$\boxed{\mathcal{AIPS}\ \text{Memo 101}}$

The Calculation of SNR in KRING's FFT stage

Ketan M. Desai

October 8, 1998

Abstract

This memo describes the SNR calculation during the FFT stage of KRING. I describe the assumed statistical properties of the visibilities used for the fringe search and show how KRING's SNR estimate differs from that in FRING. I also discuss the probability of false detection, a statistic users have often requested be reported while fringe-fitting.

1 Introduction

The estimation of residual phase variations in VLBI data is of great practical importance. Residual phase variations occur for a variety of reasons including antenna and source position errors in the correlator model, source structure, ionospheric and atmospheric effects as well as thermal noise. Linear residual variations of phase with frequency and time are called residual delays and rates. Fringe-searching or fringe-fitting is the process of estimating residual delays and rates from time/frequency data, or dynamic spectra. Schwab and Cotton[1] proposed a two stage algorithm to construct the best possible estimate of the residual phase variations using all available information. In the first stage, dynamic spectra are Fast Fourier Transformed one baseline at a time to determine an initial antenna-based model of the phase variations[2]; this is the fringe detection stage. In the second stage, a Least Squares Fit is performed using data from all baselines to simultaneously refine all model parameters; this is the fringe refinement stage. The fringe detection stage is the subject of this memo.

An important component of the Schwab & Cotton algorithm is the SNR estimate of the fringe detection. The SNR estimate in FRING is based upon simulations by B. Cotton. In the fringe-detection stage, after Fourier Transforming the dynamic spectra, peaks in the resultant delay-rate spectrum are 'searched' to find the one with the greatest amplitude. If the SNR estimated using this peak amplitude exceeds some threshold, the residual phase, rate, and delay are accepted as part of the initial model. The correctness of the SNR computation is important because it is used to determine whether or not to accept the search results for a given baseline. If the computation incorrectly reports a high SNR, erroneous model values may be passed to the fringe refinement stage. If the computation incorrrectly reports a low SNR, acceptable model values may be discarded, forcing more baselines to be searched, increasing computation time. If the computation systematically reports low SNRs, the FFT stage may fail to pass model values for some antennas, resulting in lost data. The common wisdom is that the present calculation in FRING is approximately correct at low SNRs but is an underestimate at high SNR. I present a formal analysis of SNR estimation in section 2. I will show that the common wisdom is incorrect at low SNR, this is illustrated by example in section 2. There are three implementations of the Schwab and Cotton algorithm in $\mathcal{AIPS}$, FRING, BLING/BLAPP, and KRING. The analysis in this memo has been implemented in KRING to compute SNRs during the fringe detection stage.

[1] Schwab, F. & Cotton, B., " Global Fringe-Fitting Search Techniques for VLBI", AJ, 688, **88**, 1983.
[2] See the note about baseline stacking in the next section.

The time/frequency sampling rate of a dynamic spectrum determines the delay/rate range of its FFT; the exact relation between the sampling rate and the corresponding range is determined by the Nyquist theorem. The delay-rate peak is the point with the largest amplitude inside a search window, or range of delays and rates, specified by the user. If too large a search window is specified, a noise spike may be found that, by chance, has amplitude greater than the true delay-rate peak; this is a false detection. If too small a search window is specified, the true delay-rate peak may lie outside the search window; any point found within the search window will be a false detection. The probability of false detection is of interest for deciding whether or not the search windows need to be made smaller or increased; it is not reported by FRING, BLING, or KRING. Although only a small fraction of the full Nyquist range may have been specified for the search window, FRING and BLING still must calculate full-size FFTs while KRING actually performs smaller FFTs when smaller search windows are given, decreasing computation time[3]. I present a brief analysis of the probability of false detection in section 3.

1.1 Assumptions

Signal processing of the voltages received at each antenna can, in principle, introduce biases in the visibility amplitude and more importantly in the visibility phase. Post-correlation processing can correct for amplitude and phase biases but cannot recover from all associated sensitivity losses. I assume that all amplitude and phase biases have been removed, leaving only a bias due to an additive noise component, before fringe-searching. This is a common assumption when analyzing visibility data[4]. I also assume that any sensitivity losses are properly incorporated into the weights so that each weight properly reflects the noise in its associated measurement.

The dynamic spectrum produced by the correlator and presented for fringe-searching includes amplitude and phase variations due to source structure as well as non-linear phase variations. Schwab and Cotton offered three alternative approaches to the nominal Least Squares formulation of the fringe search problem in which the model and the data are altered for reasons of efficiency or convenience. In this memo, I compare the standard pre-Cotton and Schwab method and the method actually implemented in FRING, both of which are described in the next section.

Generally, analyses of fringe-searching techniques assume that the source is a point source. This is, of course, usually not true. If the data are divided by a good model of the source, the resultant visibilities will reflect a source with unit strength at the phase center, *i.e.* a point source. I assume that such a model is available and has been divided out of the data. It must be remembered however that the model division distorts the noise distribution, magnifying the noise wherever the model is of low amplitude. If during the model division phase, the weights are also divided by the square of the model amplitude, this procedure will not bias the location of the peak in delay/rate space. The weights are not modified when model division is peformed in FRING or KRING. I do not address the effect of this omission in this memo.

Schwab and Cotton introduced stacking, a technique for increasing sensitivity when fringe-searching weak sources. The analysis in this memo does not address the subtleties introduced by stacking.

Besides estimating the residual delay and rate in the data, it is of interest to determine the signal-to-noise ratio of the final estimate and the probability of false detection. The latter is more easily calculated if it is assumed that the true residual delays and rates are actually zero, *i.e.* there are no residual phase variations present in the input visibilities. Note that as long as this assumption is not used in the fringe-searching procedure, it does not result in any loss of generality for the final results. I therefore assume that the true residual delay and rate in the data are both zero.

