Peeling the Visibility Onion
the optimum way of self-calibration

J.E. Noordam

(jnoordam@astron.nl)

Dwingeloo

15th November 2005

Verified:			
Name	Signature	Date	Rev.nr.
K. van der Schaaf	o.p.v.	15th November 2005	0.9.5

Accepted:		
Work Package Manager	System Engineering Manager	Program Manager
J.E. Noordam	C.M. de Vos	J. Reitsma
.....	.....	.....
date:	*date:*	*date:*

Distribution list:

Group:	For Information:
ASTRON	dOList
G. van Diepen (GVD)	C.M. de Vos
M. Brentjens (MAB)	
O.M. Smirnov (OMS)	
K. van der Schaaf (KvdS)	

Document revision:

Revision	Date	Section	Page(s)	Modification
0.5	2002-Feb-20	-	-	Creation
2.1	15th November 2005	??	??-??	Updated class description

Abstract

Predicting visibility values, for selfcal or source subtraction, can be very expensive (many sources, many parameters). Treating the bright sources one by one (peeling) is a promising way to deal with this problem. It is also a very useful approach for phase tracking.

1 Introduction

Self-calibration has been such a phenomenal, and above all 'easy' success, that nobody seems to have felt the need to explore it to the bottom. But because LOFAR is so much more difficult than the existing radio aperture synthesis telescopes, it has been necessary to probe a little deeper. The first step was the introduction of *'generalised selfcal'*, which has been described elsewhere [...]. Its main features are the use of a proper Measurement Equation (M.E.), the possibility of solving for arbitrary sequences of arbitrary subsets of its parameters, the specification of arbitrary constraints, and the treatment of all M.E. parameters as 2D polynomials in frequency and time. This document describes the concept of *'peeling'*, which can be considered the next big step forward. The main idea is simply to treat the brightest source in the field as if it was (almost) the only one, and solve for the subset of M.E. parameters that pertain to its direction. The brightest source is then subtracted ('peeled') from the uv-data, somewhat like peeling the outer shell of an onion. The process is repeated with the next brightest sources, each of which is peeled from the '*visibility onion*' in its turn.

One of the major advantages of peeling is processing speed, or rather a reduction of several orders of magnitude in the amount of processing that is required to calibrate LOFAR. This is rather welcome, since preliminary estimates were yielding some staggering numbers. It is achieved mainly because the expensive predict operation is needed for fewer domain cells, but there are gains in other areas as well.

Another advantage of peeling is that the extra complication (extra processing, tracking of core beams, general data organisation) of Station-Core Selfcal no longer seems necessary for LOFAR. The analysis offered here indicates that Station-Station Selfcal will work with fainter (or rather less visible) sources than was first thought. The reason is that the ionospheric phase corrections that are measured every 10 seconds on the 3-5 brightest sources can be used for ionospheric *phase-tracking*, so we always know the shape of the phase screen across the main lobe to better than a radian. This is then used to keep the residual phase errors towards fainter sources within one radian (say) so that they can be integrated longer before solving for M.E. parameters in their direction. Therefore, we only need 3-5 (rather than the original 20) sources in the field that are bright enough to give $S/N > 3$ in 10 sec. Peeling turns out to be very suitable for phase-tracking.

The many advantages of peeling have a price, of course. An analysis is presented of the influence of the other ('contaminating') sources in the field on the selfcal solution for the peeling source , and also of the influence of the noise ('noise bias'). Both are governed by an asymmetric effect, for which the name 'diode effect' is coined. Recommendations are made for strategies that minimise the influence of contaminating sources. In addition, a correction factor for the estimated noise bias is proposed.

The peeling approach suggests a natural distinction between Cat I and Cat II sources. It will be remembered that Cat I sources are so bright that they have to be predicted and subtracted with the full (and expensive) sophistication of DFT and individual corrections. The fainter Cat II sources are predicted and subtracted in groups, by means of an FFT and collective corrections (this approach is pioneered by the Haystack data simulator). Since peeling is the most effective way to predict and subtract individual

sources, all Cat I sources are 'peeling sources' by definition.

One of the consequences of peeling is that some M.E. parameters are not only associated with a station, but also with a source (direction). These *primary* parameters are used to estimate *secondary* M.E. parameters that describe 2D station beam-shapes, or the 2D ionospheric phase screen across them. These are then used in the prediction and subtraction of the many fainter (Cat II) sources in the Global Source Model, prior to making a map. They are also used to correct the residual uv-data for the center of a patch image. Since uv-data can only be corrected for a single point in the sky, the imaging quality of the many remaining Cat III sources, which are too faint ($< 5\sigma$) to be identified for the GSM, will decrease towards the edge of the patch. This problem can be kept within reasonable limits by reducing the size of the patch.

The concept of peeling is not really new. Experienced radio aperture synthesis astronomers (de Bruyn, Perley, Kassim, Cotton) have tried it in one form or another, using the existing reduction packages as a kind of 'bed of Procrustus'. The new LOFAR processing system will have the advantage that it will be specifically designed around the concepts of generalised selfcal and peeling. Especially important in this context is the idea of solving for the coefficients of 2D polynomials in freq and time, which makes it possible to use larger time-freq 'domains', i.e. larger numbers of visibilities for each selfcal solution. On the whole, the fact that 'things seem to start falling into place' might make us cautiously optimistic that we have taken at least some fortunate decisions on the dark road into the unknown.

2 Analysing the χ^2 minimisation

Self-calibration is a χ^2 (chi-squared) minimisation technique. For each visibility sample, a *selfcal equation* is generated, which equates the difference between measured and predicted values with an expression in the parameters p_k of the Measurement Equation (M.E). These equations are accumulated in a solution matrix, which is inverted to produce incremental improvements Δp_k. Since the M.E. is non-linear, several iterations are generally needed. In peeling, we concentrate on one source at a time (usually the brightest remaining one), which will be called the *peeling source*.

A measured visibility sample v_i can be written as:

$$v_i = m_i(\vec{p}) + \Delta m_i(\vec{p}) + c_i + n_i \tag{1}$$

in which $m_i(\vec{p})$ is the *predicted* value for the peeling source, using the Measurement Equation (M.E.) with the currently best known values of its parameters p_k (which include Sky Model parameters). The term $\Delta m_i(\vec{p})$ is the 'missing part' of the predicted value m_i, due to errors in the M.E. parameters. The term c_i is due to the other 'contaminating' sources in the sky, and n_i is the thermal noise. All terms are complex, of course.

When using the 'correct' values $\vec{p}_0$, the predicted value $m_i(\vec{p}_0)$ will be the correct one, and $\Delta m_i(\vec{p}_0) = 0$. This means that the function $(\Delta m_i^*(\vec{p})\Delta m_i(\vec{p}))$ will have a minimum at $\vec{p} = \vec{p}_0$. Because $(v_i - c_i - n_i)$ is independent of the p_k, we have:

$$\frac{\partial \Delta m_i}{\partial p_k} = -\frac{\partial m_i}{\partial p_k} \tag{2}$$

Therefore, the general form of a (peeling) selfcal equation is:

$$v_i - m_i(\vec{p}) = \sum_k \frac{\partial(\Delta m_i(\vec{p}) + c_i + n_i)}{\partial p_k}\Delta p_k = -\sum_k \frac{\partial m_i(\vec{p})}{\partial p_k}\Delta p_k \tag{3}$$

The idea is to estimate the incrementally better values Δp_k of the subset p_k of the M.E. parameters *in the direction of the peeling source* by minimising Δm_i. The latter is done by minimising:

$$\chi^2 = \sum_i (v_i - m_i)(v_i - m_i)^* = \sum_i (\Delta m_i + c_i + n_i)(\Delta m_i + c_i + n_i)^* \tag{4}$$

in which the sum is taken over a suitable subset of *nvis* visibilities, i.e. all the [1s, 1kHz] samples in a *domain* of [100s, 4MHz].

Obviously, this approach is only justified if c_i and n_i can be neglected. Traditionally this is done by integrating v_i to a reasonable S/N (greater than 3) , and by including sufficient sources in the Sky Model. Usually, using the 200 brightest sources is enough to approximate v_i to within a fraction of a percent, especially since the *vector sum* of the many faint sources rapidly tends to zero. However, predicting the contributions of 200 sources is rather expensive. Fortunately, a closer analysis of the various contributions to χ^2 in equ 4 suggests that this is not necessary.

We first write down the terms that depend on on the M.E. parameters $\vec{p}$ via $\Delta m_i(\vec{p})$, and thus affect the selfcal solution:

- <dMxdM> $= \sum_i (\Delta m_i^* \Delta m_i) = \sum_i \|\Delta m_i\|^2$: Positive and real. Depends on $\vec{p}$, with a minimum of zero at $\vec{p} = \vec{p}_0$. To be minimised by solving for better values of $\vec{p}$ in an (iterative) selfcal solution. See figs 4 and 3.
- <dMxC> $= \sum_i (\Delta m_i^* c_i + c_i^* \Delta m_i) = 2\sum_i \|c_i\|\|\Delta m_i\| \cos(\phi_i)$: Real. For a single contaminating source, $\phi_i = 2\pi(u_i\Delta l + v_i\Delta m + w_i\Delta n)$. The sum represents the *instantaneous synthesised beam* (or PSF), centered on the peeling source (see also section **??** below). Thus, the influence of contaminating sources on a selfcal solution is attenuated by this PSF. Even better, since it is multiplied by Δm, the contamination tends to zero as we get closer to the minimum. The PSF sidelobe level may minimised by including more visibilities, with better uv-coverage. Therefore, a 2D array is better than a 1D array like the WSRT. See also the *diode-effect* below. Since c_i is a vector sum,

the contributions of many contaminating sources will tend to average out. Ultimately, it is always possible, at a price in processing, to minimise c_i by including the brightest contaminating sources in the prediction.

