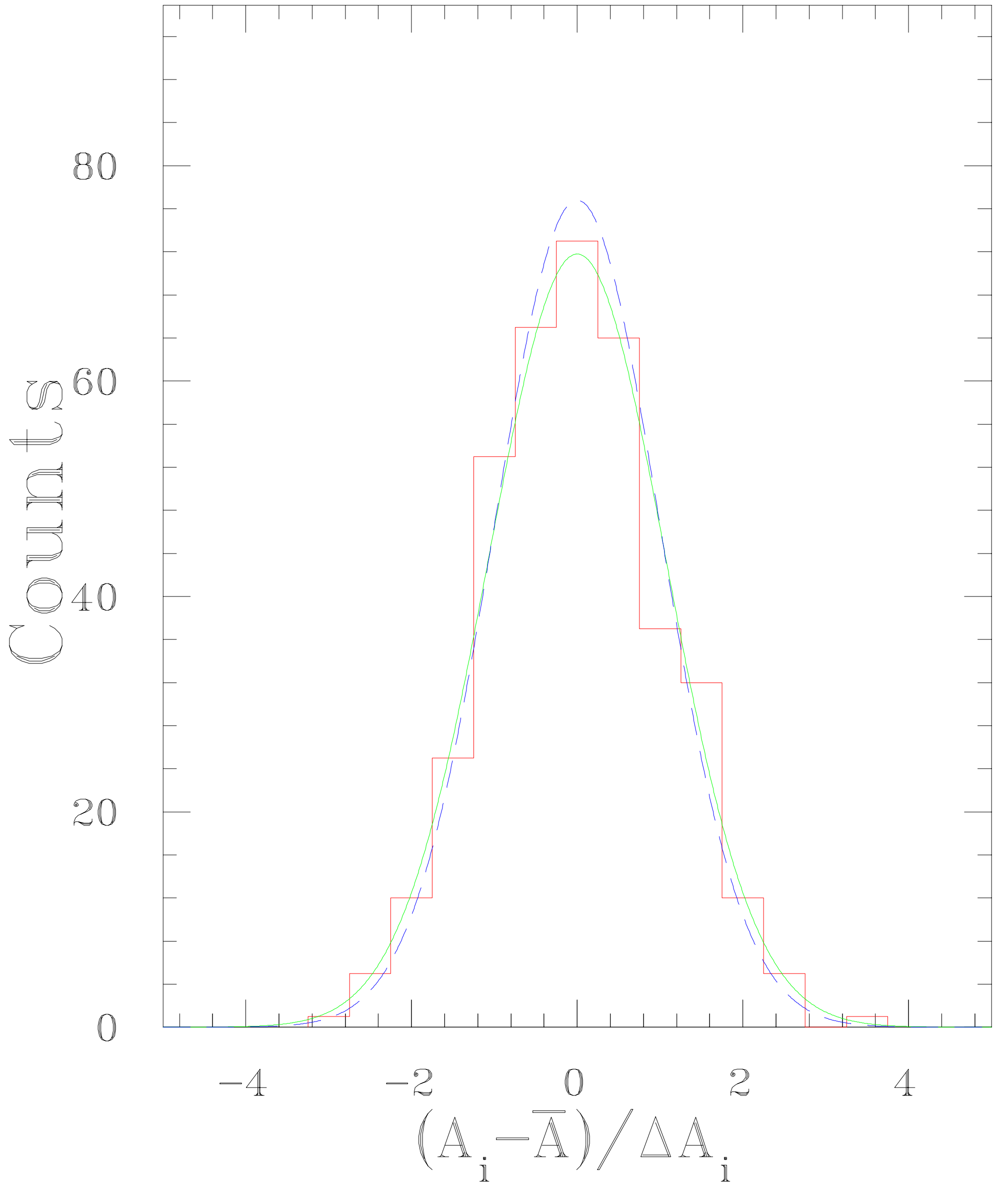


Figure 2: Histogrammed deviation of A_i from the mean in units of the statistical error for run 22057 and Cut 0. The blue dashed curve represents the expected normal distribution with $\sigma = 1$. The green curve is a fit of a normal distribution to the histogram with σ as the only variable parameter. In this case σ was 1.07 and the reduced χ^2 was 1.0