[3] In KRING, If the delay search window is smaller than the rate search window, the frequency axis is first FFT'd to delay, then only those times that lie within the the delay search window are FFT'd into rates. If the rate search window is smaller than the delay search, the order of the FFTs is reversed appropriately. This offers tremendous savings in execution time when the search is over a restricted range of delay/ rate space.

[4] see *e.g.* chapter 6 of Thompson, Moran, and Swenson, Interferometry and Synthesis in Radio Astronomy, Krieger Publishing Co., 1991

All the analyses in this memo involve averaging random quantities that are either Gaussianly distributed or satisfy the Central Limit Theorem criteria. I assume that the averaged quantities are Gaussianly distributed and so can be described entirely using the first and second moments.

In summary, I analyze the estimation of SNR and the probability of false detection in this memo using the following assumptions:

- The visibilities are well behaved; besides additive noise, the data is bias-free in amplitude and phase.
- The weights attached to each visibility properly reflect the noise associated with each measurement.
- The data reflects a point source. Any source structure has been pre-divided out of the visibility data.
- The random variables analyzed in this memo are completely characterized by their first and second moments.

1.2 Notation

The complex visibility function, V, is sampled at discrete times and frequencies and is modelled as a constant real signal S plus complex noise $n_r + jn_i$. The former is independent of time and frequency while the latter is independent between one time/frequency sample and others.

This assumes that any source structure has been divided out, as discussed in the last section. The complex noise is distributed as a Gaussian with half-width at half-max $\sigma\sqrt{\ln 8}$ in its real and imaginary parts.

The probability density distributions for the real and imaginary part of V are Gaussians centered upon S and zero respectively. The probability density distribution for the phase is obtained by integrating over all possible signal amplitudes. Here are the probability distributions for the real part, imaginary part, and phase of the complex visibilities:

$$p_{V_r}(r) = \frac{1}{\sqrt{2\pi\sigma^2}} e^{-\frac{(r-S)^2}{2\sigma^2}} \qquad p_{V_i}(i) = \frac{1}{\sqrt{2\pi\sigma^2}} e^{-\frac{i^2}{2\sigma^2}} \qquad p_\Phi(\phi) = \frac{e^{-\frac{1}{2}k^2}}{2\pi} \int_0^\infty x dx e^{-\frac{1}{2}x^2 + xk cos\phi}$$

Throughout this memo, $k = S/\sigma$ is the signal-to-noise ratio *for an individual visibility.* In some places, $\gamma = k^2/4$ is used for convenience.

I use the 'symmetric' form of the Fourier transform in this memo

$$M_U(\zeta) = \langle e^{ju\zeta} \rangle = \int_u p_U(u) e^{i2\pi u\zeta} du$$

except in section 2 where, to avoid a plethora of '2π's, I use the 'asymmetric' form:

$$M_U(\zeta) = \langle e^{ju\zeta} \rangle = \int_u p_U(u) e^{iu\zeta} du$$

The pre-Schwab and Cotton approach was to fringe-search the visibilities directly. I denote by $Y_{\omega\tau}$ the elements of the Fourier Transform of the visibilities.

$$Y_{\omega\tau} \equiv \frac{1}{TF} \sum_{\substack{f=1,F \\ t=1,T}} w_{tf} V_{tf} e^{2\pi j f\tau/D + 2\pi j t\omega/R}$$

The sums here are over T time and F frequency points, ω takes rate values between 0 and $R-1$, and τ takes delay values between 0 and $D-1$. The weights are denoted by w_{tf}. The first and second moments, and variance of w_{tf} are denoted by $\langle w \rangle$, $\langle w^2 \rangle$, and $\sigma_w^2 = \langle w^2 \rangle - \langle w \rangle^2$.

The method actually implemented by Schwab and Cotton in FRING uses a normalized form of the visibilities. I denote by $X_{\omega\tau}$ the elements of the Fourier Transform of the normalized visibilities according to the following formula.

$$X_{\omega\tau} \equiv \frac{1}{TF}\sum_{\substack{f=1,F\\t=1,T}} w_{tf}\frac{V_{tf}}{|V_{tf}|}e^{2\pi jf\tau/D+2\pi jt\omega/R} = \frac{1}{TF}\sum_{\substack{f=1,F\\t=1,T}} w_{tf}e^{j\phi_{tf}+2\pi jf\tau/D+2\pi jt\omega/R}$$

When $\omega=\tau=0$, I will write

$$X \equiv X_{00} = \frac{1}{TF}\sum_{\substack{f=1,F\\t=1,T}} w_{tf}e^{j\phi_{tf}} \qquad Y \equiv Y_{00} = \frac{1}{TF}\sum_{\substack{f=1,F\\t=1,T}} w_{tf}V_{tf}$$

I will also write X_r and X_i for the real and imaginary parts of X while $X_a^2 = X_r^2 + X_i^2$ and $X_{ri} = X_rX_i$. Similar notation applies for Y and when ω and τ are non-zero, *e.g.*, Y_r, $X_{\omega\tau r}$, *etc.*

The rate and delay indices ω and τ are assumed to be integers except as otherwise noted. The points in the delay-rate plane are also sometimes called delay-rate cells or just cells.

2 Statistics of the true delay-rate peak in the delay-rate plane

In order to derive the statistics of the true delay-rate peak in the delay-rate plane, I first compute the moments of X and Y. These moments are easily computed if I first compute the Fourier Transforms of the probability functions of X and Y.

The Fourier Transform of $p_{Y_r}(y_r)$, $p_{Y_i}(y_i)$ and $p_\Phi(\phi)$ are:

$$M_{Y_r}(\zeta) = e^{jk\zeta\sigma-\frac{1}{2}\zeta^2\sigma^2} \qquad M_{Y_i}(\zeta) = e^{-\frac{1}{2}\zeta^2\sigma^2}$$

$$M_\Phi(\zeta) = \int_{-\pi}^{\pi} d\phi p_\Phi(\phi)e^{j\zeta\phi} = e^{-\frac{1}{2}k^2}\int_0^\infty xdxe^{-\frac{1}{2}x^2}I_\zeta(kx) = e^{-2\gamma}\delta(\zeta) + \sqrt{\frac{\pi\gamma}{2}}e^{-\gamma}\left(I_{\frac{|\zeta-1|}{2}}(\gamma) + I_{\frac{|\zeta+1|}{2}}(\gamma)\right)$$

Here I_ζ is the modified Bessel function of integer order[5][6]. Note that M_Φ is even in ζ.