- $<\text{dMxN}> = \sum_i (\Delta m_i^* n_i + n_i^* \Delta m_i) = 2 \sum_i \|n_i\| \|\Delta m_i\| \cos(\phi_i)$: Real. Since the angle ϕ_i between n_i and Δm_i is random, this term will decrease by $1/\sqrt{nvis}$. Its influence on the selfcal solution is a **noise bias**, which causes an under-estimation of $\|\Delta m\|$, and thus an over-estimation of the flux of the peeling source. The noise bias is caused by the the *diode effect* (see below). It is increased when the noise is non-gaussian, e.g. after a non-linear transformation like conversion to visibility phase or amplitude, especially at low S/N.

The remaining terms do not depend on $\vec{p}$, so they do not affect the minimisation of $\Delta m(\vec{p})$:

- $<\text{CxC}> = \sum_i (c_i^* c_i) = \sum_i \|c_i\|^2$: Positive and real.
- $<\text{NxN}> = \sum_i (n_i^* n_i) = \sum_i \|n_i\|^2$: Positive and real.
- $<\text{NxC}> = \sum_i (c_i^* n_i + n_i^* c_i) = 2 \sum_i \|c_i\| \|n_i\| \cos(\phi_i)$: Real. Since the angle ϕ_i between n_i and c_i is random, this term will decrease by $1/\sqrt{nvis}$.

The various terms are illustrated in figs 4 and 3.

NB: One might argue that c_i is weakly dependent on $\vec{p}$...

3 Error propagation in M.E. parameter estimation

Just a few points. This section is to be elaborated.

We are less interested in the errors in the M.E. parameters, as in the magnitude of the uv-residuals. They determine the dynamic range.

The selfcal system endeavours to minimise the uv-residuals. When peeling in the presence of noise and contaminating sources, this may cause the 'wrong' source to be subtracted. This may lead to second-order problems downstream, but the important thing is that the brightest sources are subtracted as completely as possible.

For the estimation of the error on a particular M.E. Parameter, use only those condition equations (uv-samples) that depend on this parameter. For a source-based parameter, this number is proportional to the number of interferometers, but for a station-based parameter, it is proportional to the number of stations. Since the dependence of a visibility on a parameter is expressed by its derivative in a condition equation, the latter should obviously play a role in the analysis of the problem...

3.1 The diode effect

For an explanation of the diode effect, see fig 3 and 4. The two cross-terms that depend on $\Delta m_i(\vec{p})$ are multiplied with a factor $\cos(\phi_i)$, where ϕ_i is the angle between Δm_i and either c_i or n_i. If this factor is positive, it only steepens the well of the χ^2 minimum, but its position remains at the correct $\vec{p} = \vec{p}_0$. But if the factor is *negative,* deeper minima will be created at the wrong values of $\vec{p}$, i.e. the selfcal solution will be affected. Of course we cannot say anything specific about the error in the estimated parameter values Δp_k. But we *can* say something about the extra depth of the minimum, and thus about the resulting *under-estimation* of $< \|\Delta m\| >$, and the *over-estimation* of the flux $< \|m\| >$of the peeling source.

3.2 Noise bias

Manifestation of the diode effect. Noise bias always leads to an underestimation of Δm, and thus to an overestimation of the flux $\|m\|$ of the peeling source. This can be estimated, and thus corrected.

Correction factor (a sort of Van Vleck correction):.....

3.3 The influence of contaminating sources

Manifestation of the diode effect. See fig 4. The term c_i is of course the sum of multiple contaminating sources:

$$\sum_i \|c_i\|\|\Delta m_i\| \cos(\phi_i) =< \|\Delta m_i\| > \sum_i \sum_k \|c_{ik}\| \cos(\phi_{ik})$$

in which $\phi_{ik} = u_i \Delta l_k + v_i \Delta m_k + w_i \Delta n_k$ and Δl_k etc are the distance between the peeling source and contaminating source k. This is exactly the expression for the map pixel at the position of the peeling source, multiplied with $< \|\Delta m_i\| >$. The latter may be taken outside the sum if we move the phase centre to the position of the peeling source. Note that the contamination decreases with Δm, i.e. when we get closer to the minimum.

So in the absence of ionospheric phase errors, the effect of contaminating sources on a selfcal solution are attenuated by the *instantaneous synthesised beam*, centered at the position of the peeling source. Unfortunately, this is the synthesised beam of only those interferometers that depend on the M.E. parameters that are being solved for. Fortunately, the resulting high sidelobes are reduced again by the randomizing effect of residual ionospheric errors. So, paradoxically, peeling works better in the presence of large ionospheric phases...

4 Minimising the number of domain cells

Because we wish to images the entire main lobe of the LOFAR station beam, the uv-data samples are available with a resolution of 1 s and 1 kHz. Predicted visibility values are used for selfcal, and for subtracting sources before making an image. Since the predict operation is very expensive, it is worth our while to try and minimise the number of domain cells by making them larger.

The only *practical* way to predict an (integrated) visibility value for a domain cell is to calculate the visibility function for the centre(f,t) of the cell, and then multiply this by cell bandwidth and integration time. This will only be equal to the measured value (a true integration) if the visibility function is *linear* over the cell. Therefore, approximate linearity is our criterion for maximum cell size. Obviously, the largest cells are possible for the smoothest visibility functions, i.e. if we contrive to minimize the visibility phase variation over a domain.

Since peeling deals with sources one by one, we only have to predict the visibility of the peeling source correctly. If the phase-centre of the uv-data is shifted to the apparent position of the current peeling source, the phase variation of its visibility function over a domain will be minimised. In the absence of ionospheric or other phase effects, the apparent position of the peeling source is its nominal (RA, DEC) position. But in the presence of such errors, the minimising phase factor(t,f) is the sum of the nominal position phase shift(t,f) and any other phase errors(t,f). Therefore, these other errors must be corrected in the measured data, by phase tracking.

We distinguish two cases: prediction of Cat I sources for selfcal, and prediction for subtracting Cat I and Cat II sources from the uv-data prior to making a (patch) image.

- In selfcal, we solve for M.E. parameters, or rather for the coefficients of their 2D functions in freq and time. The minimum number of cells in a selfcal domain is equal to maximum nr of coefficients in any of the 2D functions of the relevant subset of M.E. parameters. A typical number is 12 coefficients, i.e. 2nd degree in freq and 3rd degree in time. Since high-order polynomials are unstable, the number of coefficients will rarely exceed 40. Compared to the 400.000 full-resolution cells in the typical domain menitioned above, this is a reduction by a factor of 10.000. Since the predict operation (including differences) is relatively expensive, reducing the number of cells represents a reduction of at least 1000 in the overall amount of selfcal processing. This factor alone is sufficient reason to make the peeling approach *de rigueur*.
- At some point we have to predict and subtract all relevant Cat I and Cat II sources from the uv-data, prior to making a residual image. Since there are many Cat II sources, this is potentially a very expensive operation. The size of a patch image is determined by residual ionospheric errors, and will often be (much) smaller than the primary beam. Therefore, we may integrate the measured uv-data to a lower (t,f) resolution, e.g. cells of 10 s and 64 kHz. Of course this should be done after shifting the phase center of the uv-data to the center of the patch, and after correcting the ionospheric phase for this point.

An additional advantage of using large domain cells for selfcal is that the influence of contaminating sources will be reduced, because their visibility phase will vary considerably over a large cell. Note that it is always possible to reduce their influence even further by including the brightest contaminating sources in the prediction, even though their predicted value will be less accurate.

Finally a word on subtracting peeling sources from the uv-data. Obviously, this has to be done from the full-resolution data, because these have to be used for the next peeling source. However we will only have predicted the peeling source for the much larger cells, over which its visibility function will be linear, but certainly not constant. The solution is to calculate the gradient over the cell from its neighbours, which makes it possible to subtract the correct value from the much smaller cells. Of course this must be done while full-resolution data are still shifted 'towards' the peeling source.

In summary, most of the heavy lifting in selfcal and residual imaging may be performed on a tiny number of large cells. Since the required smoothing of the visibility function can only be achieved for one source at a time, this is the area where peeling has its greatest impact. The precise saving in processing remains to be calculated, but is likely to be several orders of magnitude.

5 Ionospheric phase tracking

It has become increasingly clear that the ionospheric phase variations, both in time and over the FOV, will be much larger than a radian for most LOFAR observing frequencies. This means that some kind of semi-on-line phase-tracking will be required for all LOFAR observing modes. Since uv-data can only be corrected for one position in the sky at a time, and since different corrections are needed for different parts of the FOV, it is difficult to see how this could be done without peeling.

In designing a strategy, we use the famous VLA 74 MHz Virgo A observation by Perley and Bust (see ...) as a point of departure. In fig 7 we can distinguish two main components in the TID phase. The *large-scale component* is roughly sinusiodal, with a period of about 15 min, and a peak-to-peak variation of 20 rad. Assuming a typical TID velocity over of 500 km/h, the TID wavelength is 135 km, i.e. much larger than a typical station beam size of 30 km at the assumed altitude (300 km) of the ionospheric TID phase screen. The maximum phase variation in a particular direction (e.g. a peeling source) is 5 deg/s, or 1 radian in 10 s. It will be seen below that, even though the effects of the large-scale component look pretty daunting and spectacular, the underlying phase variations are smooth enough to be handled with phase-tracking. Paradoxically, we should be more worried about the small-scale-component, which has phase variations of up to 0.5 rad/s, but fortunately stays within a radian (see sect 6).