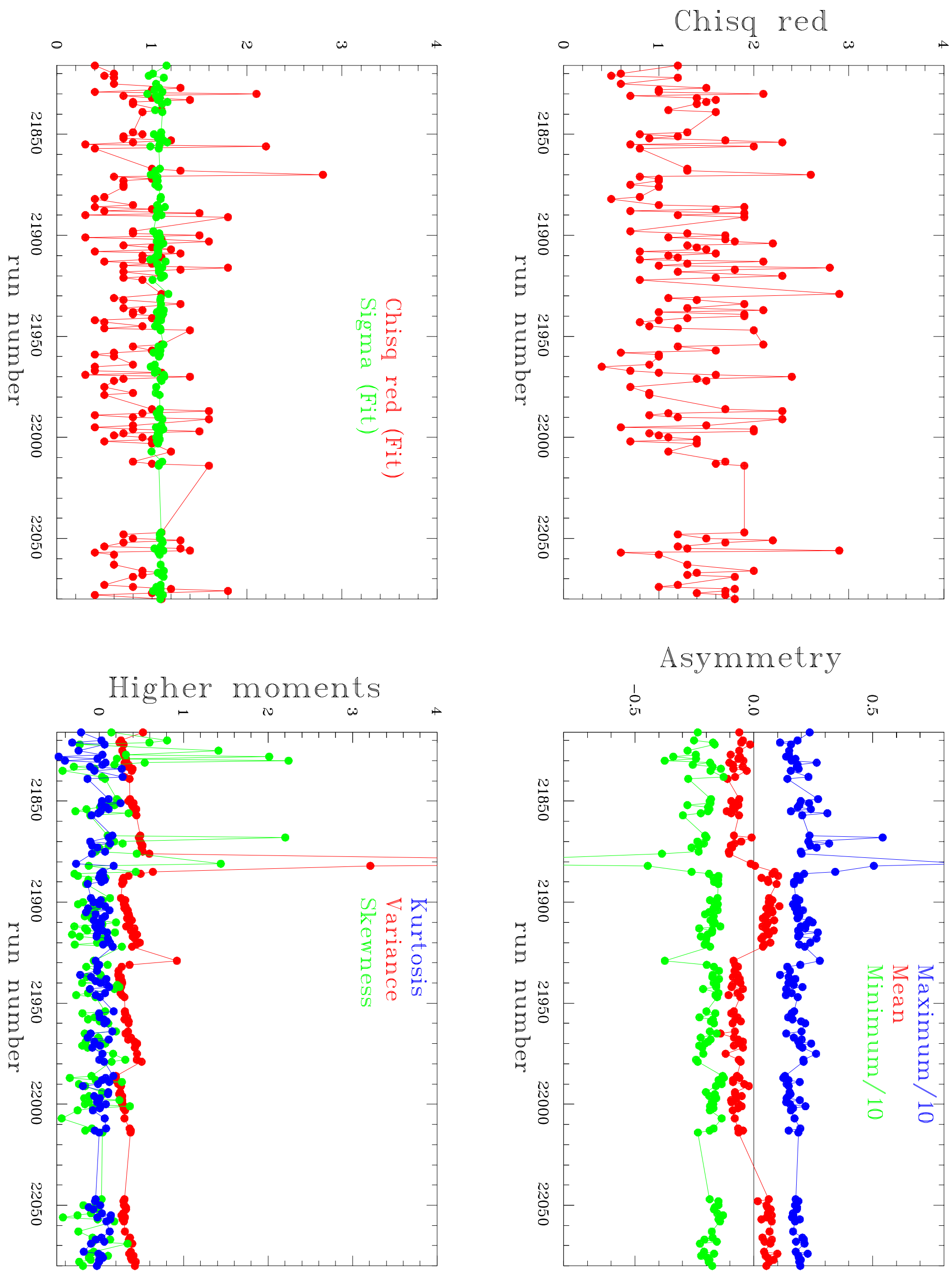


Figure 3: Statistics for Cut 0.

