A convex approximation to the likelihood criterion for aperture synthesis imaging.

Serge Meimon

Office National d'Études et de Recherches Aérospatiales

Département d'Optique Théorique et Appliquée

BP 72, F-92322 Châtillon cedex, France

Laurent M. Mugnier

Office National d'Études et de Recherches Aérospatiales

Département d'Optique Théorique et Appliquée

BP 72, F-92322 Châtillon cedex, France

Guy Le Besnerais

Office National d'Études et de Recherches Aérospatiales

Département Traitement de l'Information et Modélisation

BP 72, F-92322 Châtillon cedex, France

Aperture synthesis allows one to measure visibilities at very high resolutions by coupling telescopes of reasonable diameters. We consider that visibility amplitudes and phase are measured separately. It leads to an estimation problem where the noise model yields a non-convex data likelihood criterion. We show how to optimally approximate the noise model while keeping the criterion convex. This approximation has been validated both on simulations and on experimental data.

OCIS codes: 120.3180, 100.3020, 100.3190, 110.6770

1. Introduction

Aperture synthesis allows one to reach very high angular resolution by coupling telescopes of reasonable diameters in an interferometric array. Because current interferometers do not provide directly images, the data have to be processed through an appropriate imaging software.

The basic observables of an interferometer are the complex visibilities extracted from each fringe pattern formed by the instrument. In the absence of noise, complex visibilities amplitudes and phases are corrupted by atmospheric path length fluctuations, and by imperfect knowledge of the source position and of the interferometer geometry.

At radio wavelengths, it is usually possible to consider these errors as part of the noise, and to use directly complex visibility amplitudes and phases. On the contrary, at optical wavelengths, path length fluctuations due to atmospheric turbulence make visibility phases unexploitable. Thus, the observables of current interferometers at optical/infrared wave-

lengths are quantities independent of turbulent phases, such as squared visibilities and closure phases.

There are various ways of circumventing turbulence effects. A first one is obviously to locate the instrument where there is no turbulence, i.e. in space. In this case, complex visibilities are measurable. Secondly, if the u-v plane, i.e. the frequency coverage, is redundant enough, visibility phases can be successfully estimated from closure phases. This is the method used by Delage *et. al.* [1] to form complex visibilities from experimental squared visibilities and closure phases. However, redundancy techniques reduce the frequency coverage. Another promising way of obtaining complex visibilities with an optical interferometer in presence of turbulence is to use phase reference, as in the Very Large Telescope Interferometer (VLTI) instrument PRIMA (Phase-Referenced Imaging and Microarcsecond Astrometry) [2]. This method will allow astronomers to measure complex visibilities without constraining the u-v coverage.

Lastly, self-calibration algorithms [3] first developed for radio-interferometry, allow one to estimate both turbulent phases and the object, by alternating turbulent phases estimation steps with a known object and object reconstruction steps with known turbulent phases. The latters are strictly identical to Fourier synthesis problems without turbulence, i.e. to object reconstruction problems from noisy complex visibilities.

In this paper, we address object reconstruction from complex visibilities for both optical and radio wavelengths. The noise witnessed on complex visibilities yields a non-convex data likelihood criterion (Sect. 3.D), which makes reconstruction difficult.

After stating the interferometric data model we consider (Sect. 3), we compute an optimal approximation of it which yields a quadratic data likelihood criterion (Sect. 4). This

approximation is then validated on simulations and used to process experimental data [1] (Sect. 5).

2. Fourier synthesis

The basic observable of an interferometer is complex visibility, which can be measured from the fringe pattern obtained by combining the beams of two correctly phased telescopes. According to the Van Cittert-Zernike theorem [4], complex visibilities are related to the sky brightness distribution $x(a,b)$ through a Fourier Transform (FT):

$$V\left(\boldsymbol{\nu} = \begin{bmatrix} u \\ v \end{bmatrix}\right) = \iint x(a,b) \exp\left(-2\pi i(ua+vb)\right) \, \mathrm{d}a\mathrm{d}b \tag{1}$$

a and b being angular positions in the sky and $\boldsymbol{\nu}$ the 2D spatial frequency. For a couple of telescopes (T_1, T_2), the spatial frequency $\boldsymbol{\nu}$ is given by $\overrightarrow{\nu} = \frac{\overrightarrow{r_2} - \overrightarrow{r_1}}{\lambda}$, where $\overrightarrow{r_1}$ (resp. $\overrightarrow{r_2}$) denotes the position vector of T_1 (resp. T_2) projected onto a plane normal to the observation axis. $\overrightarrow{r_2} - \overrightarrow{r_1}$ is the corresponding baseline.

An interferometer is a device allowing to measure the Fourier Transform of an object at a set $\boldsymbol{\nu}$ of spatial frequencies. The aim of interferometry imaging is to retrieve the observed object from the set of measured Fourier samples. We adopt a Bayesian approach to solve this inverse problem, in which the first step is to design a data formation model, both accurately fitting the actual physical process and yielding a tractable estimation problem.

3. Reconstruction model

3.A. Matrix formulation

Let us suppose that the sky brightness $x(a,b)$ is discretized over a cardinal sine basis. It is thus represented by a vector of real coefficients $\boldsymbol{X} = [X_1, \ldots, X_j, \ldots, X_{N_p}]$, and equation (1) reads:

$$V(\boldsymbol{\nu}) = \sum_{j}^{N_p} h(j, \boldsymbol{\nu}) X_j, \tag{2}$$

the $h(m, \boldsymbol{\nu})$ being complex coefficients.

We derive the following matrix formulation

$$\boldsymbol{V}(\boldsymbol{X}) = \boldsymbol{H}\boldsymbol{X}, \tag{3}$$

with vector $\boldsymbol{V}$ and matrix $\boldsymbol{H}$ defined by

$$V_i = V(\boldsymbol{\nu_i})$$

$$h_{i,j} = h(j, \boldsymbol{\nu_i}),$$

where $\boldsymbol{\nu_i}$ denotes the i$^{\text{th}}$ measurement spatial frequency.

