A Method Based on Wavelet Transforms for Source Detection in Photon-Counting-Detector Images. I: Theory and General Properties

F. Damiani, A. Maggio, G. Micela, and S. Sciortino

Istituto ed Osservatorio Astronomico di Palermo "G.S.Vaiana", Piazza del Parlamento 1, 90134 Palermo, ITALY

E-mail: damiani@, maggio@, gmicela@, sciorti@oapa.astropa.unipa.it

ABSTRACT

We have developed a method based on wavelet transforms (WT) to detect sources in astronomical images obtained with photon-counting detectors, such as X-ray images. The WT is a multiscale transform which is suitable for detection and analysis of interesting image features (sources) spanning a range of sizes. This property of the WT is particularly well suited to the case where the Point-Spread Function is strongly varying across the image, and it is also effective in the detection of extended sources. The method allows to measure source count rates, sizes, and ellipticity, with their errors. Care has been taken in the assessment of thresholds for detection, in the WT space, at any desired confidence level, through a detailed semi-analytical study of the statistical properties of noise in wavelet-transformed images. The method includes the use of exposure maps to handle sharp background gradients produced by a non-uniform exposure across the detector, which would otherwise yield many spurious detections. The same method is applied to evaluate upper limits to the count rate of undetected objects in the field of view, allowing a sensitivity map for each observation to be constructed.

Subject headings: Methods: Data Analysis – Methods: Statistical – Techniques: Image Processing

1. Introduction

In most astronomical images a large fraction of sources are near the detection limit. A careful statistical treatment is then needed to determine their existence and their properties, such as accurate position, flux, size, etc. Furthermore, the reliability of the source detection method has implications for the assessment of source statistical properties, such as their luminosity function. It is therefore important to have a detection method which is both efficient for detecting sources in the field of view, and reliable in the sense that it yields only a predefined number of expected spurious detections.

A case of particular importance for X-ray, γ-ray, EUV and some UV detectors is the photon-counting regime, where the information for most sources has to be extracted from a handful of photons, and the photon statistics per image pixel is significantly different from gaussian. In this case, methods for source detection commonly used for optical images are no longer effective. Moreover, many kinds of astronomical objects are known, and some of them are observable as spatially extended sources. Examples are supernova remnants in Local Group galaxies (including the Milky Way), galaxies themselves, and clusters of galaxies; in the two latter cases the emission may be superimposed on that of other point sources. An ideal detection algorithm should be able to recognize these cases and deal properly with them.

Among all photon-counting detectors, present-day X-ray detectors (such as the PSPC and HRI on ROSAT, and the forthcoming EPIC camera on XMM) have only limited and non-uniform spatial resolution, which of course makes the analysis of X-ray images more complex. X-ray images with significantly higher resolution will be obtained with the ACIS and HRC detectors on AXAF, but the detection problems will be similar, albeit dealing with smaller angular sizes. The instrument which has already collected the largest amount of spatial X-ray data is the PSPC onboard ROSAT: its Point Spread Function (PSF) is bell-shaped with a FWHM as wide as $\sim 20''$ on axis, and becoming gradually degraded (i.e. wider) farther from the optical axis, reaching a width of a few *arcmin* at the image edge (Hasinger et al. 1993). In spite of this degradation, PSPC images still contain a considerable amount of useful information at off-axis angles in the range 20-40 arcmin, where neither vignetting nor the PSF widening are too severe. However, most current source detection algorithms for the PSPC do not efficiently detect sources so far off-axis; once again, therefore, a more versatile tool for X-ray image analysis is desirable.

In this paper we explore the capabilities of a relatively new method of source detection, based on wavelet transforms (WT). For an introduction to wavelet transforms, see e.g. Daubechies (1992) or Kaiser (1994). In general, the wavelet transform is a multi-scale transform providing a representation of the data which allows easy extraction of both the *position* and the *shape* of features, in an image or in a lightcurve. This is unlike classical Fourier analysis, which provides only shape information, and also unlike simple object-finding methods looking for peaks (maxima) in the original data, which provide only positional information. Analogous exploration of a wavelet-based X-ray source detection have been already presented by Rosati, Burg & Giacconi (1994), and by Grebenev et al. (1995); a similar technique ("matched filter") has been employed by Vikhlinin et al. (1995). In the present work, many improvements have been brought to this technique, that include a better quantified statistics of noise in WT images, the use of exposure maps to handle spatial exposure nonuniformities, the computation of upper limits using the WT, and a careful treatment of errors on flux and size of detected sources, among other things.