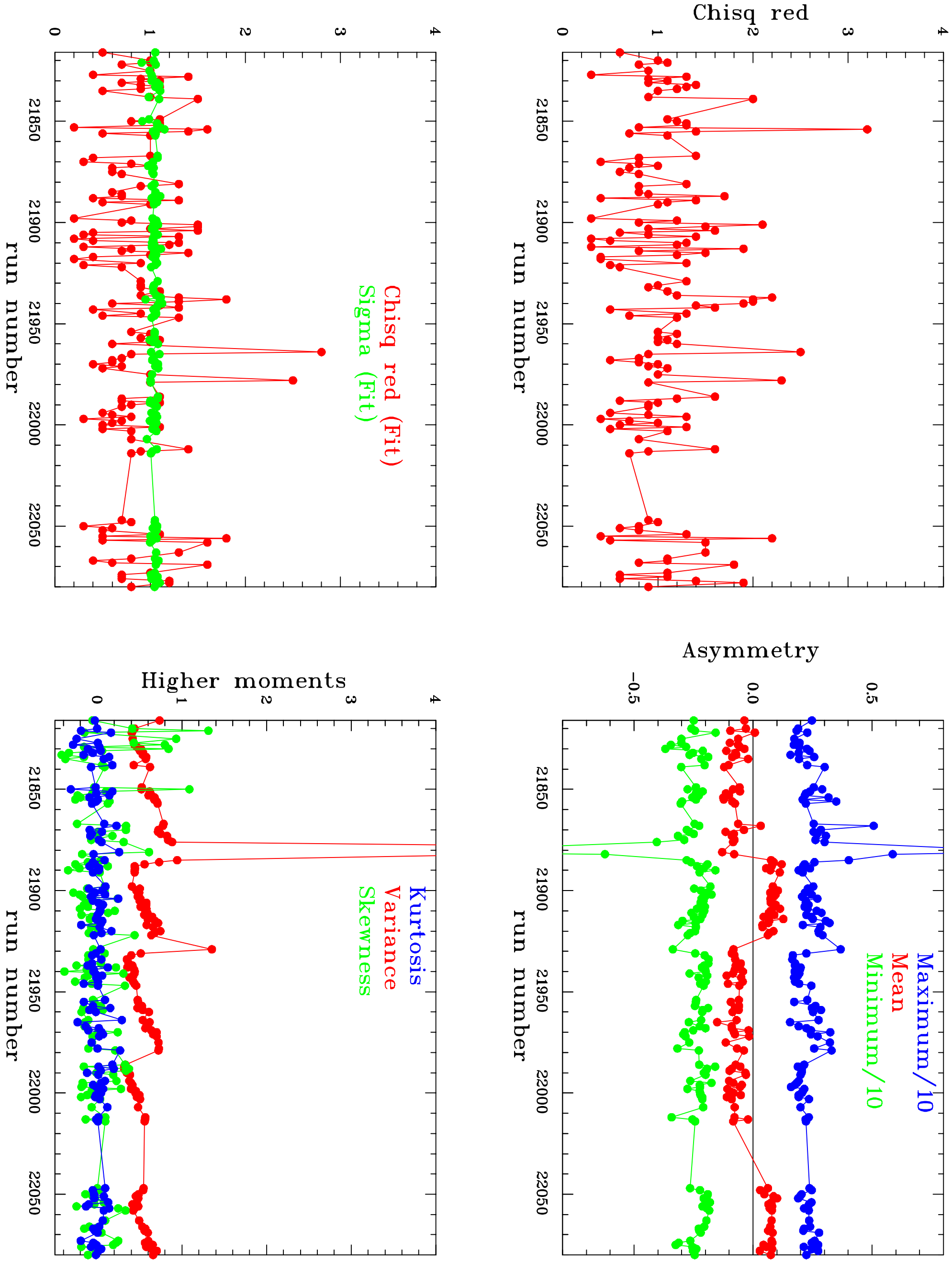


Figure 4: Statistics for Cut 1.