3.B. Noise statistics

We consider that measured visibility moduli and phases follow Gaussian distributions. Although our method generalizes to any Gaussian distribution of the visibility moduli and phases, we will assume in this paper that the cross correlations are either not available or negligible. Then the measured visibilities V_i^{meas} are linked to the "true" ones V_i by the

following model :

$$\begin{cases} |V_i^{meas}| = |V_i(\boldsymbol{X})| + b_{||,i} \\ \arg V_i^{meas} = \arg V_i(\boldsymbol{X}) + b_{\arg,i} \end{cases} \tag{4}$$

with all the noises centered, decorrelated and Gaussian. Let $\sigma_{||,i}$ the standard deviation of $b_{||,i}$, and $\sigma_{\arg,i}$ the standard deviation of $b_{\arg,i}$. Model 4 applies to the output of an unstable radio-interferometer [3]. In optical interferometry, it corresponds to the noise witnessed in various experimental settings where turbulence effects are either inexistent or sufficiently corrected (see section 1).

3.C. Bayesian estimation

Due to the poor spectral coverage, the object reconstruction is an ill-posed inverse problem and must be regularized (see Refs. [5], [6] and [7] for reviews on regularization), in the sense that some *a priori* information must be introduced in their resolution for the solution to be unique and robust to noise. In Bayesian estimation, the data likelihood $p(\text{ data}|\boldsymbol{X})$ is associated with a prior distribution $p(\boldsymbol{X})$. The "Maximum *a posteriori*" estimation is obtained by maximizing the joint probability

$$p(\boldsymbol{X}|\text{ data}) \propto p(\text{ data}|\boldsymbol{X})\, p(\boldsymbol{X}),$$

or by minimizing the opposite of its logarithm:

$$-\log p(\boldsymbol{X}|\text{ data}) = -\log p(\text{ data}|\boldsymbol{X}) \; -\log p(\boldsymbol{X}) + \text{constant term}.$$

Hence, it reduces to the minimization of a compound criterion:

$$J = J_{\text{data}} + \lambda J_{prior}, \tag{5}$$

where λ accounts for the confidence in the prior, and is called the regularization parameter. With a Gaussian prior on $\boldsymbol{X}$, i.e. if we consider that the distribution of the object $\boldsymbol{X}$ is Gaussian, J_{prior} is quadratic.

Here, we focus on the data-likelihood term, which is directly yielded by the noise model:

$$J_{data}(\boldsymbol{X}) \propto -\log(p(\boldsymbol{V}^{meas}|\boldsymbol{X})).$$

Taking into account data model (4), the data likelihood term reduces to J_1:

$$\begin{aligned} J_1(\boldsymbol{X}) = &\sum_{i=0}^{N_b-1} \left(\frac{|V_i^{meas}| - |V_i(\boldsymbol{X})|}{\sigma_{||,i}} \right)^2 \\ &+ \sum_{i=0}^{N_b-1} \left(\frac{\arg V_i^{meas} - \arg V_i(\boldsymbol{X})}{\sigma_{\arg,i}} \right)^2 \end{aligned} \tag{6}$$

with N_b the number of baselines for which the Fourier Transform of the object is measured.

3.D. A non-convex criterion

The strict convexity of the criterion is a sufficient condition of uniqueness of its minimum, and ensures the good behavior of classical minimization algorithms [8]. We show now that the functional J_1 of equation (6) is not convex. Because $\boldsymbol{V}$ is linked to $\boldsymbol{X}$ by a linear operator (see equation 3), the convexity of $\tilde{J}_1$ defined by $\tilde{J}_1(\boldsymbol{V}(\boldsymbol{X})) = J_1(\boldsymbol{X})$ is equivalent to the convexity of J_1. Because $\tilde{J}_1$ is a sum of N_b independent terms, we can deal with the case $N_b = 1$ without loss of generality. Then $\boldsymbol{V}_{meas}$ reduces to a complex number z_0, and $\boldsymbol{V}$ to a complex number z. $\tilde{J}_1$ reads:

$$\tilde{J}_1(z) = \frac{(|z| - |z_0|)^2}{\sigma_{||}^2} + \frac{(\arg z - \arg z_0)^2}{\sigma_{\arg}^2}$$

The choice of $z_1 = z_0 \exp(\frac{2i\pi}{3})$ and $z_2 = z_0 \exp(\frac{-2i\pi}{3})$ yields $\tilde{J}_1(z_1) = \frac{(2\pi/3)^2}{\sigma_{\arg}^2}$, $\tilde{J}_1(z_2) = \tilde{J}_1(z_1)$ and $z_1 + z_2 = -z_0$. Hence, we get

$$\tilde{J}_1\left(\frac{z_1+z_2}{2}\right) = \frac{(|z_0|/2)^2}{\sigma_{||}^2} + \frac{\pi^2}{\sigma_{\arg}^2}$$
$$\frac{1}{2}\left(\tilde{J}_1(z_1) + \tilde{J}_1(z_2)\right) = 0 + \frac{4\pi^2/9}{\sigma_{\arg}^2}$$

So $\tilde{J}_1\left(\frac{z_1+z_2}{2}\right) > \frac{1}{2}\left(\tilde{J}_1(z_1) + \tilde{J}_1(z_2)\right)$, which contradicts the convexity of $\tilde{J}_1$ (actually, this example shows the non convexity of both the phase term and the modulus term of $\tilde{J}_1$).

4. An equivalent additive Gaussian noise

In this section, we design an additive Gaussian approximation of the noise distribution, optimally "close" to the true one (in terms of a distance to be defined in the sequel), which yields a quadratic data likelihood criterion. We first recall the "true" distribution, then we state the general shape of any complex Gaussian distribution, expressed in a convenient basis, and we conclude by selecting the parameters of the optimal one.

4.A. Statement of the true distribution

Once again, we only have to study the complex unidimensional problem, which is generalized without any difficulty. We consider the following model:

$$\begin{cases} |z| = |z_0| + r \\ \arg z = \arg z_0 + \varphi \end{cases} \tag{7}$$

Hence, $z = (|z_0| + r)\exp[i(\arg z_0 + \varphi)]$ with r and φ following Gaussian centered distributions of variances $\mathrm{Var}(r) = \sigma_r^2$ and $\mathrm{Var}(\varphi) = \sigma_\varphi^2$. It is the model of (4).