Evaluating M_Φ at different values of ζ yields the necessary moments of X but computation of the moments of Y requires the first and second derivatives of $M_{X_r}(\zeta)$ and $M_{X_i}(\zeta)$ with respect to ζ:

$$\begin{array}{ll} M'_{Y_r}(\zeta) = (jk\sigma - \zeta\sigma^2)e^{jk\zeta\sigma-\frac{1}{2}\zeta^2\sigma^2} & M'_{Y_i}(\zeta) = -\zeta\sigma^2 e^{-\frac{1}{2}\zeta^2\sigma^2} \\ M''_{Y_r}(\zeta) = \sigma^2\left[\zeta^2\sigma^2 - k^2 - 1 - 2jk\zeta\sigma\right]e^{jk\zeta\sigma-\frac{1}{2}\zeta^2\sigma^2} & M''_{Y_i}(\zeta) = \sigma^2(\zeta^2\sigma^2-1)e^{-\frac{1}{2}\zeta^2\sigma^2} \end{array}$$

The moments of X and Y are summarized below. The analysis of X's moments is based upon the treatment found in Goodman's Statistical Optics, Appendix B[7]:

$$\begin{array}{llll} \langle X_r\rangle & = \langle w\rangle M_\Phi(1) & \langle Y_r\rangle & = -j\langle w\rangle M'_r(0) \\ \langle X_i\rangle & = 0 & \langle Y_i\rangle & = -j\langle w\rangle M'_i(0) \\ \langle X_r^2\rangle & = \frac{\langle w^2\rangle}{2TF}\left[1+M_\Phi(2)\right] + \langle w\rangle^2\frac{TF-1}{TF}M_\Phi^2(1) & \langle Y_r^2\rangle & = -\frac{\langle w^2\rangle}{TF}M''_r(0) - \langle w\rangle^2\frac{TF-1}{TF}M_r'^2(0) \\ \langle X_i^2\rangle & = \frac{\langle w^2\rangle}{2TF}\left[1-M_\Phi(2)\right] & \langle Y_i^2\rangle & = -\frac{\langle w^2\rangle}{TF}M''_i(0) - \langle w\rangle^2\frac{TF-1}{TF}M_i'^2(0) \\ \langle X_{ri}\rangle & = 0 & \langle Y_{ri}\rangle & = 0 \\ \langle X_a^2\rangle & = \frac{\langle w^2\rangle}{TF} + \langle w\rangle^2\frac{TF-1}{TF}M_\Phi^2(1) & \langle Y_a^2\rangle & = -\frac{\langle w^2\rangle}{TF}(M''_r(0)+M''_i(0)) - \langle w\rangle^2\frac{TF-1}{TF}(M_r'^2(0)+M_i'^2(0)) \end{array}$$

[5] as defined in Eqn. 9.6.20 of Abramowitz and Stegun, Handbook of Mathematical Functions, Dover Publications, 1965.
[6] The integral was evaluated using A.&S. Eqns. 9.6.26 and 11.4.31
[7] Goodman, J., *Statistical Optics*, Wiley Interscience, 1985.

Plugging in the FT's, evaluated at the apropropiate values, these moments become:

$$
\begin{array}{ll}
\langle X_r\rangle = \langle w\rangle G(\gamma) & \langle Y_r\rangle = \langle w\rangle S \\
\langle X_i\rangle = 0 & \langle Y_i\rangle = 0 \\
\langle X_r^2\rangle = \frac{\langle w^2\rangle}{2TF}\left[2-\left(\frac{\sinh\gamma}{e^\gamma\gamma}\right)\right]+\langle w\rangle^2\frac{TF-1}{TF}G^2(\gamma) & \langle Y_r^2\rangle = \frac{\langle w^2\rangle}{TF}(\sigma^2+S^2)+\langle w\rangle^2\frac{TF-1}{TF}S^2 \\
\langle X_i^2\rangle = \frac{\langle w^2\rangle}{2TF}\left[\left(\frac{\sinh\gamma}{e^\gamma\gamma}\right)\right] & \langle Y_i^2\rangle = \frac{\langle w^2\rangle}{TF}\sigma^2 \\
\langle X_{ri}\rangle = 0 & \langle Y_{ri}\rangle = 0 \\
\langle X_a^2\rangle = \frac{\langle w^2\rangle}{TF}+\langle w\rangle^2\frac{TF-1}{TF}G^2(\gamma) & \langle Y_a^2\rangle = \frac{\langle w^2\rangle}{TF}(S^2+2\sigma^2)+\langle w\rangle^2\frac{TF-1}{TF}S^2
\end{array}
$$

where $G(\gamma) = M_\Phi(1) = \sqrt{\frac{\pi\gamma}{2}}e^{-\gamma}\left(I_0(\gamma)+I_1(\gamma)\right)$. Note that the cross-correlations X_{ri} both vanish so that X_r and X_i are uncorrelated.

2.1 The limiting cases of weak and strong signal

The limiting forms of the moments are summarized below for the weak and strong signal limits.