In this section, we outline a phase-tracking procedue for the large-scale TID component. A maximum phase-rate of 0.1 rad/s suggests that there should be at least one *primary* tracking source in the FOV, i.e. one that can be tracked independently because it is bright enough to give $S/N > 3$ in 10 s. Using the regular peeling approach, such a source allows us to solve for the ionospheric phase in that direction with an accuracy of better than a radian, and to acquire and maintain phase-lock. However, since the

phase variation over a single FOV can be many radians, we need more information to track the phase in other directions. Fortunately, the TID scale is usually much larger than the FOV at the TID altitude, so that the 2D phase screen will be approximately flat across the FOV (and even the near sidelobes!). A minimum of 3-5 *primary* sources is needed to sample such a shape.

From the primary M.E. parameters in the direction of the 3-5 primary sources, we solve every 10 s for *secondary* M.E. parameters that describe the 2D phase screens(f,t) across the main lobes of the various station beams. These are used to track the phases in the direction of fainter *secondary* sources, so that these remain within 1 rad of the correct value while we integrate for 100-1000 s (boosted tracking). We then solve for the specific phases in the direction of these secondary sources, which are used subtract them with maximum accuracy, and to refine the screen shape.

NB: Phase-tracking will also be needed for bright sources in the station beam sidelobes. It is vitally important that the screen is flat enough to extend 'boosted' tracking to greater distances. We should also realise that the sidelobes farther from the main lobe will have more frequency structure, which complicates matters.

We now propose the following phase-tracking procedure (see also fig 8):

- Each pointing direction is treated separately and independently. The ionosphere is treated as a collection of virtually independent 2D phase-screens centered on the main lobe of individual stations. A screen is approximately flat across the main lobe, but can have large phase gradients. It may be extrapolated to the sidelobes of the same station beam, but not to other station beams. NB: It is not expected in this stage that 3D tomography will have sufficient predictive power to be used in LOFAR calibration. Therefore, it only plays a role as a byproduct for ionosphere scientists.
- In order to preserve sources at the edge of the field, the uv-data is available with the full resolution of 1 s and 1 kHz. A typical (f,t)-domain of 100 s and 4 MHz contains 400.000 full-resolution cells.
- Identify the bright (Cat I) sources that are to be used for phase-tracking. Depending on the ionospheric conditions and the required accuracy, this may range from a minimum of only 3-5 primary sources in the main lobe, to a much larger number of primary and secondary sources in main lobe and sidelobes. Remember that a primary source is defined as being bright enough to give $S/N > 3$ in 10s, and is used to estimate the shape of the 2D phase-screen to better than a radian. The latter allow the fainter secondary sources to be integrated longer before helping to refine the screen shape.
- Set up a solver for each of the sources. Define initial domains for each of them, with the full bandwith of 4-32 MHz, and a timelength of 10 s. These are dynamic domains, which will grow in timelength until get too long for a low-order time-polynomial. Then they are split into two overlapping ones. The optimal length depends on the ionospheric conditions.
- Collect uv-data for 10s. Then do the following, for all interferometers simultaneously:
  - [Any multiplicative or additive baseline-dependent effects must be applied here, before any station-based effects.]

  - Some uv-plane effects (e.g. bandpass) are applied to the full-resolution uv-data.
  - Peel off the phase-tracking sources, in order of brightness:
    * If it is a primary source, and if we are not already phase-tracking it, acquire phase-lock on it (see sect 8). Note that the necessity to maintain phase-lock on a number of primary sources will probably limit our freedom somewhat to move LOFAR beams rapidly over the sky, or even in frequency.
    * Smooth its visibility function over the domain by shifting the phase-centre of the full-resolution uv-data to its position, and by applying the best known ionospheric phases. The latter will be *extrapolated* in time, using the current coefficients of the 2D functions(f,t) that describe the ionospheric phase from the relevant stations in the direction of the peeling source.
    * [This is an excellent point to do some simple but effective flagging of transient or spectral RFI].
    * Integrate the uv-data in freq and time to get fewer larger cells over the same domain. The optimal cell size can be determined dynamically from the smoothness of the smoothed visibility of the peeling source. The latter depends on the size of residual phase errors after phase-tracking, which are equivalent to an error in apparent position.
    * Predict its visibility for the minimum number of cells. This could be either a re-calculation for all cells, or only for the cells corresponding to the last 10s of the dynamically growing domain. (Re-using the existing solution matrix is efficient). Supply one equation per predicted cell to the relevant solver, which adds them to its solution matrix.
    * If its solution matrix is 'ripe', solve for (an improvement of) its ionospheric phases, as seen from all stations. This will be every 10 s for the primary sources, after which a new estimate is made of the shapes of the 2D phase screens. A solution for the fainter secondary sources is made every 100-1000s.
    * Subtract the peeling source from the full-resolution uv-data. As explained in section 4, this can be done using the large cells.
  - Repeat for all peeling sources.
- Repeat every 10s.

A phase-tracking procedure like this should be running at all times during LOFAR observations, using at least 3-5 primary sources. This is necessary to condition the data for passing on to a pulsar processor or transient detector, or VLBI. But it is also a very powerful tool for on-line monitoring of the proper functioning of LOFAR itself, and for measuring ionospheric conditions. Essentially the same procedure can then be used off-line with more sources, to get or refine the specific phases in the direction of all the sources (in main lobe and side-lobes) that are bright enough to cause dynamic range problems.

Solutions for (subsets of) other M.E. parameters can be inserted in this procedure, either on-line or off-line. Most of them, but not all, will reside on the predict-side of the tree. One that must probably be included on-line is the station beam gain in the direction of the peeling sources. This is not only to be able to subtract the peeling sources properly, but also to monitor the LOFAR station beam-shapes.

This phase-tracking procedure is more complicated than anything we are used to with existing instruments (*except for VLBI fringe-fitting?*). However, it is difficult to deny that something like it will be necessary for LOFAR, so it is good to have at least one that might work. It would be an interesting challenge to devise one that works without peeling...

6 Summary of dynamic range limitations

This section is not very polished yet. It is not even directly to peeling. It merely contains a number of considerations, to be elaborated later.

The dynamic range in LOFAR images will not be limited by the Cat I sources, but by the fainter Cat II sources in between. The Cat I sources are subtracted optimally, using the near-perfect values of the M.E. parameters in their own direction. The Cat II sources are subtracted using the interpolated values of the beamshape and ionospheric phase screen, which may be slighty wrong (see fig 1). The error will be larger in the case of the ionospheric phase, because it is not guaranteed to be smooth like the beamshape. However, while these ionospheric phase residuals are random and will average out over time, the smaller beamshape residuals may still have the largest effect since they are systematic, i.e. correlated over longer time intervals.

NB: Cat II sources are predicted and subtracted in groups of sources that are close together. Since these 'patches' can be chosen to be arbitrarily small (at a cost in processing), this is not a potential dynamic range limitation. *NB: It is important to realise that we only need enough information to solve for the 2D phase screens, and NOT to solve for the fluxes of all Cat II sources. The latter will be refined over time, while they are being observed many times by LOFAR.*

Proposition: Any source that is bright enough to cause discernible dynamic range problems, is bright enough to solve the problem. More particularly: any Cat II source that causes discernible dynamic range problems is bright enough to be treated as a Cat I source, and be peeled. The latter not only allows us to subtract this particular Cat II source more accurately, but it also adds another sampling point to the estimation of station beamshapes and ionospheric phase screens.

Cat III sources are also a potential dynamic range limitation. They are not subtracted from the uv-data because they are too faint ($< 5\sigma$) to be identified and put into the Global Sky Model. Since uv-data can only be corrected for a single point in the sky, the maging errors, mostly due to residual ionospheric phase errors, will increase towards the patch edge. Fortunately, there is some indication that the resulting PSF degradation can be calculated, which suggests that at least the brightest Cat III sources can be

deconvolved (they can also be put into the GSM as Cat II sourc, of course). More fundamentally, the PSF sidelobes of the many invisible Cat III sources just below the noise will cause *side-lobe confusion*. This should be tackled first of all by minimising the PSF sidelobe level, but perhaps there are ways to distinguish it from thermal noise by statistical means.

The ultimate limitation on the dynamic range will be the artifacts that cannot easily be distinguished from real brightness. In anticipation of reaching that level after a few years of LOFAR operation, we should develop new methods to measure 'map reliability'. Such methods are likely to be statistical (like in the case of COBE), but they should incorporate as much *a priori* knowledge about the system as possible. Very likely, the capability of trying different Measurement Equations will prove to be a powerful tool.

Bright sources coming in via station sidelobes, especially the very extended galaxy, will be a major challenge.

Learning the art of LOFAR calibration will take several years, and the involvement of lots of bright people. The best we can do in this stage is to create a processing system that is as flexible and policy-free as possible, and to encourage its use with existing radio telescopes like WSRT, GMRT, VLA and JIVE. When LOFAR comes on-line, it is probably best to start observing at relatively high frequencies, far from the galaxy or with the galaxy under the horizon, and gradually increase the level of difficulty.

The small-scale TID component....

7 The processing advantages of peeling

NB: Indicate with respect to what.