The probability distribution of z is

$$\begin{cases} p\left(z = (|z_0| + r)\exp\left[i(\arg z_0 + \varphi)\right]\right) = f(r,\varphi) \\ f(r,\varphi) \propto \exp\left[-\frac{1}{2}\left(\frac{r^2}{\sigma_r^2} + \frac{\varphi^2}{\sigma_\varphi^2}\right)\right] \end{cases} \tag{8}$$

We want to approximate this distribution by an additive one. So, we have to recast model (7) in an additive one:

$$z = z_0 + B \tag{9}$$

and we choose to write B as

$$B = (x + iy)\exp\left(i\ \arg z_0\right). \tag{10}$$

Identification of (8) and (9,10) yields

$$\begin{cases} x = (|z_0| + r)\cos\varphi - |z_0| \\ y = (|z_0| + r)\sin\varphi \end{cases} \tag{11}$$

It is simple to see that x and y are the coordinates of B in the Cartesian basis $(\boldsymbol{u}_x, \boldsymbol{u}_y)$, corresponding to the canonical $(\Re, \Im)$ one, rotated by angle $\arg z_0$ (see Figure 1).

4.B. Statement of a complex Gaussian distribution

A complex noise is Gaussian if its vector representation in Cartesian coordinates is Gaussian. We choose the aforementioned Cartesian basis $(\boldsymbol{u}_x, \boldsymbol{u}_y)$. The change of basis is achieved by a rotation matrix $\mathcal{R}(\arg z_0)$, with

$$\mathcal{R}(\psi) = \begin{bmatrix} \cos\psi & \sin\psi \\ -\sin\psi & \cos\psi \end{bmatrix}. \tag{12}$$

Any additive Gaussian noise distribution can be written $p(z = z_0 + (x+iy)\exp[i\ \arg z_0]) = f_g(x,y)$, with:

$$f_g(x,y) = \frac{1}{\sqrt{2\pi \det \Sigma}} \exp\left\{ -\frac{1}{2} \begin{bmatrix} x-\bar{x} \\ y-\bar{y} \end{bmatrix}^t \Sigma^{-1} \begin{bmatrix} x-\bar{x} \\ y-\bar{y} \end{bmatrix} \right\} \tag{13}$$

with Σ a symmetric positive definite matrix.

We now compare it to the true distribution of (x,y) stated in Equations (8) and (11).

4.C. Kullback-Leibler divergence minimization

In order to choose the additive Gaussian distribution closest to the true one, we have to define a distance between two distributions. A convenient and well known one is the Kullback-Leibler divergence, defined by:

$$\delta(f_1, f_2) = \int f_1 \log\left(\frac{f_1}{f_2}\right).$$

Note that technically, this divergence is not a distance, because it is not symmetric. It is however often used as a discrepancy measure of f_1 w.r.t. f_2 because it is positive and equal to 0 only for $f_1 = f_2$. $\delta(f_1, f_2)$ is the expectation of the "log-distance" between two distributions $\log\left(\frac{f_1}{f_2}\right)$, w.r.t. the probability distribution f_1. To fit a Gaussian distribution f_g on the true distribution f, it is therefore natural to minimize $\delta(f, f_g)$ rather than $\delta(f_g, f)$.

As proved in the Appendix, the minimization of $\delta(f, f_g)$ yields the following optimal

parameters :

$$\begin{cases} \bar{x} = \mathrm{E}_f\{x\} = |z_0| \left[\exp\left(-\frac{\sigma_\varphi^2}{2}\right) - 1\right] \\ \bar{y} = \mathrm{E}_f\{y\} = 0 \\ \Sigma = \mathrm{Diag}\{\sigma_1^2, \sigma_2^2\} \\ \sigma_1^2 = \mathrm{E}_f\left\{(x-\bar{x})^2\right\} = \frac{|z_0|^2 + \sigma_r^2}{2}[1 + \exp\left(-2\sigma_\varphi^2\right)] - |z_0|^2 \exp\left(-\sigma_\varphi^2\right) \\ \sigma_2^2 = \mathrm{E}_f\left\{(y-\bar{y})^2\right\} = \frac{|z_0|^2 + \sigma_r^2}{2}[1 - \exp\left(-2\sigma_\varphi^2\right)] \end{cases} \tag{14}$$

The radial bias $\bar{x}$ can be estimated from z and σ_φ as $\bar{x} \approx |z| \left[\exp\left(-\frac{\sigma_\varphi^2}{2}\right) - 1\right]$. We shall note m the complex bias of coordinates $(\bar{x}, \bar{y})$.

4.D. Two Gaussian approximations

Circular approximation This simple isotropic Gaussian approximation, inherited from Radio Imaging, is obtained by setting σ_1 and σ_2 in Equation (14) to the same value. Such an approximation is valid in Radio Imaging with stable interferometers, and has been also used in optical interferometry [9]. However, it is not adapted to noise distributions in which the modulus standard deviation is different from the phase standard deviation, which is often the case in optical interferometry. We show here how to design an approximation specifically dedicated to process optical interferometry data.

Optimal approximation Instead of a circular approximation, we propose a second order expansion of the optimal Gaussian approximation stated in Equation (14), i.e. we consider

that $\sigma_\varphi/2\pi$ and $\sigma_r/|z_0|$ are small w.r.t. 1:

$$\begin{cases} \bar{x} = 0 \\ \bar{y} = \mathrm{E}_f\{y\} = 0 \\ \sigma_1^2 = \sigma_r^2 \\ \sigma_2^2 = |z_0|^2\sigma_\varphi^2 \end{cases} \tag{15}$$

Why choose the optimal approximation? The contours of the distribution of z around z_0 are plotted in figures 2 and 3 for the true noise statistics, for the optimal Gaussian approximation (more precisely, its second order expansion) and for the circular one. In Fig. 3, the radial noise level, i.e in the direction $\boldsymbol{u}_x$, given by $\frac{\sigma_r}{|z_0|}$, is greater than the one in the direction $\boldsymbol{u}_y$ given by σ_φ, whereas it is the opposite in Fig. 2.

For both configurations, these contour maps illustrate that our approximation fits better the true distribution.