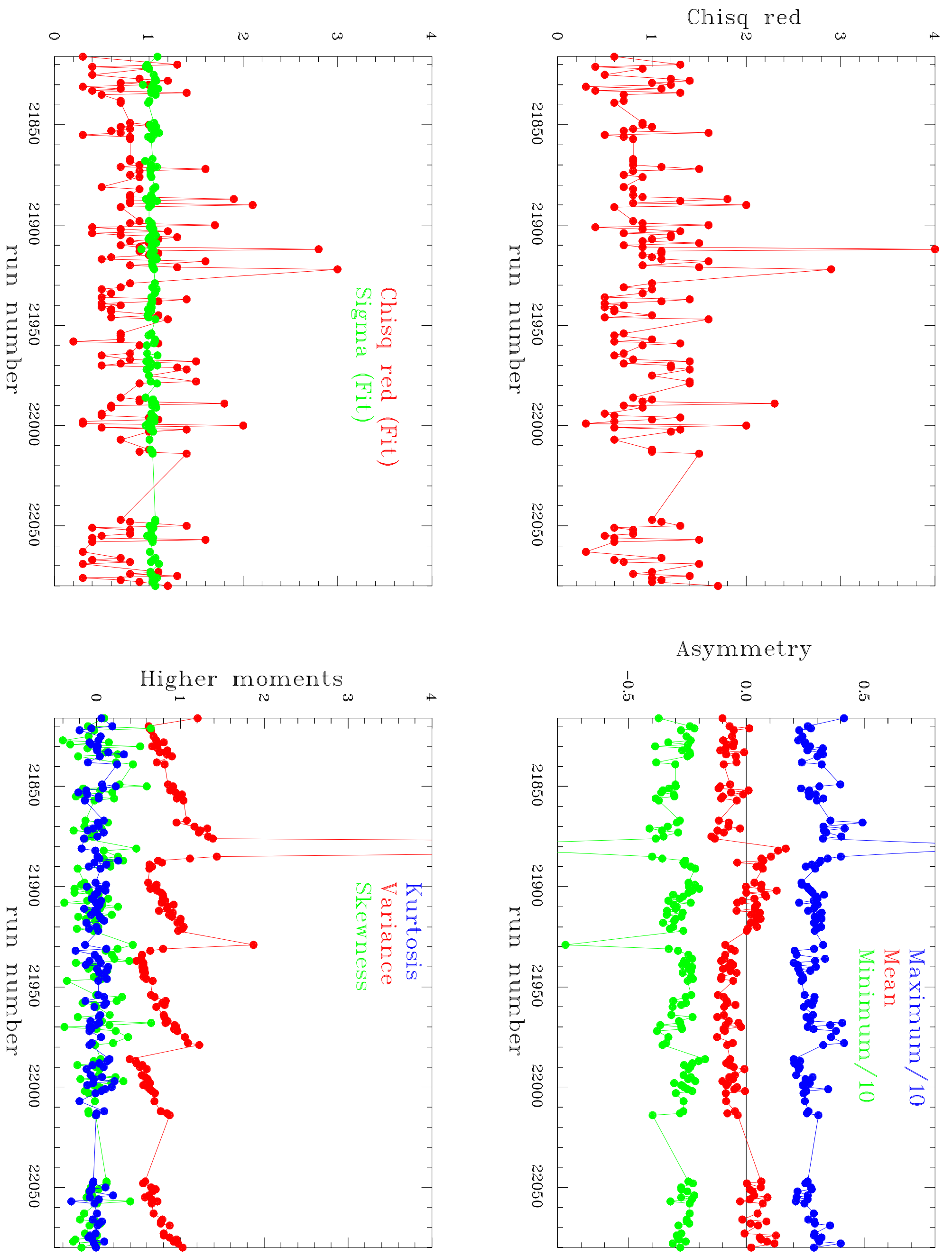


Figure 5: Statistics for Cut 2.

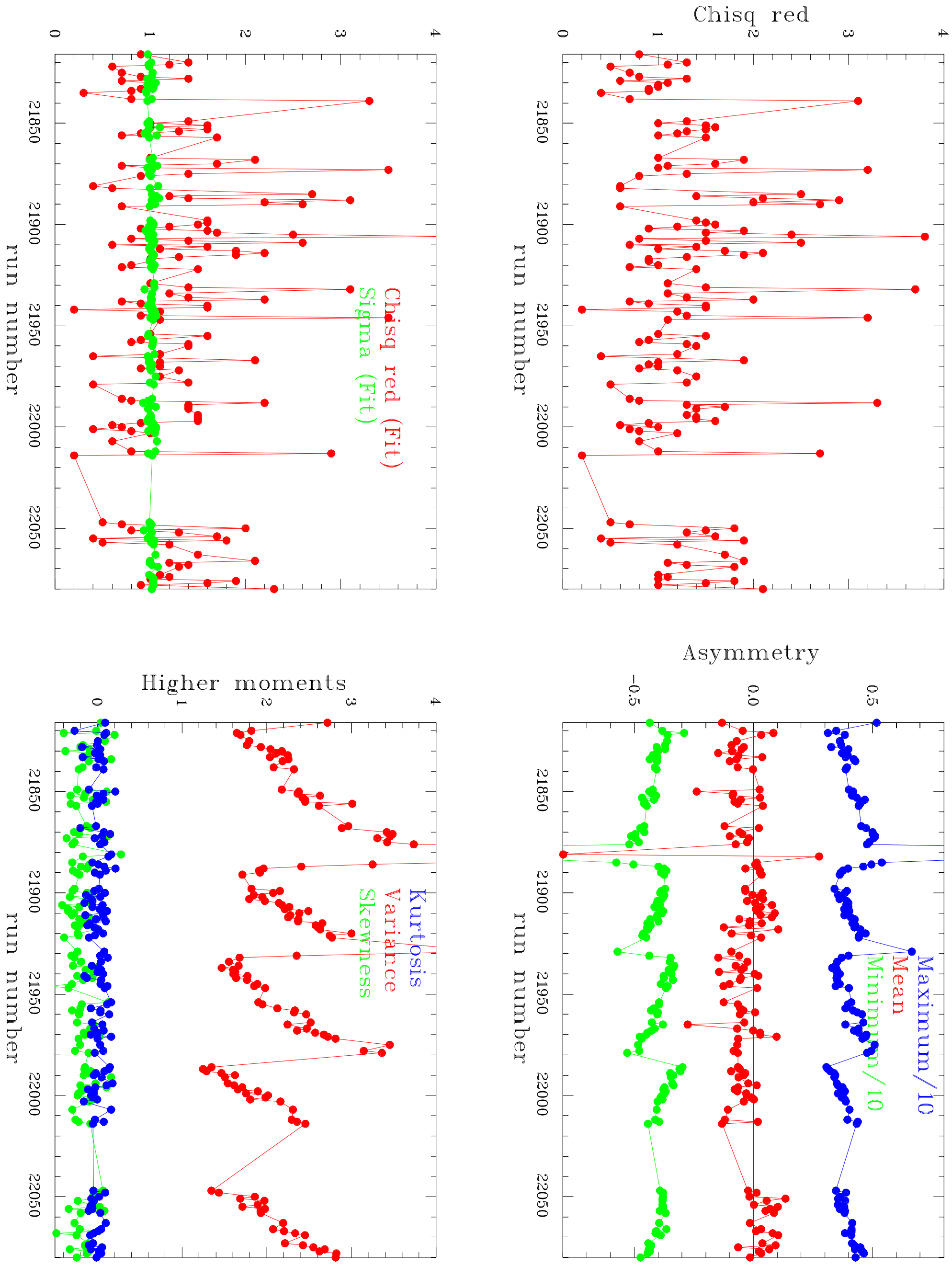


Figure 6: Statistics for Cut 3.