The peeling approach offers very significant advantages in terms of processing speed and memory use. In fact, this was our first reason to turn to peeling, because a more traditional approach of selfcal looked to be prohibitively expensive in the case of LOFAR. Since it is difficult in this stage to calculate absolute numbers for the various operations, we will limit ourselves to a list of improvements factors:

- The very expensive prediction operation is needed for only one (or a few at most) sources per solution. For traditional selfcal, 200-500 sources would have to be included to predict a LOFAR visibility value v_i with an accuracy better than 1%. Estimated improvement factor: 100-500.
- Prediction is needed for far fewer (f,t) domain cells, due to shifting of the phase-centre to the position of the peeling source. Estimated improvement factor: 1000-10000.
- When the phase-centre is shifted to the peeling source, multiplication with the Fourier phase factor is done on the measured uv-data, and not on the many derivatives. Estimated improvement factor:

10-50.

- Fewer derivatives need to be calculated per predict (very expensive). Estimated improvement factor: 10.
- When solving for fewer parameters simultaneously, fewer iterations tend to be necessary. Estimated improvement factor: 2-3.
- Not every selfcal solution requires the full set of iterations. Successive iterations can be done on a gradually increasing domain (10 sec at a time), in which the new equations are added to the existing matrix. Estimated improvement factor: 10.
- Inversion of more, but smaller matrices. $N^3 > n(\frac{N}{n})^3 = N^3/n^2$. Estimated improvement factor: 100.

It is clear that the success of peeling hinges on the minimisation of the influence of the contaminating sources c_i on the selfcal solution. Fortunately, it is always possible to minimise c_i to arbitrarily low values by including more contaminating sources in the prediction process. The price to be paid is not only the prediction of extra sources, but also the extra number of domain cells needed to predict sources away from the phase centre (which is placed on the peeling source). However, the prediction of the contaminating sources does not have to be as accurate as for the peeling source itself.

8 Miscellaneous

The following points have come up since the original peeling document was distributed on 17 March. They will be discussed at the project meeting, and integrated into the next version.

8.1 Nr of stations

Given a total nr of dipoles, the χ^2 noise is independent of the number of stations. This is because $\sqrt{nvis} \propto nstat$.

8.2 Contamination $c_i = c_i(\vec{p})$?

In section 2, we stated that the contributrion c_i of contaminating sources does not depend on the M.E. parameters $\vec{p}$. In reality, there is of course a weak dependence...

8.3 Parameter orthogonality

Peeling one source at a time: solving for a small subset of M.E. parameters in its direction. Parameters can still be dependent, but there will always many more equations than unknowns. However, the total number of parameters may still be too large. So are we being naive, or very clever, or both? What is the stabilising influence of doing the brightest source first?

NB: (JPH) after self-alignment, seven (!) parameters remain...

8.4 Robustness in the presence of RFI

When using largisch domains, with many (1s, 1kHz) uv-samples, we can ignore the rms value of $(v_i - m_i)$ and ignore the samples where that deviate by more than 3σ (say). Ignoring is very simple: just do not generate a selfcal equation for that sample.

8.5 Flagging while peeling

This is related to the robustness issue above, of course.

Criterion is adapted to the (relative) brightness of the peeling source. Flags mapped onto measured uv-data (cells).

8.6 The use of uv-plane effects

Traditional selfcal solves only for uv-plane effects, i.e. M.E. parameters that are valid for the entire field. The advantage is that there is much less processing required, and that the image made from the corrected uv-data is the final image. With LOFAR this is no longer possible, because different sources in the field require different corrections. First of all, this requires that as many sources as possible are subtracted from the uv-data before making an image from the residuals. But in order to get a useful residual image we still have to apply a correction to the residual uv-data. Usually, this will be the correction that is valid for the center of the 'patch', which is a small part of the primary beam.

For this reason, and for use with existing instruments like WSRT, we need to implement MeqTrees for uv-plane effects. These might be applied to the measured data, either before or after integration, or as part of the predict process. In any case, one may solve for their parameters in exactly the same way as for other M.E. parameters, and without knowing on which side they reside. . See fig 8.

NB: When solving for uv-plane effects, equations are accumulated for more than one peeling source.

8.7 Extended sources

Peeling sources are likely to be extended. Since we are shifting the phase-center to the peeling source, its contribution to the visibility does no longer depend on its position in the field. This might offer processing advantages (to be explored). However, a potential obstacle is that image-plane effects may vary across the source, and therfore have to be applied before the DFT. See also fig 8.

8.8 Acquiring phase-lock

As explained in section 5, one of the reasons for phase tracking is that the ionospheric phase varies by many radians during an observation. But before we can start tracking, we need a procedure to acquire phase-lock, i.e. to remove 2π phase ambiguities. This can be done in the following way:

- Start with a single interferometer (2 stations).
- Include an extra station. Vary the ionospheric phase in the direction of the brightest source (a primary M.E. parameter) by steps of 2π until all ifrs are consistent with each other. If the source is bright enough, this is quite dramatic, and easily detected. Track the phase of this station from here on.
- Repeat this process until all stations are included.

The process can be speeded up by continuously solving for a 2D phase-screen above the phase-locked part of the array, and to use this to predict the phase above the next station to be included. This should reduce the search range to only a few 2π steps. For this reason, the process should start with the stations near the center of the array, and work its way outwards.

In practice, we will have to phase-track more than one source in the field. In that case we need a *separate* procedure like the one above for each source, which have to be made consistent with each other. *(to be elaborated).*

8.9 Acquiring Faraday lock

Faraday rotation varies strongly with freq (and time). If not modelled correctly, they will average out any polarised flux over a largish domain.

Estimate per ifr: maximise the average value over the time-freq domain.

Consistency over all ifrs.

Constraint: known freq dependence (use MeqParm solution mask)

8.10 Sky frame

The peeling approach has serendipitously shown the way towards a simple and natural way to describe variable station beamshapes and ionospheric refraction and magnification. Since we derive the 2D beamshapes and ionospheric phase screens from gains and phases in the direction of peeling sources, it was natural to describe them in the sky-frame coordinates, i.e. in terms of RA and DEC. It turns out that this is all one needs to know to predict visibilities correctly. An Earth-frame concept like apparent position due to refraction is completely useless, because it only describes the linear term of a phase-screen over the entire array, just like 'magnification' only describes the second-order term. In contrast, a set of mutually consistent (phase-locked) 2D phase screens over the station main lobes covers all. The same is true for beamshapes described in Earth-frame coordinates. An excellent example of how different ways to look at the same problem can make a huge difference in its apparent complication.

9 Do we still need Station-Core Selfcal (SCS)?

For a long time we believed that we needed the extra sensitivity (5 times, because the Core is the equivalent of 25 stations), but most of all the narrow Core beams with low sidelobes, to get a handle on LOFAR calibration. Fortunately, this no longer seems necessary.

The implementation of Station-Core Selfcal would complicate the processing considerably. Not only because we need a separate system next to Station-Station Selfcal, with its own data management etc. But also because the core beams are so narrow that they would have to actively track the calibrator sources at which they are pointed. And finally, it is not clear whether some of the assumptions that would have to be made to get a SCS solution would not propagate into the final result.

The answer to the question is no (see also the conclusions below). Why did we think we needed it in the first place?

(*to be elaborated*)

10 Conclusions

The most important reason for developing a 'peeling' version of self-calibration for LOFAR is the potentially (very) large reduction in processing for calibration. This is of vital importance, because the LOFAR Measurement Equation has many more parameters than the M.E. of traditional radio telescopes.

It probably has more than the future Square Km Array (SKA), so that the LOFAR processing approach is very relevant for SKA.

It turns out that peeling also has attractive secondary features. A summary of the advantages of peeling is:

- A very considerable reduction in processing (see section 7).
- Closed-loop (and thus optimal) subtraction of the brightest sources, using their own primary M.E. parameters, rather than interpolating the beamshapes and ionospheric phase-screen described by secondary M.E. parameters. In other words, the brightest sources are subtracted in such a way that their residuals are minimised, which is much more important for the dynamic range than subtracting them 'correctly'.
- Peeling is essential for ionospheric phase-tracking. The phase difference between different sources in the same field is so large (many radians) that phase-tracking is needed on more than a single source. Since the tracking correction has to be applied to the measured data, a source can only be tracked after the brighter sources have been peeled off.
- Peeling guarantees parameter orthogonality, since it deals with small subsets of M.E. parameters at a time. (This might be a bit of a spurious argument).
- Peeling simplifies the processing system in many ways (*to be elaborated*). For instance, it offers a more 'natural' way to acquire (ionospheric) phase-lock, or to calibrate the All-Sky Monitor (ASM).

A rather useful byproduct of peeling is the analysis offered in section 2, which leads to a better understanding of its strengths and limitations of selfcal, and helps to develop optimal calibration strategies. It also provides a framework for understanding aspects of traditional selfcal with various existing instruments that were less easy to understand before.

An important result of this analysis is that fewer bright calibrators are needed for the calibration of LOFAR (see also fig 6). Firstly, Station-Station baselines are sensitive enough after all. But secondly, we can get away with considerably fainter calibration sources to estimate the ionospheric beamshapes and phase screens. The reason is that, after acquiring *phase-lock* on the brightest source (using an integration time of 10 sec), the other sources may be kept within a radian by *phase-tracking* the brightest source, which allows us to integrate longer. Our earlier decision to solve for polynomial coefficients, and smoothness in applying the corrections, is crucial here. This approach largely removes the sticky problem of source visibility at baselines longer than 50 km. The tentative conclusion that LOFAR probably does not need Station-Core Selfcal is a big relief, and may have implications for the design of the instrument (e.g. no sub-cores). Of course the central core remains an important observational mode, but it is no longer required for calibration. On the contrary, the core itself should be calibrated with the help of Station-Station Selfcal, and peeling.