In the weak signal limit, the moments have the following functional forms:

$$
\begin{array}{ll}
\langle X_r\rangle = \langle w\rangle\sqrt{\frac{\pi}{8}}k & \langle Y_r\rangle = \langle w\rangle k\sigma \\
\langle X_i\rangle = 0 & \langle Y_i\rangle = 0 \\
\langle X_r^2\rangle = \left[\langle w^2\rangle+\langle w\rangle^2(TF-1)\frac{\pi}{4}k^2\right]\frac{1}{2TF} & \langle Y_r^2\rangle = \left[\langle w^2\rangle+\langle w\rangle^2(TF-1)k^2\right]\frac{1}{2TF}2\sigma^2 \\
\langle X_i^2\rangle = \langle w^2\rangle\frac{1}{2TF} & \langle Y_i^2\rangle = \langle w^2\rangle\frac{1}{2TF}2\sigma^2 \\
\langle X_{ri}\rangle = 0 & \langle Y_{ri}\rangle = 0 \\
\langle X_a^2\rangle = \left[2\langle w^2\rangle+\langle w\rangle^2(TF-1)\frac{\pi}{4}k^2\right]\frac{1}{2TF} & \langle Y_a^2\rangle = \left[2\langle w^2\rangle+\langle w\rangle^2(TF-1)k^2\right]\frac{1}{2TF}2\sigma^2
\end{array}
$$

In the strong signal limit, the moments have the following functional forms:

$$
\begin{array}{ll}
\langle X_r\rangle = \langle w\rangle & \langle Y_r\rangle = \langle w\rangle S \\
\langle X_i\rangle = 0 & \langle Y_i\rangle = 0 \\
\langle X_r^2\rangle = \left[\langle w^2\rangle+\langle w\rangle^2(TF-1)\right]\frac{1}{TF} & \langle Y_r^2\rangle = \left[\langle w^2\rangle+\langle w\rangle^2(TF-1)\right]\frac{1}{TF}S^2 \\
\langle X_i^2\rangle = \langle w^2\rangle\frac{1}{k^2TF} & \langle Y_i^2\rangle = \langle w^2\rangle\frac{1}{k^2TF}S^2 \\
\langle X_{ri}\rangle = 0 & \langle Y_{ri}\rangle = 0 \\
\langle X_a^2\rangle = \left[\langle w^2\rangle+\langle w\rangle^2(TF-1)\right]\frac{1}{TF} & \langle Y_a^2\rangle = \left[\langle w^2\rangle+\langle w\rangle^2(TF-1)\right]\frac{1}{TF}S^2
\end{array}
$$

At low SNRs, the square of the delay-rate peak's amplitude is slightly lower when using normalized visibilities instead of the traditional visibilities (as compared to the 'off-peak' rms noise). This may result in some small loss in sensitivity for weak sources.

2.2 Estimating SNR from the statistics of the delay-rate peak

Assuming that the fringe-search has found the origin as the highest point in the delay-rate plane, we are still not assured that the amplitude of the peak at the origin is the amplitude of the true delay-rate peak. This is because the amplitude at the origin is one realization of the possible values at the origin. Still, this realization is the best estimate we have of the true delay-rate peak's amplitude. It is easier to estimate the SNR (k) using the square of the amplitude of the delay-rate peak found by the fringe search.

If the expressions for $\langle X_a^2\rangle$ and $\langle Y_a^2\rangle$ are rearranged, we have:

$$
\frac{\langle X_a^2\rangle TF-\langle w^2\rangle}{\langle w\rangle^2(TF-1)} = G^2(\frac{k^2}{4}) \qquad \frac{\langle Y_a^2\rangle TF/\sigma^2-2\langle w^2\rangle}{\langle w^2\rangle+\langle w\rangle^2(TF-1)} = k^2
$$

Note that in each expression, the left-hand sides are composed entirely of known quantities with the possible exception of σ^2. On the right-hand sides, $G^2(\frac{k^2}{4})$ and k^2 can be inverted, numerically if not directly.

The second equation can be simply inverted to determine the SNR (k), but requires estimation of the noise, σ. The first equation involves the inverse of G which must be carried out numerically, or using a look-up table, but involves only known quantities.

When an *a priori* estimate of the noise level is not available, the SNR is easier to estimate when the input visibilities are normalized to unit amplitude. This procedure also renders the SNR estimate less susceptible to amplitude errors in the source model.

2.3 Comparison of the present analysis with that in FRING

The mapping of the squared-amplitude to SNR as implemented in FRING is here compared to that derived in the previous section.

Figure 1 plots the mapping function currently used in FRING and that derived in this memo for use in KRING. Given an observed peak value on the y-axis, the SNR is to be read off of the x-axis. Note that the x-axis is the SNR *per visibility* and should be multiplied by the square root of the number of visibilities to determine the SNR of the fringe search, the number that is reported by FRING and KRING. For example, in a solution interval of 10 minutes, with 2 second integrations and 16 channels across an IF, 4800 visibilities are used in the fringe search. Suppose a delay-rate peak squared-amplitude of 0.15 is seen. This corresponds to an SNR *per visibility* of 0.65. The actual SNR reported by FRING or KRING would be $0.65 \times \sqrt{4800} = 45$. For a larger delay-rate peak, FRING would report a larger SNR than KRING. For a smaller delay-rate peak, FRING would report a smaller SNR than KRING.

Figure 2 shows the overestimation factor for FRING versus KRING. Continuing the example of the last paragraph, if FRING reported an SNR of 5, that corresponds to an squared-amplitude of ≈ 0.002 for which KRING would report an SNR of 7.25.

3 Probability of false detection

There are two ways in which a fringe-search results in a false detection. The first is when there is actually no signal but the fringe-search yields a 'detection' with significant estimated SNR. The second is when there is a signal present with some delay and rate residual but the fringe-search yields a 'detection' at a different delay and rate residual.

The traditional analysis of the probability of false detection assumes that signal is present only in one cell in delay-rate space. In this case the two types of false detection described above are identical. If any cell besides the one containing signal is chosen by the fringe-search procedure, it is a false detection. Analysis of the probability of false detection is described in many places[8]. I begin by summarizing the traditional analysis, then present a similar analysis using normalized visibilities, and conclude with a more general analysis in which signal is allowed to be present in more than one cell in the fringe-search. In the latter analysis, the root causes of the two types of false detection can be differentiated.