Although the WT has wider application in the temporal and spectral domains, we will limit ourselves to characterize the spatial properties of each detected object simply by deriving a single number as a measure of its size. A companion paper (Damiani et al. 1996, hereafter Paper II) presents extensive tests of the performance of our algorithm for the case of ROSAT PSPC X-ray images.

In section 2 we outline the basic characteristics of our method. Then, in section 3 we study the statistical properties of the wavelet transform, which allow us to establish confidence levels for the existence of detected sources. Later (section 4), we present a complete description of our algorithm, for both detecting sources and evaluating upper limits on a list of positions. Section 5 is a summary of the results obtained.

2. Properties of the Method

The wavelet transform of a data set allows the selection of particular features of interest in the data. Various kinds of wavelet transforms exist, depending on what data properties are being investigated, and they rely on different "generating wavelets"; common to all types of generating wavelets is their dependence on position and length scale, and also their having zero mean and finite norm. In the case of images in two or more dimensions, there may be as many independent scale parameters as the number of dimensions.

The WT of a two-dimensional image $f(x, y)$ is defined as:

$$w(x,y,a) = \iint g\left(\frac{x-x'}{a}, \frac{y-y'}{a}\right) f(x',y')\ dx'dy' \tag{1}$$

where $g\left(\frac{x}{a}, \frac{y}{a}\right)$ is the generating wavelet, x and y are the pixel coordinates and a is the scale parameter. In our case, $f(x, y)$ describes the number of photons recorded in each pixel of the image; we are interested in detecting discrete sources of emission, having in most cases a nearly bell-shaped spatial profile, and therefore we choose for g a function having a similar shape (if we had interest to find a localized periodicity in the data, we would have chosen an oscillating g, and so on). An useful property of every function g is that it is a *localized* function, namely it is appreciably different from zero only in a finite neighborhood of the position (x, y) where it is evaluated. Therefore, the value of $w(x, y, a)$ defined in equation (1) will be affected only by the values of $f(x, y)$ in positions within a few times the scale parameter a. The evaluation of $w(x, y, a)$ at different scales a permits the study of structures in the data $f(x, y)$ of various spatial sizes. The choice of a unique scale parameter for both x and y axes arises from the assumption that most sources are *isotropic*, and may be relaxed to study anisotropic sources (see section 4.7).

Throughout this work, we use the two-dimensional "mexican hat" wavelet (see Figure 1):

$$g\left(\frac{x}{a}, \frac{y}{a}\right) \equiv g\left(\frac{r}{a}\right) = \left(2 - \frac{r^2}{a^2}\right)\ e^{-\frac{r^2}{2a^2}} \qquad (r^2 = x^2 + y^2) \tag{2}$$

which has all properties required for a generating wavelet, and is significantly different from zero only for r less than a few times a (e.g. $|g| \geq 10^{-4}$ for $r \leq 5\ a$). Our definition has a slightly different normalization (by a factor a) from that given by other authors (e.g. Rosati et al. 1994), but is more convenient because g is dimensionless.

This chosen g ensures that the WT of a function $f(x,y) = c_1 + c_2x + c_3y$ (a tilted plane) is zero. Therefore, the WT will be zero for both a constant or uniform gradient local background (within $\sim 5a$). While the WT is not sensitive to gradients in the data with this choice of g, it is sensitive to *second derivatives* of $f(x,y)$, e.g. local maxima or minima, or even "step functions", making it suitable for detecting sources in an image (local maxima of $f(x,y)$). This is clear if we note that (integrating by parts, and provided that the function $f(x,y)$ goes to zero at infinity)

$$w(x,y,a) = \iint \left(2 - \frac{r^2}{a^2}\right) e^{-\frac{r^2}{2a^2}} f(x',y')\, dx'dy' = -a^2 \iint e^{-\frac{r^2}{2a^2}} \left(\frac{\partial^2 f}{\partial x'^2} + \frac{\partial^2 f}{\partial y'^2}\right) dx'dy' \quad (3)$$

where $r^2 = (x - x')^2 + (y - y')^2$.

Unfortunately, second derivatives can arise from instrumental artifacts, for example due to shadows cast by the entrance window supporting structures or "ribs" in proportional counters, or inter-chip gaps in CCD-based detectors. We deal with this differently in section 4.2.