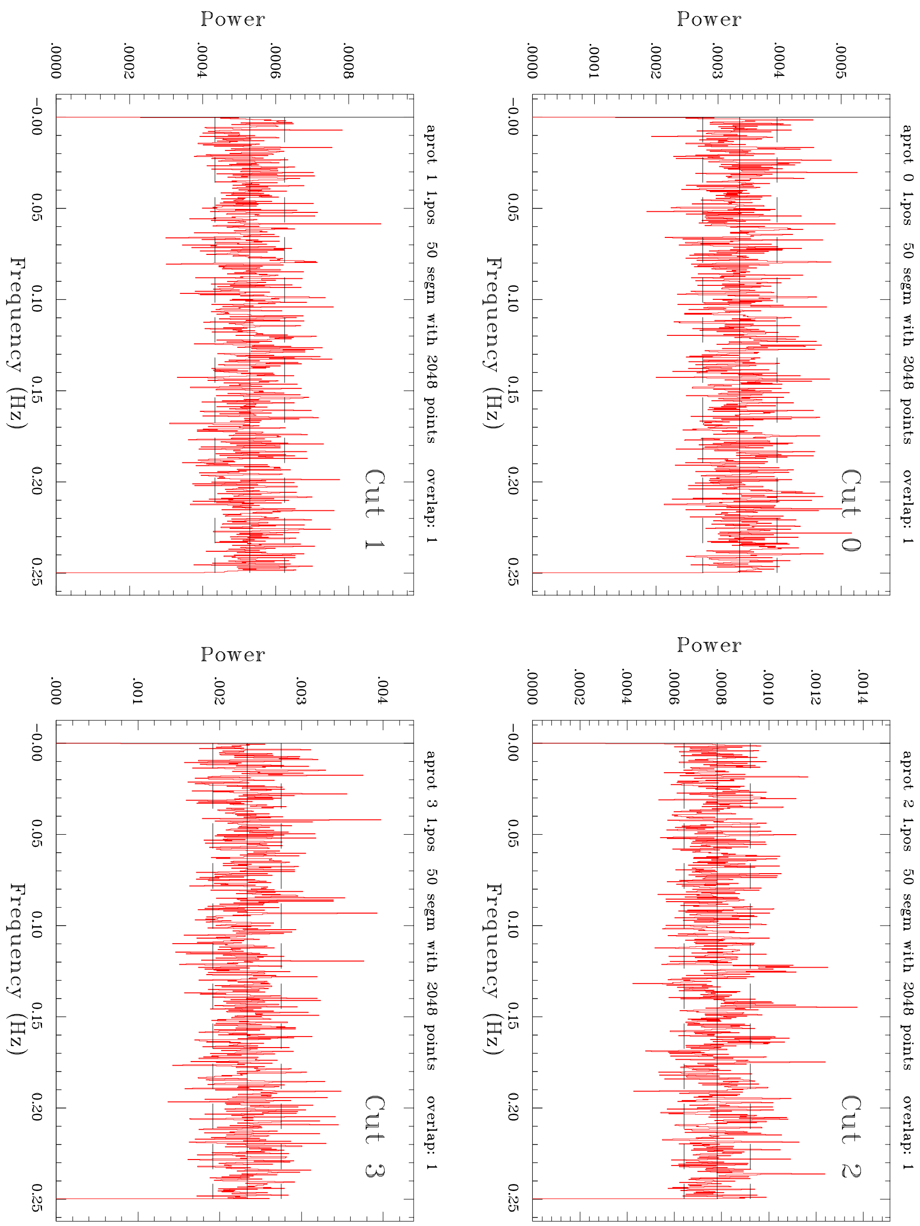


Figure 7: Power spectrum of the helicity pair asymmetries for the four different coincidence time cuts (Cut 0 to Cut 3) and for positive target polarization. The data was divided into 50 overlapping segments with 2048 data points each. The solid line indicates the average power and the area bounded by the dashed lines is the region of one standard deviation.