3.1 Traditional analysis of the probability of false detection (PFD)

I begin by reviewing the traditional analysis of the PFD, in which the input visibilities Y are directly Fourier Transform'ed.

First, we need the statistics of points in the delay-rate plane when signal is and isn't present. Since Y is formed by averaging Gaussianly distributed random variables, its real and imaginary parts obey the central limit theorem. The statistics of the real and imaginary parts of Y are described by:

[8] see *e.g.* Thompson, Moran, & Swenson

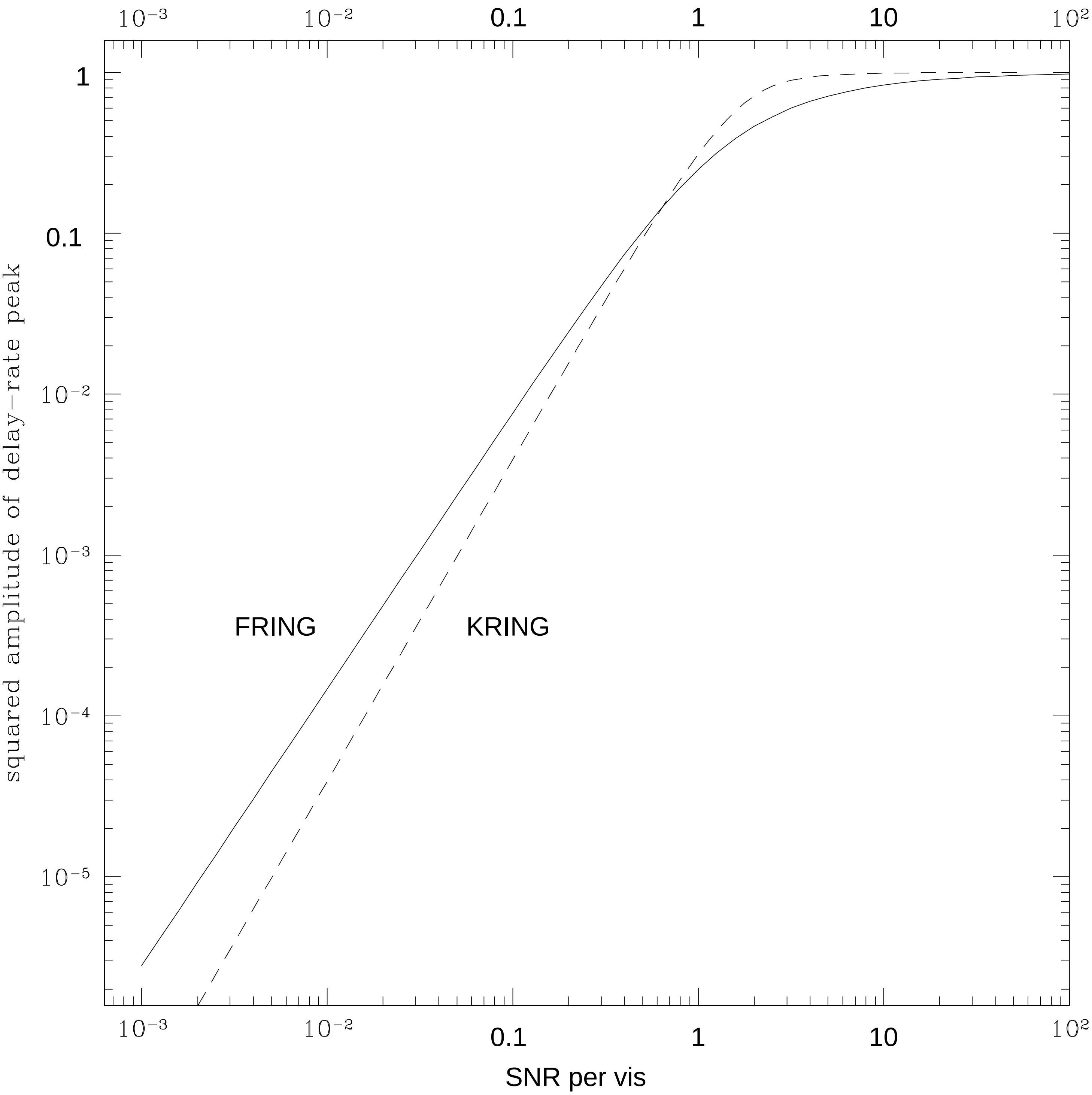

Figure 1: The solid line shows the mapping of squared-amplitude used by FRING to estimate SNR while the dashed line shows the mapping derived in this memo. The squared-amplitude of the delay-rate peak on the y-axis is to be mapped to an SNR on the x-axis. The abcissa is the SNR *per visibility* and should be multiplied by the square root of the number of visibilities used in the fringe search to obtain the value actually reported by FRING/KRING.