4.E. N dimension case

With our Gaussian approximation, the data-likelihood for one measurement V_0^{meas} is

$$J_g(\boldsymbol{X}) = -2\log f_g(V_0^{meas} - V_0(\boldsymbol{X})).$$

With (12) and (13), we get:

$$J_g(\boldsymbol{X}) = \|V_0^{meas} - V_0(\boldsymbol{X}) - m_0\|^2_{\Sigma_{0,\mathcal{R}}} \tag{16}$$

with

$$\Sigma_{0,\mathcal{R}} = \mathcal{R}(-\arg V_0^{meas})\,\Sigma\,\mathcal{R}(\arg V_0^{meas}) \tag{17}$$

and

$$\|.\|^2_{\Sigma_{0,\mathcal{R}}} l = \begin{bmatrix} \Re(.) \\ \Im(.) \end{bmatrix}^t \Sigma^{-1}_{0,\mathcal{R}} \begin{bmatrix} \Re(.) \\ \Im(.) \end{bmatrix} \tag{18}$$

This expression can be easily generalized for N measurements:

$$J_g(\boldsymbol{X}) = \sum_{i=0}^{N-1} \|V_i^{meas} - V_i(\boldsymbol{X}) - m_i\|^2_{\Sigma_{i,\mathcal{R}}} \tag{19}$$

$$= \|\boldsymbol{V}^{meas} - \boldsymbol{V}(\boldsymbol{X}) - \boldsymbol{m}\|^2_{\Sigma} \tag{20}$$

Σ being bloc diagonal, with its blocks equal to the $\Sigma_{i,\mathcal{R}}$.

5. Validation on simulations and on experimental data

In this section, we compare "circular" approximation and our optimal Gaussian approximated distribution, denoted as " elliptic", in terms of reconstruction performances. To do so, we use either "circular" or " elliptic" noise model to build the data likelihood term, which we associate with the same prior term (see section 5.A.2) in a Bayesian reconstruction process.

Although our model clearly fits better the noise distribution, its performances are highly dependent on the noise outcome affecting the data. Hence, we will generate a hundred noise outcomes, in order to assess the average gain induced by our approximation.

We will then show that our method performs satisfactorily on real data.

5.A. Simulations

5.A.1. Simulated data

The data we process simulate VLTI measurements when observing an object corresponding to the model of the Ru Lupus Micro-jet developed by Paulo Garcia et al. [10]. The

frequency coverage (Fig 5b) has to be chosen rich enough to highlight the differences between reconstructions. Indeed, our method focuses on the data likelihood, which affects all the more the reconstruction quality as there are many data. The frequency coverage corresponds to six nights of observation of the same source with 3 telescopes of the VLTI, with 20 measurements each night. As already mentioned, we consider that the effects of turbulence are corrected enough to be included in the noise. The complex visibilities are corrupted by noise according to model (4), with $\sigma_{||,i} = |V_i^{meas}| \times 4.65\%$ and $\sigma_{\arg,i} = 0.27$ radians for all i.

5.A.2. Regularization and constraints

We choose a Gaussian and shift-invariant prior distribution for $\boldsymbol{X}$[11], so the distribution of its Fourier Transform $\tilde{\boldsymbol{X}}$ is a Gaussian distribution with a diagonal covariance matrix, and the diagonal components are the values of the object Power Spectrum Density (PSD) $PSD(\boldsymbol{\nu})$. Thus, the prior term reads :

$$J_{prior}(\boldsymbol{X}) = \sum_{\boldsymbol{\nu}} \frac{\left|\tilde{\boldsymbol{X}}(\boldsymbol{\nu}) - \tilde{\boldsymbol{X}}_m(\boldsymbol{\nu})\right|^2}{PSD(\boldsymbol{\nu})}.$$

The mean object $\tilde{\boldsymbol{X}}_m$ is assumed to be constant, with its flux equal to the measured flux, i.e. the null frequency measured visibility.

The PSD model chosen is the function

$$PSD(\boldsymbol{\nu}) = \frac{K}{\left(\frac{|\boldsymbol{\nu}|}{\rho_0}\right)^p + 1}.$$

The parameters K, p and ρ_0 are estimated by a maximum likelihood on the data.

As noted in [11], K plays the role of the regularization parameter λ (See Eq. 5), and can

be directly estimated from the data. Thus, the method is completely unsupervised, i.e. **no parameter has to be set by the user**.

Reconstruction uses a BFGS-method (Broyden-Fletcher-Goldfarb-Shanno) software OP-VMLM, designed by Eric Thiébaut[12], and is performed under positivity constraint. To compare the circular approximation with our method, we compute for each noise outcome an Improvement of the Root Mean Square Error (IRMSE) in decibels (dB). A positive IRMSE means a better reconstruction with our method. Figure (4) shows the IRMSE repartition histogram for the 100 noise outcomes.

The improvement is 4 dB in average, and 95% of the reconstructions have an IRMSE of more than 2dB. We can conclude that our elliptic approximation performs much better than the circular one, in terms of reconstructed image quality.

As mentioned before, reconstructions are performed with $\lambda = 1$. To measure the influence of λ on the IRMSE, we have processed the same data with $\lambda = 0.1$ and $\lambda = 10$. Table (1) provides IRMSE means and standard deviations over the 100 reconstructions for different λ.

For a variation of a decade around the nominal λ value, we still witness a clear reconstruction improvement with our method.

5.A.3. Reconstructions

To further illustrate the interest of using our method, we show in Fig 5 typical reconstructions for both methods: we have selected among 100 noise outcomes the one yielding an IRMSE close to the mean value. Our method obviously helps reducing the noise , yielding an Improvement of the Root Mean Square Error (IRMSE) worth 4 dB in average.

In next section, we process real experimental data in order to demonstrate the efficiency of our method with realistic noise distribution and frequency coverage.

5.B. Validation on experimental data

5.B.1. Experimental setup

Experimental data were graciously supplied by Laurent Delage and François Reynaud and correspond to the experiment described in Ref. [1] . The object is made of four stars of various magnitudes, and is observed through a fiber link interferometer featuring 61 frequency measurements. The data model used corresponds to system (4), because only the standard deviation of measurements are provided.

5.B.2. Regularization

Reconstructions are done under positivity constraint. We also use the quadratic regularization term described in section 5.A.2

5.B.3. Reconstruction

Fig. (6) shows the contour maps of the true object and the restored one. The 4 structuring elements are correctly reconstructed, although quadratic regularization slightly oversmoothed them. Table (2) shows that our reconstruction is correct in terms of relative positions of the peaks. We here validate that our method is efficient and usable on experimental data.