Let us consider a source with N_{src} total counts and with a gaussian shape of width σ_{src}, namely $s(r) = \frac{N_{src}}{2\pi\sigma_{src}^2} e^{-\frac{r^2}{2\sigma_{src}^2}}$; its wavelet transform is (see Figure 2):

$$w(r,a) = \frac{N_{src}}{\left(1 + \frac{\sigma_{src}^2}{a^2}\right)^2} \left(2 - \frac{r^2}{a^2 + \sigma_{src}^2}\right) e^{-\frac{r^2}{2(a^2+\sigma_{src}^2)}}, \quad (4)$$

which, as a function of r, has the same form as the generating wavelet $g(r/a)$, provided that a is replaced by the quadrature sum $\sqrt{a^2 + \sigma_{src}^2}$. The peak of this function occurs for $r = 0$, where the wavelet transform amplitude depends on a as:

$$w_{peak}(a) = \frac{2N_{src}}{\left(1 + \frac{\sigma_{src}^2}{a^2}\right)^2}. \quad (5)$$

Such dependence allows to derive the width σ_{src} of a gaussian source once its WT has been computed at several scales a. It is useful to understand the asymptotic behaviour of $w_{peak}(a)$ (henceforth denoted simply as $w(a)$) as the ratio σ_{src}/a assumes very small or very large values. If the WT is computed at a scale a much larger than the source size ($\sigma_{src}/a \to 0$) the source shape is "seen" as a δ-function by the transform, and therefore, for a fixed source width σ_{src}, $w(a)$ approaches a finite value $w(a) = 2N_{src}$ as $a \to \infty$. If the source is much larger than the region where the generating wavelet is significantly different from zero ($\sigma_{src}/a \to \infty$), the transform sees the source as an almost flat background, and $w(a)$ tends to zero as $(\sigma_{src}/a)^{-4}$. This property of the WT turns out to be very useful if the studied image contains a superposition of sources of different sizes (e.g. point and extended sources), since the transform $w(a)$ for scales comparable to the size of the smallest sources will be affected very little by the underlying larger source, which is seen as only an enhanced background emission.

If we consider the function $y(a) \equiv w(a)/a$, we see that this tends to zero for both $a \to 0$ and $a \to \infty$ (for a fixed σ_{src}), as a^3 and a^{-1}, respectively (see Figure 3). This function has therefore a maximum for a finite value of a, namely

$$a_{max} = \sqrt{3}\sigma_{src}, \quad (6)$$

allowing us to measure the source width σ_{src}. The existence of such a maximum of $y(a)$ explains why we stated that the WT at a given scale is particularly sensitive to structures of about that size. The value of the transform for $a = a_{max}$ allows an estimate of source counts, since $w(a_{max}) = (9/8)\ N_{src}$.

Sources in the image $f(x, y)$ are therefore defined as local maxima in the function $w(x, y, a)$, at a given scale, which cannot be due to background fluctuations with a certain assigned probability level.

3. Statistical Properties of Noise in Wavelet-Convolved Images

In order to establish the statistical significance of a candidate source detection we need to know the statistical properties of the WT, and in particular of the probability distribution $P(w)$ of the WT values due purely to the fluctuations of a uniform background. While the mean value of $P(w)$ is necessarily zero since $g(r/a)$ has zero mean, the distribution variance increases with the image background intensity.

We derive in detail in Appendix A the probability distribution of WT values resulting from wavelet-transforming a spatially uniform photon background, with Poisson noise. Such a distribution $P(w)$ is not easy to compute. Other authors (e.g. Starck et al. 1996, and references therein) make use of an approximate treatment of the statistics of the WT, based on the transformation of the analyzed image into another one with unit variance. This is not suitable when there are only a few photons per image pixel.

The probability distribution $P(w)$ depends on both the background density b and the wavelet scale a, but only through the dimensionless combination $q \equiv b \times a^2$ (number of photons per square scale, or "photon density"), as is easy to understand, since the generating wavelet g only depends on the ratio r/a. Correspondingly, we denote as $P_q(w)$ the probability distribution of the WT noise. In the derivation of the distributions $P_q(w)$ (Appendix A) we will also make use of the number n of photons falling within a radius $5a$, related to q by $n = 25\pi q$. The probability distributions $P_n(w)$ for n=4, 16, 256, and 4096 (q=0.051, 0.204, 3.26, and 52.1 respectively) are shown in Figure 4. As it may be seen from this Figure, $P_n(w)$ is highly asymmetrical for low n ("discrete photon limit"), while it approaches a gaussian distribution (also shown in Figure 4) as soon as n becomes at least a few hundreds ("continuous limit"). The transition between the two limits is very smooth: to be safely in the continuous limit requires that n is as large as *a few* $\times 10^4$.