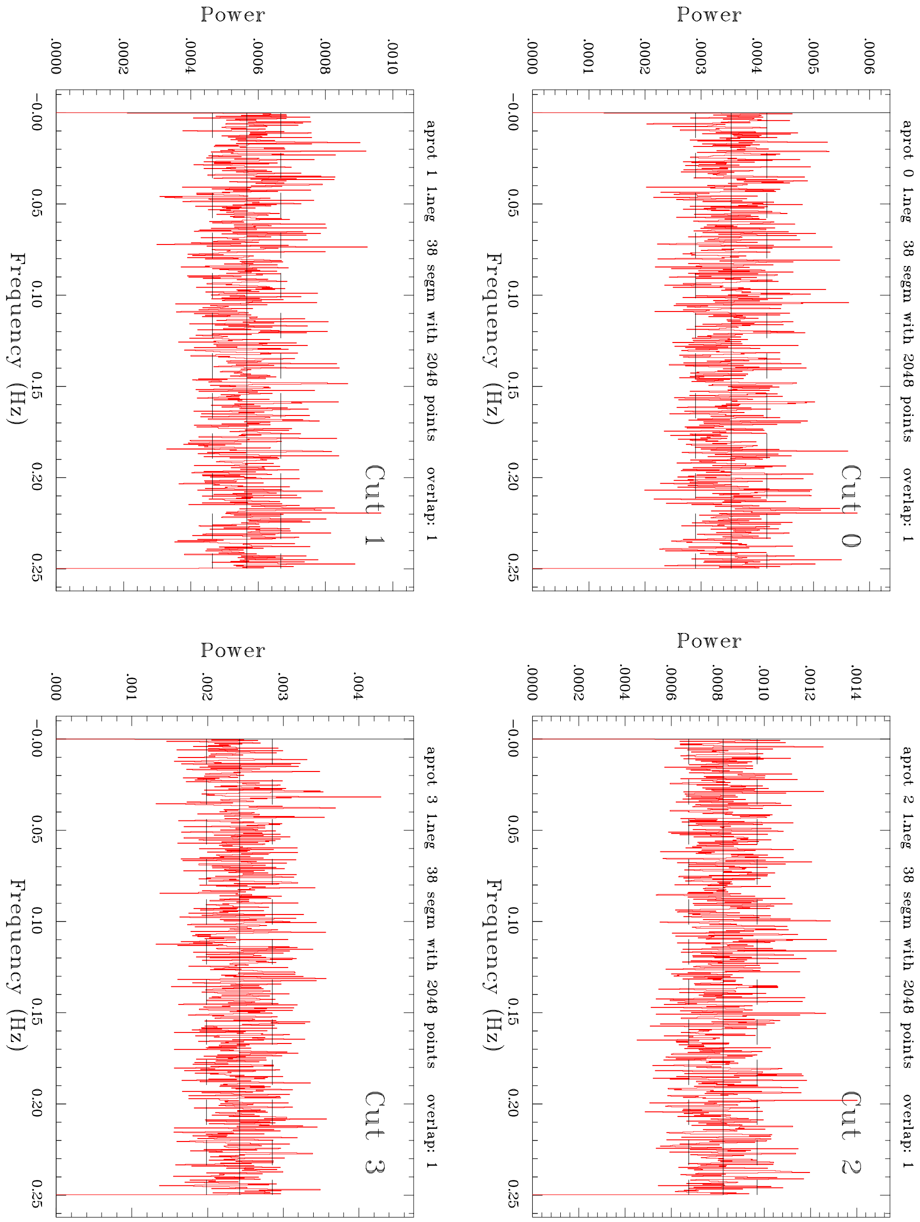
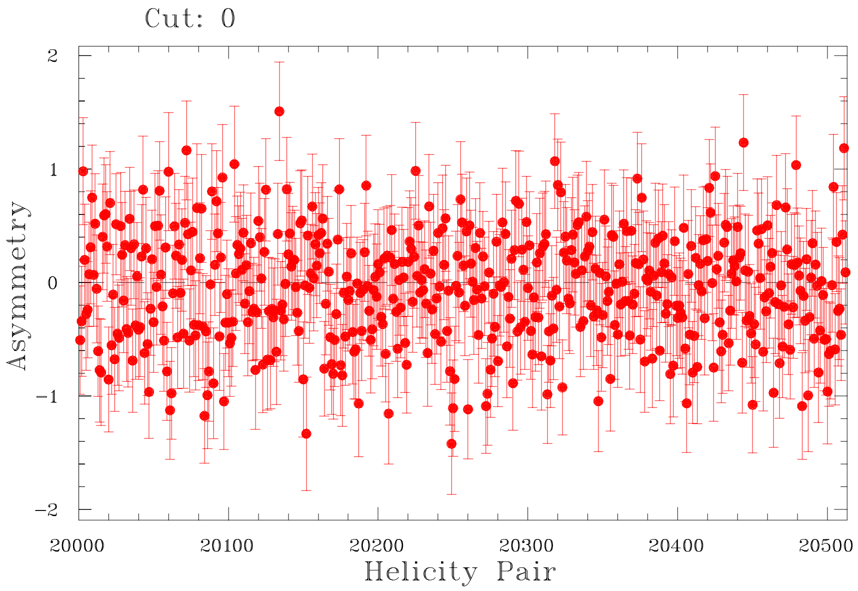


Figure 8: Power spectrum of the helicity pair asymmetries for the four different coincidence time cuts (Cut 0 to Cut 3) and for negative target polarization. The data was divided into 38 overlapping segments with 2048 data points each. The solid line indicates the average power and the area bounded by the dashed lines is the region of one standard deviation.