6. Concluding comments

We have designed an accurate data-likelihood criterion, which closely mimics the noise model while keeping the criterion convex. Our method performed satisfactorily both on simulated data and on experimental material. However, more sophisticated regularization should be investigated. Additionally, this paper did not address how to deal with closure phases instead of visibility phases. This can be done by using "Self-Calibration" methods, which alternate transfer function estimation steps with object reconstruction steps[13, 9]. We are currently developing an original self-calibration procedure which uses the likelihood approximation techniques developed in this paper.[14]

A. Kullback-Leibler Distance Optimization

We show here that for any given distribution $f(\boldsymbol{X})$, the Gaussian distribution defined by:

$$g(\boldsymbol{X}) = \frac{1}{\sqrt{2\pi \det \Sigma}} \exp\left[-\frac{1}{2} P(\boldsymbol{X})\right]$$

with Σ a symmetric positive definite matrix and

$$P(\boldsymbol{X}) = \left(\boldsymbol{X} - \bar{\boldsymbol{X}}\right)^t \Sigma^{-1} \left(\boldsymbol{X} - \bar{\boldsymbol{X}}\right)$$

which reaches the minimum of the Kullback-Leibler Distance $\delta(f, g)$ is such that:

$$\begin{aligned} \bar{\boldsymbol{X}} &= \mathrm{E}_f\left\{\boldsymbol{X}\right\} \\ \Sigma &= \mathrm{E}_f\left\{\left(\boldsymbol{X} - \mathrm{E}_f\left\{\boldsymbol{X}\right\}\right)\left(\boldsymbol{X} - \mathrm{E}_f\left\{\boldsymbol{X}\right\}\right)^t\right\} \\ &= \mathrm{Var}(\boldsymbol{X}) \end{aligned} \tag{21}$$

This property may result from general results of probability theory, but we provide here a compact and self-contained proof.

A.A. *Definition of the Kullback-Leibler Distance*

The distance $\delta(f,g)$ is defined by

$$\begin{aligned}\delta(f,g) &= \int f(\boldsymbol{X}) \log \frac{f(\boldsymbol{X})}{g(\boldsymbol{X})} \mathrm{d}\boldsymbol{x} \\ &= -\mathrm{E}_f\left\{\log g\right\} + cst\end{aligned}$$

So

$$\delta(f,g) = \frac{1}{2}\left(\mathrm{E}_f\left\{P(\boldsymbol{X})\right\} + \log\det\Sigma\right) + cst$$

A.B. *First order terms*

$$\begin{aligned}\frac{\partial\delta(f,g)}{\partial\bar{X}} = 0 \Rightarrow \frac{\partial \mathrm{E}_f\left\{P(\boldsymbol{X})\right\}}{\partial\bar{X}} &= 0 \\ \Rightarrow \mathrm{E}_f\left\{\frac{\partial P(\boldsymbol{X})}{\partial\bar{X}}\right\} &= 0 \\ \Rightarrow \mathrm{E}_f\left\{2\Sigma^{-1}\left(\boldsymbol{X}-\bar{\boldsymbol{X}}\right)\right\} &= 0 \\ \Rightarrow \mathrm{E}_f\left\{\left(\boldsymbol{X}-\bar{\boldsymbol{X}}\right)\right\} &= 0 \\ \Rightarrow \mathrm{E}_f\left\{\boldsymbol{X}\right\} &= \bar{\boldsymbol{X}}\end{aligned}$$

A.C. *Second order terms*

$$\begin{aligned}&\frac{\partial\delta(f,g)}{\partial\Sigma} = 0 \\ &\Rightarrow \frac{\partial}{\partial\Sigma}\left[\mathrm{E}_f\left\{P\right\} + \log\det\Sigma\right] = 0 \\ &\Rightarrow \mathrm{E}_f\left\{\frac{\partial P}{\partial\Sigma}\right\} + \frac{\partial\log\det\Sigma}{\partial\Sigma} = 0 \\ &\Rightarrow \mathrm{E}_f\left\{-\Sigma^{-t}\left(\boldsymbol{X}-\bar{\boldsymbol{X}}\right)\left(\boldsymbol{X}-\bar{\boldsymbol{X}}\right)^t\Sigma^{-t}\right\} + \Sigma^{-t} = 0\end{aligned}$$

Because Σ is symmetric, we have: $\Sigma^{-t} = \Sigma^{-1}$, so

$$\begin{aligned}
&\frac{\partial \delta(f,g)}{\partial \Sigma} = 0 \\
&\Rightarrow \Sigma^{-t} \mathrm{E}_f \left\{ \left(\boldsymbol{X} - \bar{\boldsymbol{X}}\right) \left(\boldsymbol{X} - \bar{\boldsymbol{X}}\right)^t \right\} \Sigma^{-t} = \Sigma^{-t} \\
&\Rightarrow \mathrm{E}_f \left\{ \left(\boldsymbol{X} - \bar{\boldsymbol{X}}\right) \left(\boldsymbol{X} - \bar{\boldsymbol{X}}\right)^t \right\} = \Sigma \Sigma^{-t} \Sigma \\
&\Rightarrow \Sigma = \mathrm{E}_f \left\{ \left(\boldsymbol{X} - \bar{\boldsymbol{X}}\right) \left(\boldsymbol{X} - \bar{\boldsymbol{X}}\right)^t \right\} \\
&\Rightarrow \Sigma = \mathrm{Var}(\boldsymbol{X})
\end{aligned}$$

which concludes the proof.

A.D. 2-dimensional case

$$\begin{aligned}
g(x,y) &= \frac{1}{\sqrt{2\pi \det \Sigma}} \exp -\frac{1}{2} P \\
P(\bar{x},\bar{y}) &= \begin{bmatrix} x-\bar{x} \\ y-\bar{y} \end{bmatrix}^t \Sigma^{-1} \begin{bmatrix} x-\bar{x} \\ y-\bar{y} \end{bmatrix}
\end{aligned}$$

is such that:

$$\begin{aligned}
\bar{x} &= \mathrm{E}_f \{x\} \\
\bar{y} &= \mathrm{E}_f \{y\} \\
\Sigma &= \mathrm{E}_f \left\{ \begin{bmatrix} x-\bar{x} \\ y-\bar{y} \end{bmatrix} \begin{bmatrix} x-\bar{x} \\ y-\bar{y} \end{bmatrix}^t \right\}
\end{aligned} \tag{22}$$

B. Acknowledgments

The data processed in section 5.B were supplied by Laurent Delage and François Reynaud. The authors want to express their special thanks to Eric Thiébaut for his support and for letting them use his minimization software.