It will take some more thinking and experimenting before the advantages and limitations of peeling are fully understood. Therefore, it is fortunate that the LOFAR team can draw on the experience with the WSRT, where there is a strong tradition of delving into the kind of 'luxury' problems that are now a '*conditio sine qua non*' for LOFAR calibration.

Acknowledgements

Some of the ideas described in this document were born, and staggered towards a measure of maturity, during intensive discussions within the LOFAR processing team in Dwingeloo, especially with Michiel Brentjens. Our LOFAR colleagues in Haystack and NRL have been a major source of encouragement to write them up (or down). Many subsequent engagements with members of the international LOFAR team, but also with Ger de Bruyn, George Miley and Wim Brouw, have considerably improved the current version, and will hopefully continue to do so in the future.

References

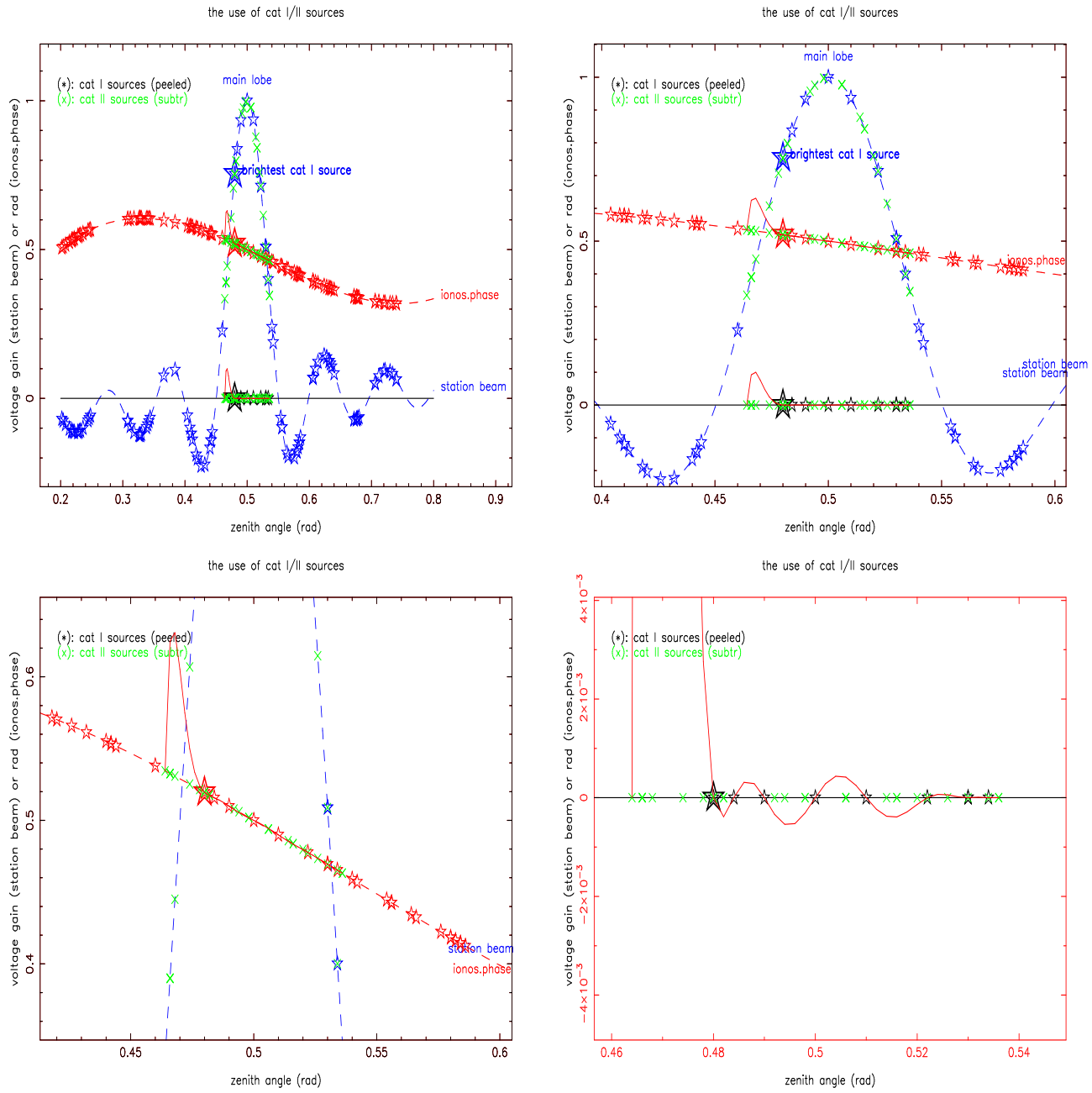


Figure 1: *The peeling process in a nutshell. The 3-5 brightest sources in the main lobe are used to acquire phase-lock, and to phase-track. For each time-freq domain, we 'peel off' 100-500 bright (Cat I) sources one by one, in order of brightness. For each 'peeling source', we solve for a subset of* primary *M.E. parameters, i.e. instrumental and ionospheric parameters in its direction, and then subtracts it from the residual uv-data. The utilised selfcal model only consists of the peeling source itself, and perhaps a few of the next brightest sources. The primary M.E. parameters, which are associated with both a station and a source (direction), are used first of all to subtract the bright peeling sources with the greatest possible accuracy. In addition, the ones associated with sources in the main lobe are also used to estimate* secondary *M.E. parameters which are only associated with a station. (As always, both primary and secondary M.E. parameters are functions of freq and time). Typically, secondary parameters are the coefficients of smooth functions that describe the shape of (the main lobe of) a station voltage beam, or the ionospheric phase screen across it. These are then interpolated to predict and subtract the many fainter Cat II sources before*