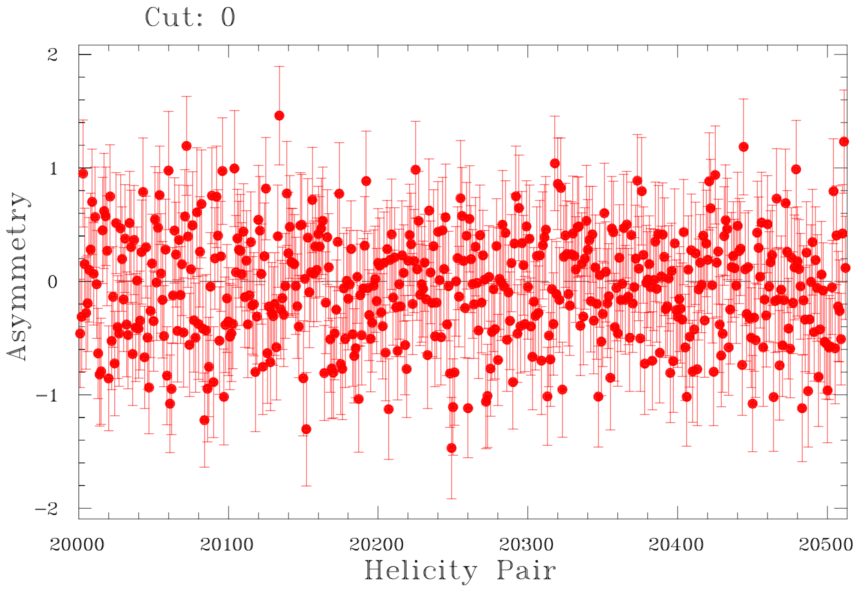


Figure 9: The same arbitrarily chosen sequence of helicity pair asymmetries (coincidence time cut 0) twice, one of them superimposed with a sine function of 0.1 Hz frequency and amplitude of 0.05. By eye it is not possible to recognize which one is biased.