Corresponding author Serge Meimon can be reached at Serge.Meimon@onera.fr

References

1. L. Delage, F. Reynaud, and E. Thiébaut, "Imaging laboratory test on a fiber linked telescope array," Opt. Commun. **160,** 27–32 (1999).
2. A. Quirrenbach *et al.*, "PRIMA: Study for a Dual Beam Instrument for the VLT Interferometer," in *Astronomical Interferometry*, 3350 (1998).
3. T. J. Cornwell and P. N. Wilkinson, "A new method for making maps with unstable radio interferometers," Mon. Not. R. Astr. Soc. **196,** 1067–1086 (1981).
4. J. W. Goodman, *Statistical optics* (John Wiley & Sons, New York, 1985).
5. D. M. Titterington, "General structure of regularization procedures in image reconstruction," Astron. Astrophys. **144,** 381–387 (1985).
6. *Approche bayésienne pour les problèmes inverses*, J. Idier, ed., (Hermès, Paris, 2001).
7. G. Demoment, "Image Reconstruction and Restoration: Overview of Common Estimation Structures and Problems," IEEE Trans. Acoust. Speech Signal Process. **37,** 2024–2036 (1989).
8. W. Press, B. Flannery, S. Teukolsky, and W. Vetterling, *Numerical Recipes in C* (Cambridge University press, 1988).
9. A. Lannes, "Weak-phase imaging in optical interferometry," J. Opt. Soc. Am. A **15,** 811–824 (1998).
10. P. J. V. Garcia, S. Cabrit, J. Ferreira, and L. Binette, "Atomic T Tauri disk winds heated by ambipolar diffusion. II. Observational tests," Astron. Astrophys. **377,** 609–

616 (2001).

11. J.-M. Conan, L. M. Mugnier, T. Fusco, V. Michau, and G. Rousset, "Myopic Deconvolution of Adaptive Optics Images using Object and Point Spread Function Power Spectra," Appl. Opt. **37,** 4614–4622 (1998).
12. E. Thiébaut, "Optimization issues in blind deconvolution algorithms," in *Astronomical Data Analysis*, 4847 (2002).
13. G. Le Besnerais, "Méthode du maximum d'entropie sur la moyenne, critères de reconstruction d'image et synthèse d'ouverture en radio-astronomie," Thèse de doctorat, Université de Paris-Sud, Orsay, 1993.
14. S. C. Meimon, L. M. Mugnier, and G. Le Besnerais, "A novel method of reconstruction for weak-phase optical interferometry," in *New frontiers in stellar interferometry*, W. A. Traub, ed.,**5491,** 909–919 (2004), Date conférence : June 2004, Glasgow, UK.

Table Captions:

- Table.1 :Influence of regularization parameter on IRMSE
- Table.2 :Relative positions and flux of the 3 faintest star w.r.t. the brightest one

Table 1. Influence of regularization parameter on IRMSE

λ	mean(IRMSE)	Std. Dev. (IRMSE)
0.1	2.6	1.8
1	4.1	1.4
10	7.8	0.9

Table 2. Relative positions and flux of the 3 faintest star w.r.t. the brightest one

	Position error (w.r.t. main star diameter D)		Intensity ($-2.5\log_{10}$ flux)	
star #	$\Delta x/D$	$\Delta y/D$	true	reconstructed
1	0	0	1.33	$1.07 \pm 19\%$
2	0.13	0.05	1.89	$1.72 \pm 9\%$
3	0.03	0.06	2.20	$1.95 \pm 11\%$
4	0.01	0.08	2.44	$2.41 \pm 2\%$

Figure Captions:

- Fig.1 :Polar and Cartesian coordinate systems in $\mathbb{C}$.
- Fig.2 : Noise distribution contour lines, for $\frac{\sigma_r}{|z_0|} < \sigma_\varphi$
- Fig.3 :Noise distribution contour lines, for $\frac{\sigma_r}{|z_0|} > \sigma_\varphi$
- Fig.4 :IRMSE repartition histogram
- Fig.5 : Simulation results : (a) True object , (b) u-v coverage : 360 frequencies, (c) reconstruction with elliptic approximation and (d) reconstruction with circular approximation. 256×256 pixels. Pixel size : 0.2 marcsec.
- Fig.6 :True object (left) and restored one (right). Contour levels : $10\%, 20\%, \ldots, 100\%$ of the maximum.
- Fig.7 :True object (left) and restored one (right). D is the diameter of the main star, used in table2.
- Fig.8 :Experimental frequency coverage.

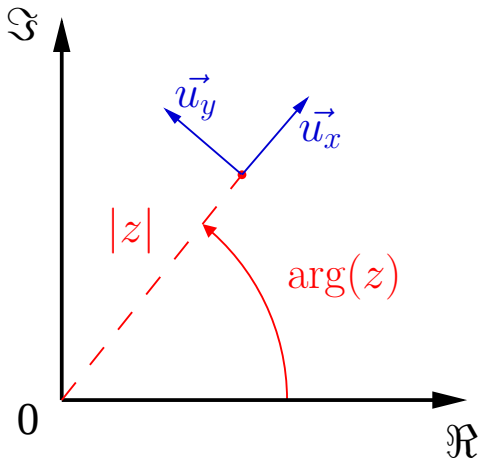

Fig. 1. Polar and Cartesian coordinate systems in $\mathbb{C}$.

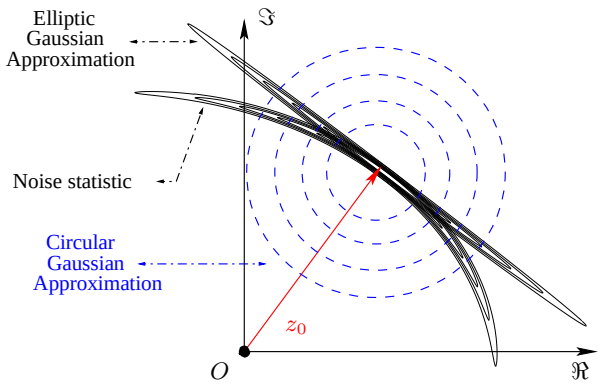

Fig. 2. Noise distribution contour lines, for $\frac{\sigma_r}{|z_0|} < \sigma_\varphi$