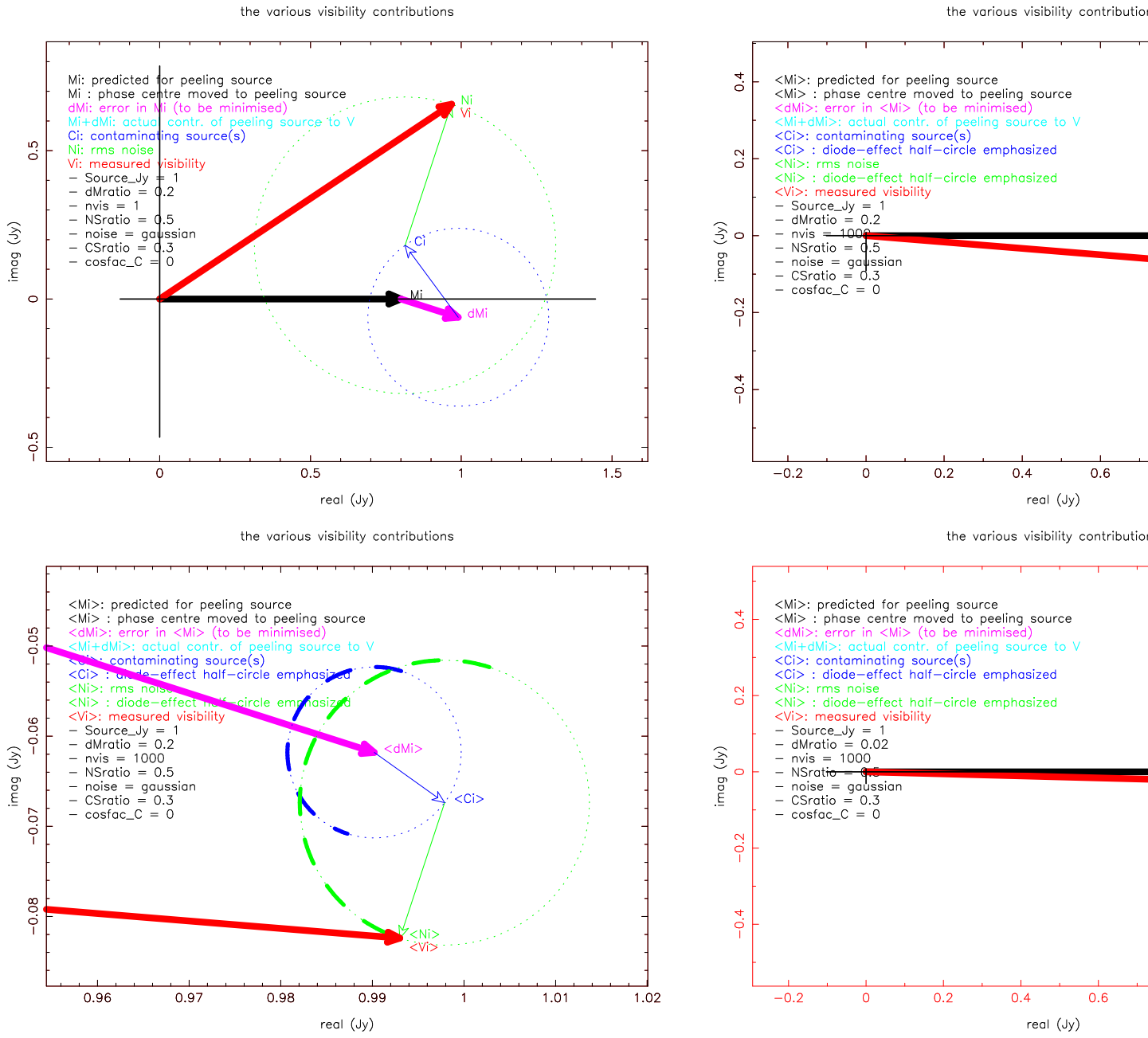
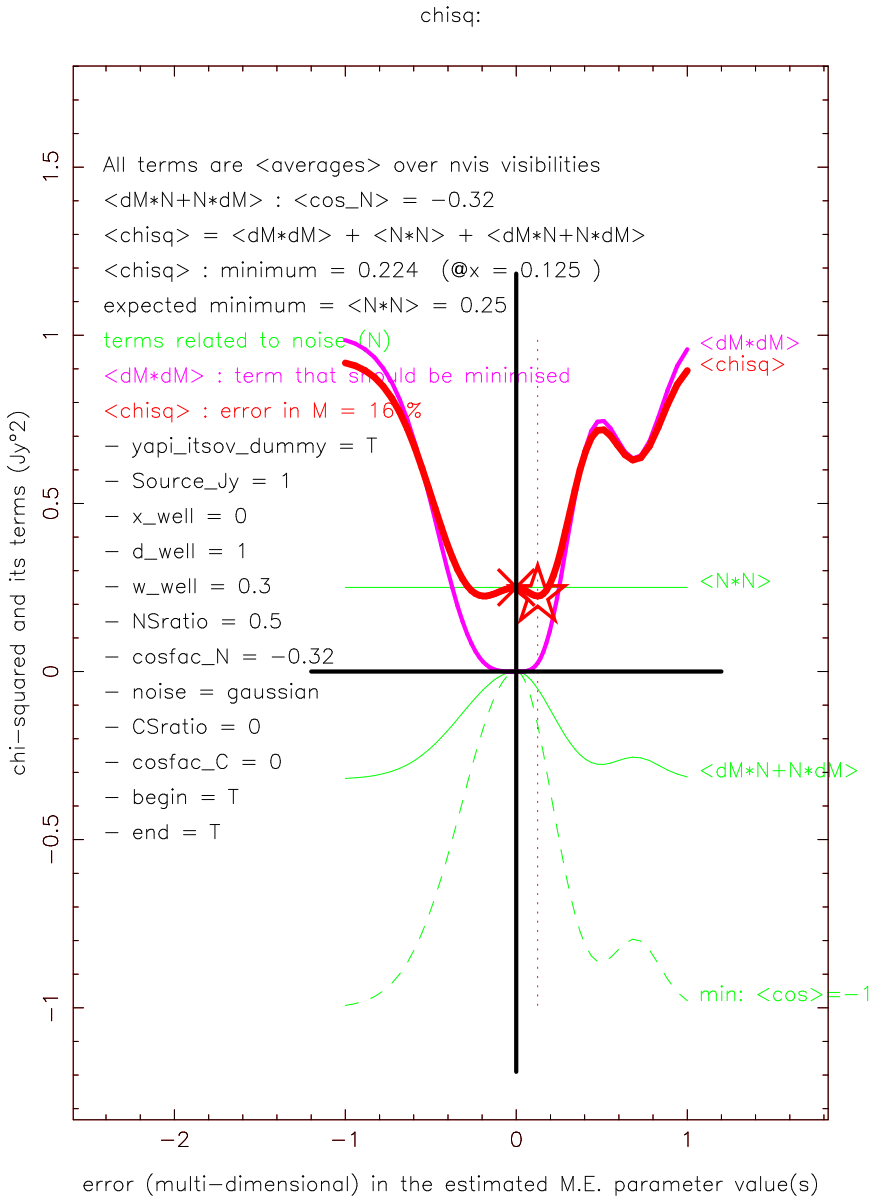


Figure 2: *Pictogram of the various (complex) contributions to the measured visibility* $v_i = m_i + \Delta m_i + c_i + n_i$. *The predicted value* m_i *for the peeling source is real and positive because the phase centre has been moved to its position to minimise the number of domain cells for which prediction (expensive!) is required. The top right panel shows that the relative contributions* n_i *and* c_i *decrease with nvis, the number of visibility samples included in a selfcal solution. The magnified part in the bottom left panel illustrates that* $< \Delta m_i >$ *is not very different from* Δm_i *because it is associated with the peeling source. But the* $< c_i >$ *and* $< n_i >$ *will be smaller and have a different orientation. Their half-circles where their angle with* Δm *will cause* $\cos(\phi_i) < 0$ *have been emphasized. The object is to find those values* $\vec{p} = \vec{p}_0$ *for the subset of parameters that pertains to the peeling source, that minimise* Δm_i *(bottom right panel). The contribution* c_i *is of course the (vector) sum of many contaminating sources.*



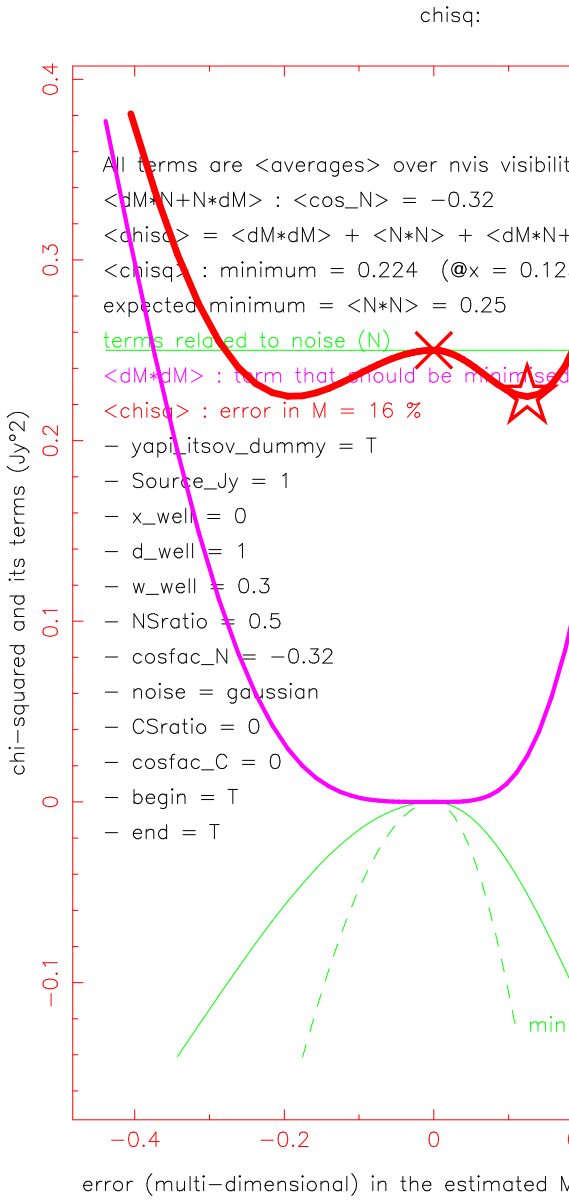


Figure 3: *The influence of noise on χ^2 minimisation, or selfcal solutions. The many-dimensional $\vec{p}-$space, i.e. the space of M.E. parameter values, is represented horizontally in one dimension, and the (arbitrary) function $\Delta m(\vec{p})$ has a minimum of zero at $\vec{p}-\vec{p}_0=0$. The terms described in section 2 are indicated, except the ones depending on c_i, which is taken to be zero for this discussion. The* diode effect *refers to the observation that the term* $<\text{dMxN}> = \sum_i (\Delta m_i^* n_i + n_i^* \Delta m_i) = 2\sum_i \|n_i\|\|\Delta m_i\| \cos(\phi_i)$ only shifts position of the minimum of the χ^2 surface if $\cos(\phi) < 0$. If that happens, the new minimum