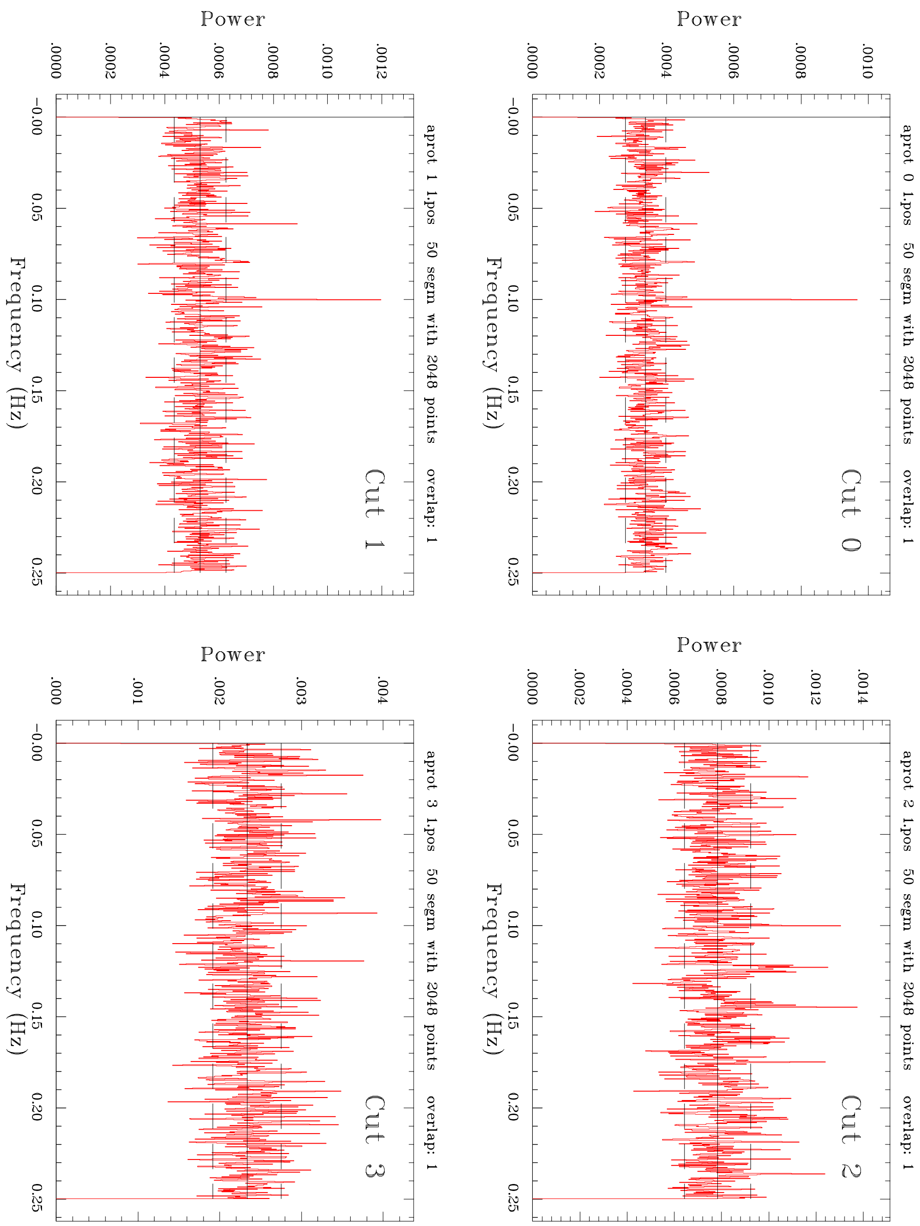


Figure 10: Power spectrum (analog to fig. 7) of helicity pair asymmetries biased by a sine function of 0.1 Hz frequency and amplitude of 0.05 (see fig. 9).

A Calculation of the asymmetry and its statistical error

From yields ($N^{\pm}$) asymmetries (A) were calculated for each helicity pair, taking life time (LT), charge (Q) and target polarization (T_p) into account:

$$A = \frac{N^- - rN^+}{N^- + rN^+} \cdot \frac{1}{\mid T_p \mid} \quad ; \quad r = \frac{Q^- \cdot LT^-}{Q^+ \cdot LT^+} \tag{1}$$

Life Time and charge are calculated from scaler values (S):

$$LT = \frac{S_{e*Btrigger}}{S_{e*Bpretrigger}} \quad ; \quad Q = Gain_{BCM1}(S_{BCM1} - Offset_{BCM1} \cdot S_{Time})$$

The statistical error (ΔA) is given as

$$\Delta A = \sqrt{\frac{r(N^- + N^+)}{(N^- + rN^+)^2}(1 - A^2)} \cdot \frac{1}{\mid T_p \mid} \tag{2}$$

For $r = 1$ this simplifies to

$$\Delta A = \sqrt{\frac{1}{(N^- + N^+)}(1 - A^2)} \cdot \frac{1}{\mid T_p \mid}$$

This form is shown just to make it easier to recognize that the error in A is given essentially by the number of counts but additionally attenuated by A. However this introduces a certain subtlety if we want to use this error as the standard deviation in order to see if the distribution of the asymmetries is normal and has the expected width. The asymmetries are expected to distribute around the true asymmetry with a standard deviation which is given by eq. 2 but with the true value of A in it. The true value of A is small in our case. However the measured asymmetry of the single helicity pair can become large due to the low statistics. This is in particular true for Cut 3 with only a few counts per helicity packet. Using the measured A in eq. A thus leads to an underestimation of the standard deviation. The fact that the true A is small ($\simeq 0.02$) and the influence in eq. A thus marginal ($< 10^{-3}$) legitimates us to calculate the error of A as

$$\Delta A = \sqrt{\frac{r(N^- + N^+)}{(N^- + rN^+)^2}} \cdot \frac{1}{\mid T_p \mid} \tag{3}$$

(which is eq. 2 with $A = 0$) in order to avoid underestimation of the standard deviation.

B Calculation of moments

Each run has N values of A_i with errors ΔA_i.
All calculated moments are weighted with w_i:

$$w_i = \frac{1}{\Delta A^2} \qquad W = \sum_{i=1}^{N} w_i$$

Mean:

$$\bar{A} = \frac{1}{W} \sum_{i=1}^{N} w_i A_i$$

Variance:

$$Var(A) = \frac{N}{W(N-1)} \sum_{i=1}^{N} w_i (A_i - \bar{A})^2$$

Standard deviation:

$$\sigma = \sqrt{Var(A)}$$

Squewness:

$$Skew(A) = \frac{1}{W} \sum_{i=1}^{N} w_i \left(\frac{A_i - \bar{A}}{\sigma} \right)^3$$

Kurtosis:

$$Kurt(A) = \left[\frac{1}{W} \sum_{i=1}^{N} w_i \left(\frac{A_i - \bar{A}}{\sigma} \right)^4 \right] - 3$$

Squewness characterizes the degree of asymmetry of a distribution around its mean. It is 0 for a normal distribution. A positive (negative) Squewness indicates an asymmetric tail extending out towards more positive (negative) values.
Kurtosis characterizes the peakedness or flatness of a distribution relative to the normal distribution. It is 0 for a normal distribution, positive for the outline of the Matterhorn and negative for the outline of the excrements of a cow.
In real life squewness and kurtosis will never be 0, even if the distribution

is Gaussian. In order to be useful estimators, we need to have an idea of their standard deviations, which unfortunately depend rather critically on the shape of the underlying distribution. For a normal distribution the standard deviation of the Skewness is $\simeq \sqrt{6/N}$ and for the kurtosis $\simeq \sqrt{24/N}$. Thus in our case for a typical run with a few hundred data points the standard deviation of the squewness (kurtosis) is expected to be $\simeq$0.1 (0.2). In practical applications one should only start to believe in a deviation from the gaussian distribution if Squewness or Kurtosis are several or many times these values.