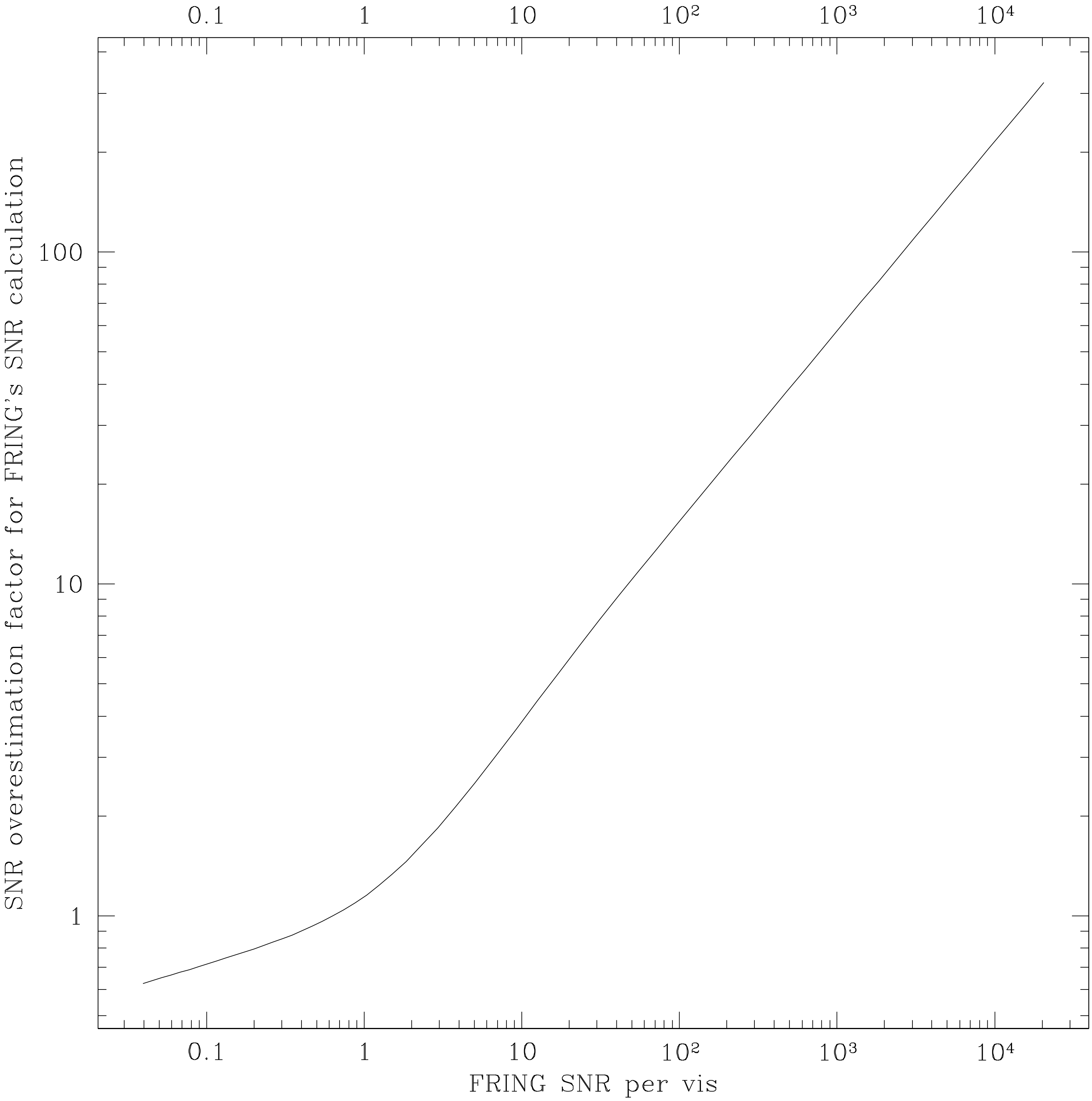

Figure 2: The factor by which the SNR reported by FRING exceeds that derived in this memo for a given fringe-search.

$$P_{Y_r}(y_r) = \frac{1}{\sqrt{2\pi}\sigma_r} e^{-\frac{(y_r - \langle w \rangle S)^2}{2\sigma_r^2}} \qquad P_{Y_i}(y_i) = \frac{1}{\sqrt{2\pi}\sigma_i} e^{-\frac{y_i^2}{2\sigma_i^2}}$$

where $\sigma_r^2 = \frac{\langle w^2 \rangle \sigma^2 + \sigma_w^2 S^2}{TF}$ and $\sigma_i^2 = \frac{\langle w^2 \rangle \sigma^2}{TF}$. It can be shown that Y_r and Y_i are un-correlated[9].

In the absence of signal, these probability distributions become:

$$P_{Y_r}(y_r)|_{k=0} = \frac{1}{\sqrt{2\pi}\sigma_a} e^{-\frac{y_r^2}{2\sigma_a^2}} \qquad P_{Y_i}(y_i)|_{k=0} = \frac{1}{\sqrt{2\pi}\sigma_a} e^{-\frac{y_i^2}{2\sigma_a^2}}$$

where $\sigma_a^2 = \frac{\langle w^2 \rangle \sigma^2}{TF}$.

What is the probability that the squared amplitude of an off-center delay-rate cell, $y^2_{\omega\tau a}$, is less than the squared amplitude of the delay-rate origin, y^2_{00a} ? This is just:

$$P(y^2_{\omega\tau a} < y^2_{00a}) = \int\limits_{y^2_{\omega\tau a} < y^2_{00a}} dy_r dy_i P_{Y_r}(y_r) P_{Y_i}(y_i) = 1 - e^{-\frac{y^2_{00a}}{2\sigma_a^2}}$$

Now what is the probability that jointly, for a search window in the delay-rate plane consisting of D delays by R rates, no point exceeds the true delay-rate peak in amplitude-squared?

$$P(y^2_{00a} > y^2_{\omega\tau a} : \forall\omega \neq 0, \tau \neq 0) = \frac{1}{2\pi\sigma_r\sigma_a} \int dy_r dy_i e^{-\frac{(y_r - \langle w \rangle S)^2}{2\sigma_r^2} - \frac{y_i^2}{2\sigma_a^2}} \left[1 - e^{-\frac{y_r^2 + y_i^2}{2\sigma_a^2}} \right]^{RD-1}$$

This is the probability of a true detection. Subtracting it from 1 yields the PFD:

$$PFD = \sum_{n=1}^{RD-1} \frac{\Gamma(RD+1)}{\Gamma(n+1)\Gamma(RD-n+1)} (-1)^{n+1} \frac{\sigma_a e^{-\frac{\langle w \rangle^2 S^2}{2} \frac{n}{\sigma_a^2 + n\sigma_r^2}}}{\sqrt{(1+n)(\sigma_a^2 + n\sigma_r^2)}}$$

Setting $\langle w \rangle = \langle w^2 \rangle = 1$, which corresponds to not using any weights for the input visibilities, this is the PFD of Eqn. 9.61 in Thompson, Moran, and Swenson.