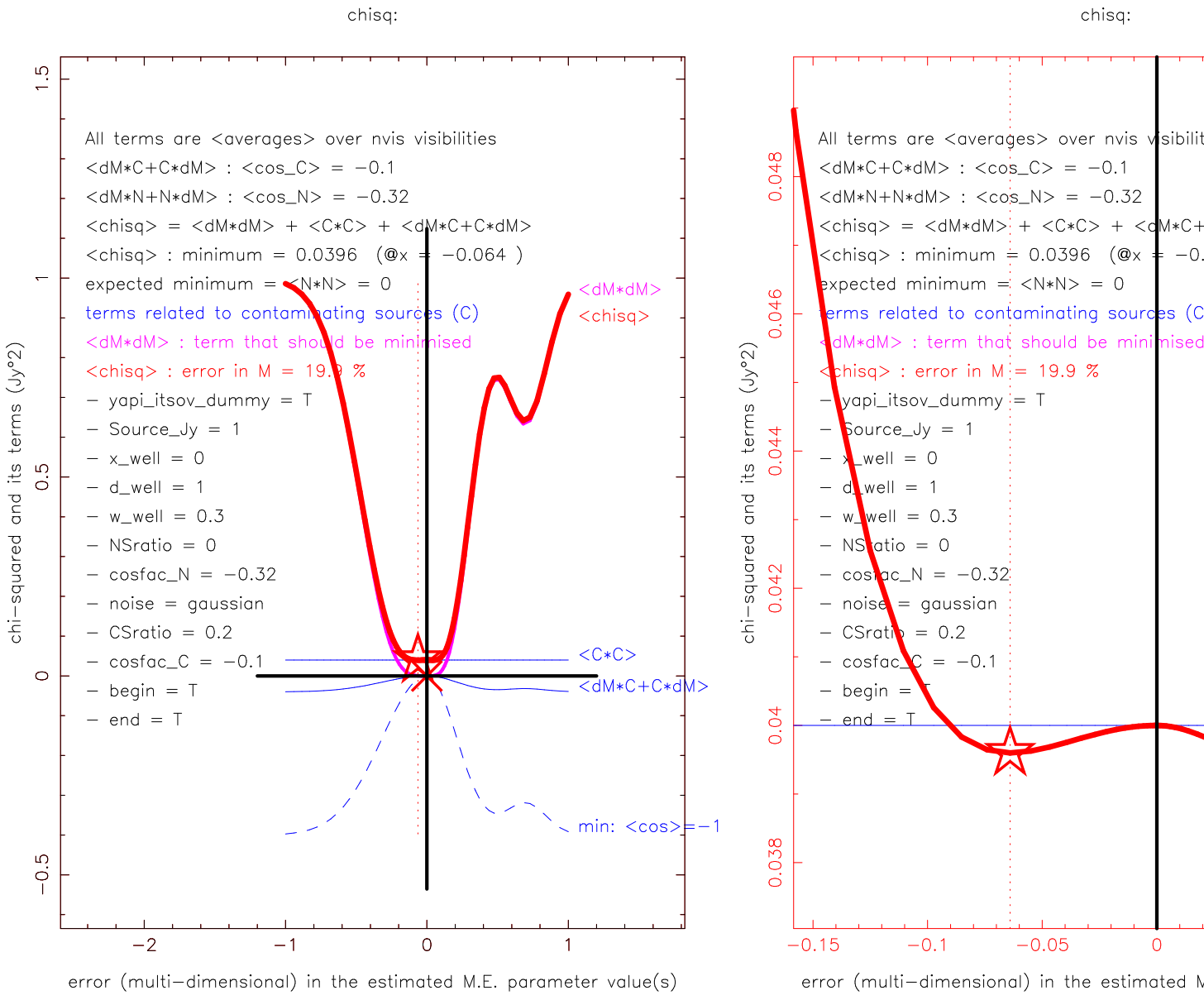


Figure 4: *Influence of contamination sources on a selfcal solution. This is similar to the noise bias, and also causes a 'diode' effect. The only difference is that the sign of the contribution to χ^2 is determined by the instantaneous synthesised beam (PSF), rather than random fluctuations. See fig 5 and its caption for more details.*

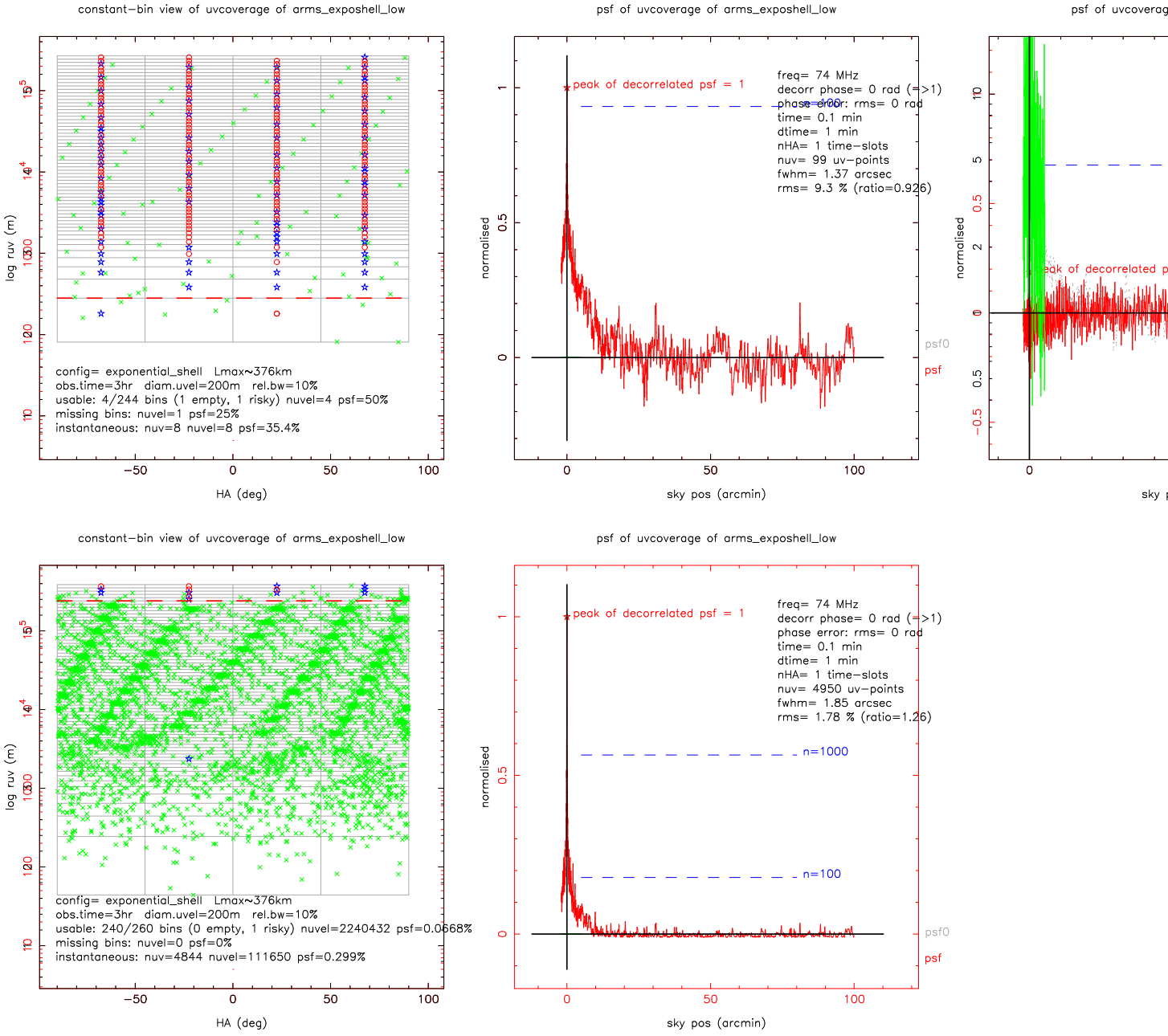


Figure 5: *A somewhat undersampled crossection through the instantaneous synthesised beam (PSF) of LOFAR. As explained in the text, the contribution to χ^2 of a 'contaminating' source is attenuated with the PSF centered at the position of the 'peeling source'. Unfortunately, this is a PSF with rather high sidelobes, because for each M.E. parameter it is the PSF of only the subset uv-samples that depend on that particular parameter. For a station-based M.E. parameter this is equal to the number of stations (top row of panels). For comparison, the bottom row of panels show the instantaneous uv-coverage and PSF of all interferometers.*

Because of the diode effect, the contaminating contribution will only affect the selfcal solution if it is [illegible] would suggest that contaminating sources close to the peeling source, where the PSF is positive, would have little effect. However, the large ionospheric phase errors over the FOV tend to scramble the PSF to something closer to the one in the top-right panel. IN any case, if a contaminating source causes trouble, it can always be included in the prediction.