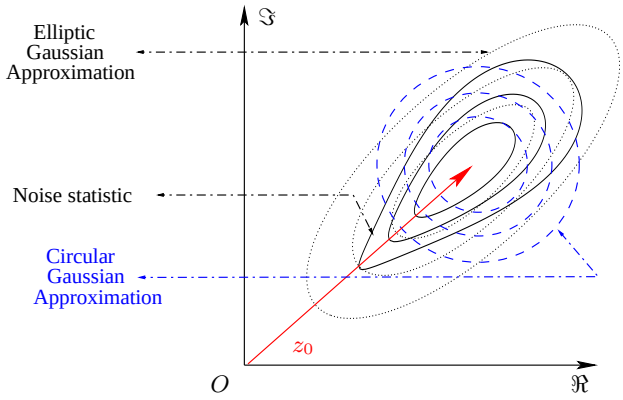

Fig. 3. Noise distribution contour lines, for $\frac{\sigma_r}{|z_0|} > \sigma_\varphi$

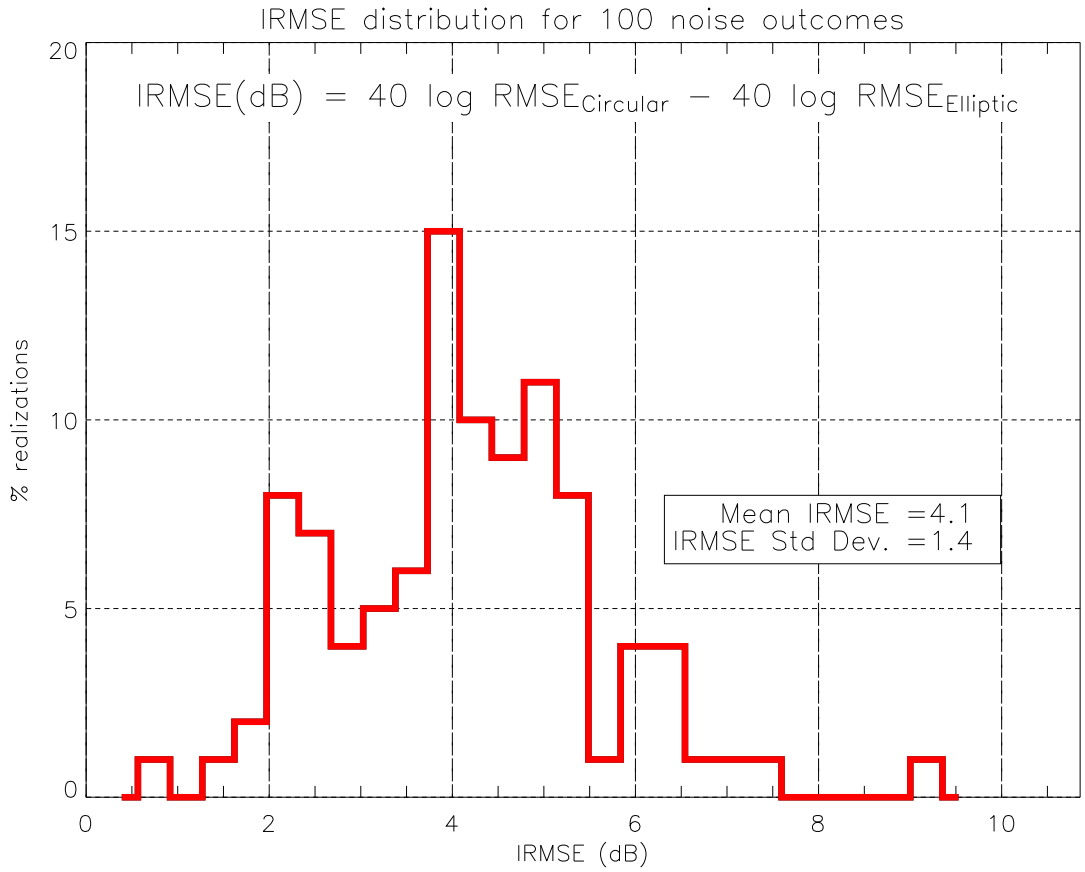

Fig. 4. IRMSE repartition histogram

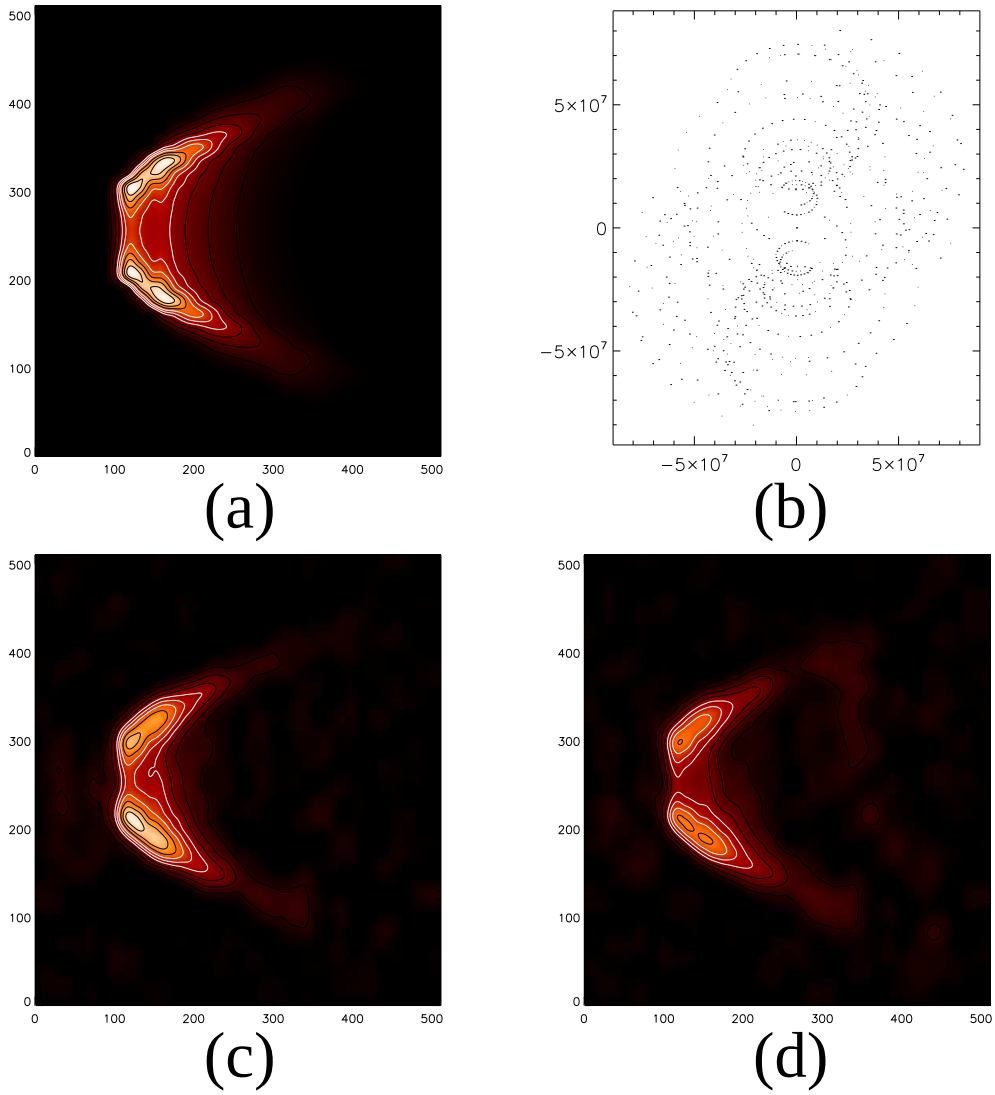

Fig. 5. Simulation results : (a) True object , (b) u-v coverage : 360 frequencies, (c) reconstruction with elliptic approximation and (d) reconstruction with circular approximation. 256×256 pixels. Pixel size : 0.2 marcsec.