3.2 PFD using normalized visibilities

I now calculate the PFD when the input visibilities are first normalized to unit amplitude to form X which is then Fourier Transform'd.

Again, we need the statistics of points in the delay-rate plane when signal is and isn't present. As described in Appendix B of Goodman, the real and imaginary parts of X obey the central limit theorem even if the signal disappears and the phase of X is uniformly distributed. The statistics of the real and imaginary parts of X are:

$$P_{X_r}(x_r) = \frac{1}{\sqrt{2\pi}\sigma_r} e^{-\frac{(x_r - \langle w \rangle G(\gamma))^2}{2\sigma_r^2}} \qquad P_{X_i}(x_i) = \frac{1}{\sqrt{2\pi}\sigma_i} e^{-\frac{x_i^2}{2\sigma_i^2}}$$

where $\sigma_r^2 = \frac{\langle w^2 \rangle (1 - \frac{\sinh\gamma}{2\gamma e^\gamma}) - \langle w \rangle^2 G^2(\gamma)}{TF}$ and $\sigma_i^2 = \frac{\langle w^2 \rangle}{2TF} (\frac{\sinh\gamma}{2\gamma e^\gamma})$.

[9] see Appendix B of Goodman's Optics book

In the absence of signal, these probability distributions become:

$$P_{X_r}(x_r)|_{k=0} = \frac{1}{\sqrt{2\pi}\sigma_a} e^{-\frac{x_r^2}{2\sigma_a^2}} \qquad P_{X_i}(x_i)|_{k=0} = \frac{1}{\sqrt{2\pi}\sigma_a} e^{-\frac{x_i^2}{2\sigma_a^2}}$$

where $\sigma_a^2 = \frac{\langle w^2 \rangle}{2TF}$.

What is the probability that the squared amplitude of an off-center delay-rate cell, $x_{\omega\tau a}^2$, is less than the squared amplitude of the delay-rate origin, x_{00a}^2 ? This is just:

$$P(x_{\omega\tau a}^2 < x_{00a}^2) = 1 - e^{-\frac{x_{00a}^2}{2\sigma_a^2}}$$

Now what is the probability that jointly, for a search window in the delay rate plane consisting of D delays by R rates, no point exceeds the true delay-rate peak in amplitude-squared?

$$P(x_{00a}^2 > x_{\omega\tau a}^2 : \forall\omega \neq 0, \tau \neq 0) = \frac{1}{2\pi\sigma_r\sigma_i} \int dx_r dx_i e^{-\frac{(x_r - \langle w \rangle G(\gamma))^2}{2\sigma_r^2} - \frac{x_i^2}{2\sigma_i^2}} \left[1 - e^{-\frac{x_r^2 + x_i^2}{2\sigma_a^2}}\right]^{RD-1}$$

So, the probability of false detection is

$$PFD = \sum_{n=1}^{RD-1} \frac{\Gamma(RD+1)}{\Gamma(n+1)\Gamma(RD-n+1)} (-1)^{n+1} \frac{\sigma_a^2 e^{-\frac{\langle w \rangle^2 G^2(\gamma)}{2} \frac{n}{\sigma_a^2 + n\sigma_r^2}}}{\sqrt{(\sigma_a^2 + n\sigma_i^2)(\sigma_a^2 + n\sigma_r^2)}}$$

3.3 Generalized analysis of the PFD

I now analyze the statistics for the generalized case when signal is present in more than one cell in the delay-rate search. I also discuss the conditions under which signal occupies more than one cell in the delay-rate search.

First, we need the statistics of an arbitrary point in delay-rate space in terms of the input visibilities.

$$X_{\omega\tau} = \frac{1}{TF} \sum_{\substack{t=1,T \\ f=1,F}} w_{tf} e^{j(\phi_{tf} + 2\pi t\omega/R + 2\pi f\tau/D)} \quad Y_{\omega\tau} = \frac{1}{TF} \sum_{\substack{t=1,T \\ f=1,F}} w_{tf} V_{tf} e^{j2\pi t\omega/R + j2\pi f\tau/D}$$

What are the moments of the real and imaginary parts of $X_{\omega\tau}$ at an arbitrary point in delay-rate space?

$$\langle X_{\omega\tau r} \rangle = M_\Phi(1) \langle w \rangle Q_{\omega\tau} \cos\psi$$

$$\langle X_{\omega\tau i} \rangle = M_\Phi(1) \langle w \rangle Q_{\omega\tau} \sin\psi$$

$$\langle X_{\omega\tau ri} \rangle = \langle X_{\omega\tau r} \rangle \langle X_{\omega\tau i} \rangle + HQ_{2\omega 2\tau} sin2\psi$$

$$\langle X_{\omega\tau r}^2 \rangle = \langle X_{\omega\tau r} \rangle^2 + \frac{\langle w^2 \rangle - \langle w \rangle^2 M_\phi^2(1)}{2TF} + HQ_{2\omega 2\tau} cos2\psi$$

$$\langle X^2_{\omega\tau i}\rangle = \langle X_{\omega\tau i}\rangle^2 + \frac{\langle w^2\rangle - \langle w\rangle^2 M^2_\phi(1)}{2TF} - HQ_{2\omega 2\tau}cos2\psi$$

$$\langle X^2_{\omega\tau a}\rangle = \langle X_{\omega\tau r}\rangle^2 + \langle X_{\omega\tau i}\rangle^2 + \frac{\langle w^2\rangle - \langle w\rangle^2 M^2_\phi(1)}{TF}$$

where $\psi = \pi\omega\frac{T-1}{R} + \pi\tau\frac{F-1}{D}$, $H = \frac{\langle w^2\rangle M_\Phi(2) - \langle w\rangle^2 M^2_\phi(1)}{2TF}$ and $Q_{\omega\tau} = \frac{sinc(2\pi\omega T/R)sinc(2\pi\tau F/D)}{sinc(2\pi\omega/R)sinc(2\pi\tau/D)}$.