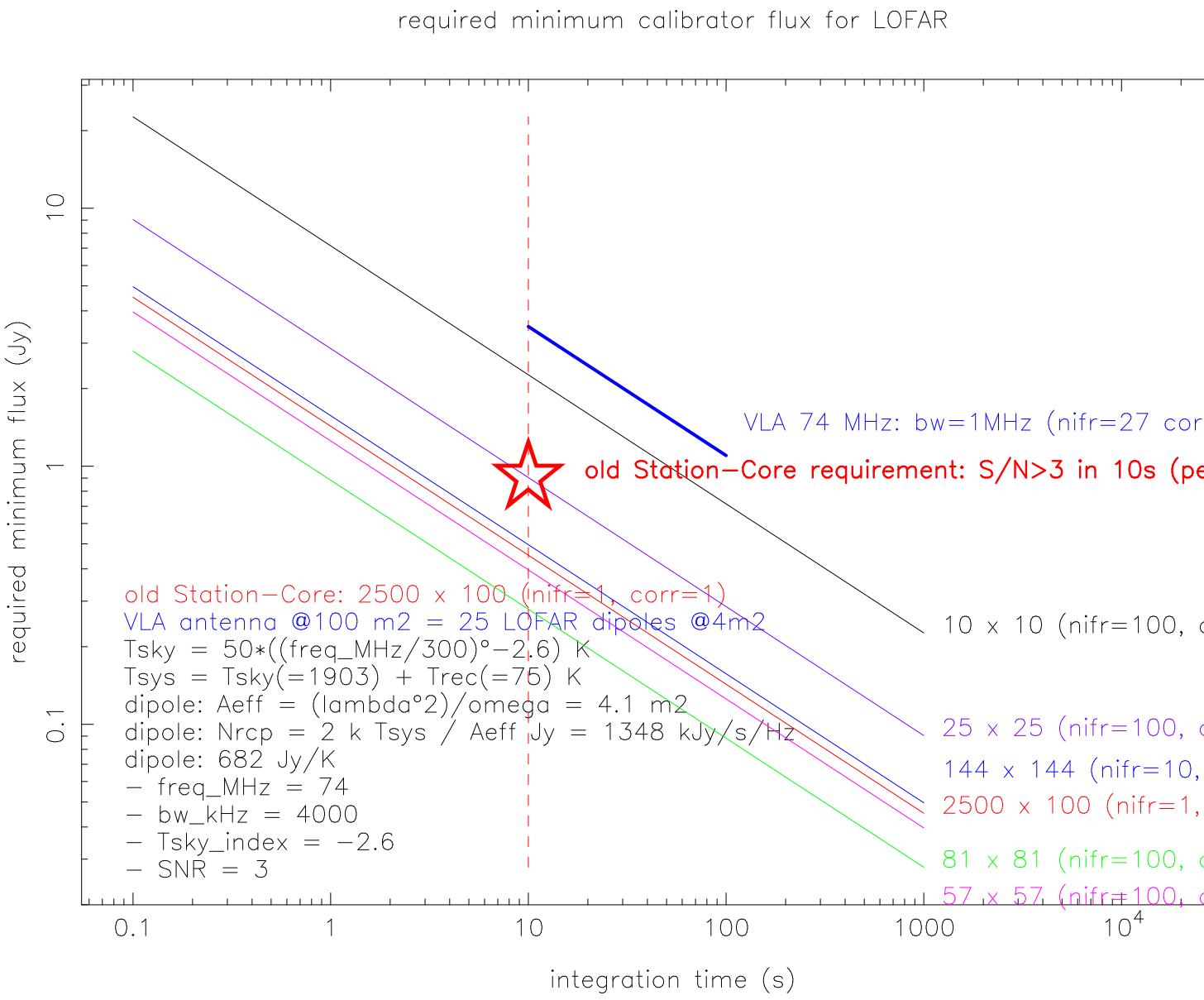


Figure 6: *The required minimum flux of a (point-source) calibrator source for LOFAR. The 'old' requirement for Station-Core Selfcal (SCS) demanded a S/N greater than 3 in 10 s for each visibility, for 20 sources per station beam. This was too severe, for several reasons. Firstly, the S/N may be increased with* $\sqrt{nifr}$*, in which* $nifr$ *is the number of visibilities participating in a selfcal solution that depend on a particular M.E. parameter. In practice,* $nifr$ *will be equal to the nr of stations for station-based parameters. The numbers used are indicated in the plot. Each interferometer contributes four visibilities (correlations). Secondly, only 3-5 sources in the field have to be bright enough to give a solution in 10 s, i.e. the time in which the ionospheric phase changes less than one radian. If the estimated ionospheric phases in the direction of these sources are used to correct the phases in the direction of the other sources, the latter can be kept within a radian of the 'correct' value for a much longer time (up to 100-1000 s, depending on the conditions). Since we are solving for the coefficients of freq-time polynomials, we may integrate for t[illegible] for each solution. As the plot indicates, this makes it possible to use much fainter sources to estimate the station beam-shapes, and the ionospheric phase screens across them. The plot shows the required minimum flux for Station-Core Selfcal (2500x100), and for Station-Station Selfcal for stations with 144, 81, 57, 25 and 10 elements each. One conclusion is that the extra complication of Station-Core Selfcal does not seem necessary for calibration (although the Core remains an important feature of LO-*

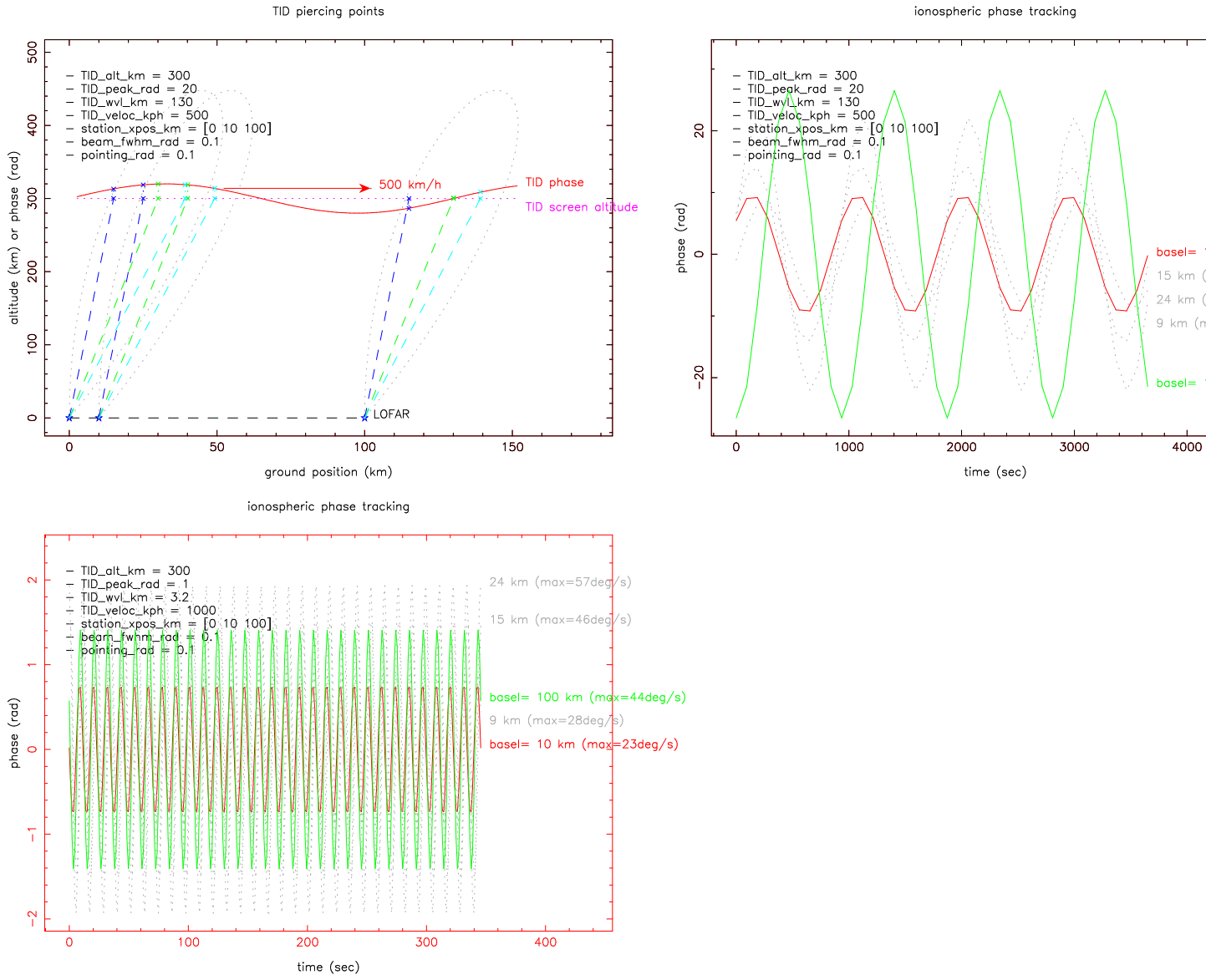


Figure 7: *A Travelling Ionospheric Disturbance (TID) can cause large and rapid phase variations. The somewhat idealised TID in this illustration would reproduce the phase variations that were measured during a 74 MHz VLA observation of Virgo A by Perley and Bust, in which the source moved by about 1 degree, but did not distort very much. Assuming a TID velocity of 500 km/h, the large-scale component has a wavelength of 135 km (much larger than the VLA, but smaller than LOFAR), and a period of 10-15 min (top right panel). Note the large phase differences between sources in the same FOV, whose rays pierce the TID screen at widely separated points. Because of its smooth behaviour, this large-scale component lends itself to phase-tracking. The maximum phase-rate is about one radian in 10 s.*
The small-scale component in the Virgo A phase (botton left panel) is potentially more troublesome. Assumed a velocity of 1000 km/h (?) it has a 'TID' wavelength of 3 km and a period of 10-20 sec. Because it varies too rapidly to be tracked or solved for, it will affect the subtraction of cat II sources, and the imaging of sources. Fortunately, the total error is less than one radian most of the time, and tends to average out over longer periods.

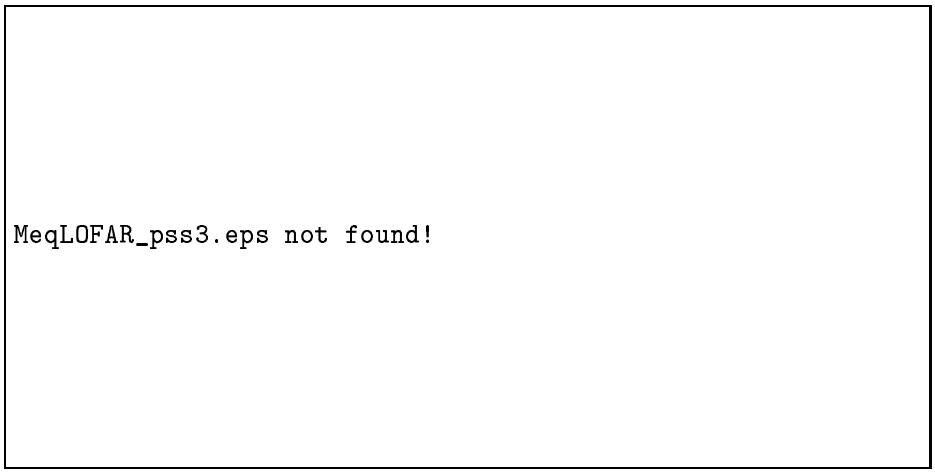


Figure 8: *Implementation of the LOFAR Measurement Equation, using peeling. Note the reduction in the nr of domain cells that is made possible by minimizing the visibility phase variation of the peeling source over the (t,f) domain. This is a combination of shifting the phase center to the position of the peeling source, and correcting for ionospheric phase errors. The application of uv-plane efffects is needed to get residual (patch) images that are properly corrected for at least the center of the patch.*