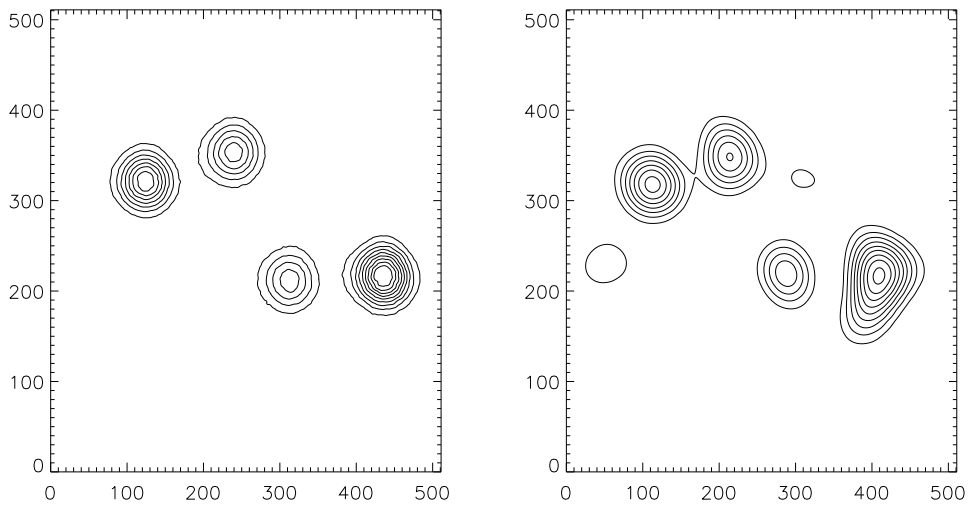

Fig. 6. True object (left) and restored one (right). Contour levels : $10\%, 20\%, \ldots, 100\%$ of the maximum.

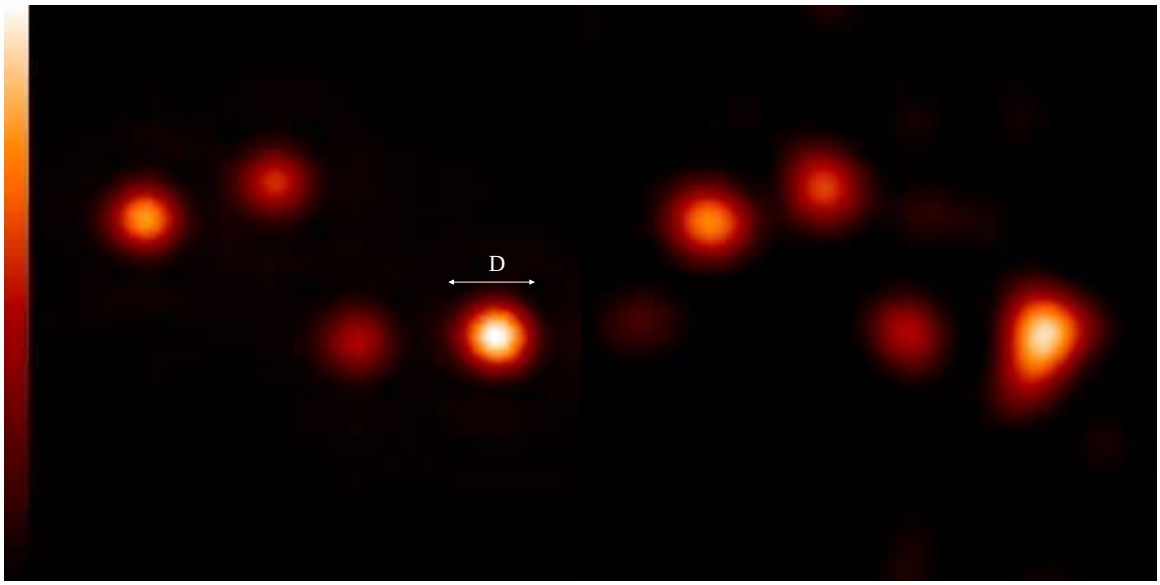

Fig. 7. True object (left) and restored one (right). D is the diameter of the main star, used in table2.

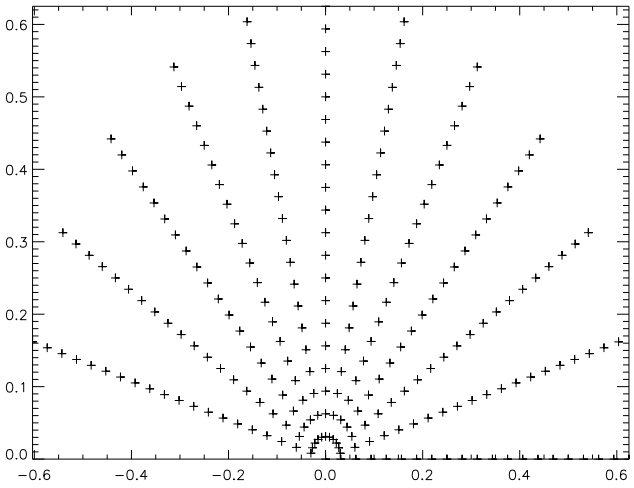

Fig. 8. Experimental frequency coverage.