If the Fourier-transforms are carried out with no over-sampling, $T = R$ and $F = D$ and $Q_{\omega\tau} = \delta_{\omega 0}\delta_{\tau 0}$. The moments then simplify considerably into the results of section 2.

In particular, note that the first moments become

$$\langle X_{\omega\tau r}\rangle = M_\Phi(1)\langle w\rangle\delta_{\tau 0}\delta_{\omega 0} \qquad \langle X_{\omega\tau i}\rangle = 0 \qquad \langle X_{\omega\tau ri}\rangle = 0$$

Implicit here is the assumption that ω and τ are integers.

The assumption that signal is present in only one cell in delay-rate space is equivalent to the assumption that the FFT is not oversampled, and that true delay-rate residual lies at the center of one of the searched FFT cells. The former is not true in practice while the latter is not true in general. The FFT is zero-padded, in FRING and KRING, by a factor of at least 4 and up to 16.

The statistics of $X_{\omega\tau}$ are those of two correlated Gaussian random variables:

$$P_X(x_r, x_i) = \frac{1}{2\pi\sigma_{\omega\tau r}\sigma_{\omega\tau i}(1-\rho^2)} e^{-\frac{(x_r - \langle X_{\omega\tau r}\rangle)^2}{2\sigma^2_{\omega\tau r}(1-\rho^2)} + \frac{(x_i - \langle X_{\omega\tau i}\rangle)(x_r - \langle X_{\omega\tau r}\rangle)\rho}{\sigma_{\omega\tau r}\sigma_{\omega\tau i}(1-\rho^2)} - \frac{(x_i - \langle X_{\omega\tau i}\rangle)^2}{2\sigma^2_{\omega\tau i}(1-\rho^2)}}$$

where

$\rho = \frac{\langle X_{\omega\tau ri}\rangle - \langle X_{\omega\tau r}\rangle\langle X_{\omega\tau i}\rangle}{\sigma_{\omega\tau r}\sigma_{\omega\tau i}}$, $\sigma^2_{\omega\tau r} = \sigma^2_a + HQ_{2\omega 2\tau}cos2\psi$, $\sigma^2_{\omega\tau i} = \sigma^2_a - HQ_{2\omega 2\tau}cos2\psi$ and $\sigma^2_a = \frac{\langle w^2\rangle - \langle w\rangle^2 M^2_\phi(1)}{2TF}$.

What is the probability that the squared amplitude of an off-center delay-rate cell, $x^2_{\omega\tau a}$, is less than the squared amplitude of the delay-rate origin, x^2_{00a}?

$$P(x^2_{\omega\tau a} < x^2_{00a}) = \int\limits_{x^2_a < x^2_{00a}} dx_r dx_i P_X(x_r, x_i) = \frac{1}{2\pi\sigma_{x\omega\tau}\sigma_{y\omega\tau}} \int\limits_{x^2_r + x^2_i < x^2_{00a}} dx_r dx_i e^{-\frac{(x_r - x_0)^2}{2\sigma^2_{x\omega\tau}} - \frac{x^2_i}{2\sigma^2_{y\omega\tau}}}$$

where $x_0 = M_\Phi(1)\langle w\rangle Q_{\omega\tau}$, $\sigma^2_{x\omega\tau} = \sigma^2_a + HQ_{2\omega 2\tau}$, $\sigma^2_{y\omega\tau} = \sigma^2_a - HQ_{2\omega 2\tau}$.

What is the probability that $x^2_{\omega\tau a}$ will jointly not exceed x^2_{00a} for all $\omega\tau$ in the search space?

$$1 - PFD = \prod_{\omega\tau} \frac{1}{2\pi\sigma_{x\omega\tau}\sigma_{y\omega\tau}} \int\limits_{x^2_r + x^2_i < x^2_{00a}} dx_r dx_i e^{-\frac{(x_r - x_0)^2}{2\sigma^2_{x\omega\tau}} - \frac{x^2_i}{2\sigma^2_{y\omega\tau}}}$$

This probability is too difficult to evaluate and is currently a matter of active research.

4 Conclusions

I have presented a formal analysis of the method by which the SNR can be estimated for the results of a fringe-search. This analysis shows that the SNR estimate in FRING is an underestimate at low SNR and an overestimate at high SNR. The former is contrary to the conventional wisdom. I have also shown that, if the input visibilities are normalized to unit amplitude before the fringe-search is carried out, the SNR can be estimated directly from the phases of the resultant visibilities which contain only information about the ratio of the signal to the noise, not either one separately. The procedure discussed in section 2 of this memo has been implemented in KRING.

Use of the normalized visibilities comes at a slight cost in sensitivity. However, there is a cost, in that the probability of false detection is slightly greater when using normalized visibilities .

Traditionally, the 'probability of false detection' is computed assuming that signal is present in only one cell in delay-rate space. When zero-padded Fourier transforms are used, this assumption is not valid. More generally, whenever the true delay/rate residual is not an integral number of search cells away from the starting delay/rate point, there will be signal present in many cells in the delay/rate plane, even if the Fourier transforms are not padded. The traditional analysis of the probability of false detection is generalized in section 3.3; there, signal is allowed in more than one cell in delay-rate space. The analysis in section 3.3 can be used to formally derive the probability of false detection when zero-padding is used before performing the FFTs.

5 Future Directions

When padded Fourier transforms are used, the delay-rate peak is more likely to lie at the center of a delay-rate cell, decreasing the probability of false detection. On the other hand, padding the Fourier transforms increases the probability of false detection because a noise spike close to the true peak may be found that has greater amplitude. A more complete analysis of the effects of zero-padding upon the probability of false detection is needed. Also, alternatives to zero-padding such as tapering and their effects upon the probability of false detection should be analyzed.

The PFDs of sections 3 remain to be calculated using numerical methods.

6 Acknowledgements

I would like to acknowledge Walter Brisken, Bryan Butler, and Michael Rupen, Vivek Dhawan, Fred Schwab, and Leonid Kogan for their helpful comments in preparing this memo.