The higher quantiles of the distributions $P_q(w)$ define thresholds w_0 for acceptance of a measured WT amplitude $w \geq w_0$ as not due to background fluctuations at a certain confidence level. On the basis of the distributions $P_q(w)$ derived through MonteCarlo simulations (Appendix A), we have looked for a convenient parametrization of the threshold w_0 for fluctuations of w at a probability level corresponding to $k\sigma$ of a gaussian distribution, expressed as a function of k and q, $w_0 = w_0(k, q)$. To do this, we have noted that the difference between the values of w_0 yielded by our simulations at each k (= number of gaussian σ) and the corresponding continuous-limit value $k\sigma_w = k\sqrt{2\pi q}$ is nearly a constant, independent of q, say $\Delta w(k)$, for $-1 \leq \log q \leq 4$ (a range comprising the vast majority of values we have found in present-day X-ray images); then, we have found empirically that $\Delta w(k)$ is very well approximated by a quadratic polynomial in k, $\Delta w(k) = c_1 + c_2 k + c_3 k^2$ (with $c_1 = -0.2336$, $c_2 = 0.0354$, and $c_3 = 0.1830$), in the range of k that could be explored with our simulations, namely $0.5 \leq k \leq 5.5$. Therefore, we approximate

$$w_0(k,q) = k\sqrt{2\pi q} + (c_1 + c_2 k + c_3 k^2), \tag{7}$$

shown as dotted lines in Figure 5, which is seen to reproduce the simulation results (dots) quite well. This approximation may also be used to estimate the threshold w_0 at still higher k by extrapolation, which however is not essential for source detection purposes since usually a threshold between $3-5\sigma$ is sufficient to yield very few spurious detections. The formula given in equation (7) predicts correctly that for fixed k the probability levels approach those of a gaussian distribution for increasing q, i.e. in the continuous limit. Finally, for a fixed q, and assuming that it holds for all k, equation (7) may be inverted to yield the "significance level" k as a function of w_0 and q, i.e. the confidence level in terms of gaussian σ's that the given local excess (detection) can be generated by a random background fluctuation.

4. The Algorithm

We now describe the steps which our algorithm goes through to produce a final list of detected sources (Figure 6). After having computed the exposure map for the given observation (step 1), the reference background map is computed by a suitable smoothing of the image (step 2). Then, a local median filter is applied to recompute the background at every scale of a set a_k (step 3), to minimize the effect of point sources on the background determination. The wavelet transform $w(x,y,a)$ of the image $f(x,y)$ is then computed on a grid of positions (x_i,y_j) and for each scale a_k (step 4). Spatial maxima (peaks) of $w(x_i,y_j,a_k)$ are selected if their heights are above the expected background at a chosen significance level (step 5). Problems arise near underexposed regions in the studied images (such as the PSPC "ribs"), since the apparent sharp background gradient they generate can be mistakenly detected as a source by the wavelet method. To solve this problem we use appropriate exposure maps (section 4.2).

For each scale a, groups of detected sources lying closer than a minimum distance are merged into a single source. After having built a list of sources detected at all scales a, a cross-identification of sources detected at consecutive scales is made, so that $w(x_0,y_0,a)$ may be evaluated as a function of a for a given source lying at (x_0,y_0), at least for those scales a where the source is detected. This allows us to derive the intensity and extent of the source (step 6).

Having determined the scale a_{max} yielding the maximum significance for a given detection, we refine the spatial grid where we search the maximum of $w(x,y,a_{max})$, to derive a more accurate position (step 7). Since source positions may have slightly changed after this best-position step, we repeat the source merging and cross-identification steps on the new source positions. At the improved positions, source parameters are more accurately derived (step 8).

At this stage, we have an almost final list of detections. However, we have found that, in some (rare) cases, a weak source is not detected when another stronger source lies nearby (but the two are still clearly resolved by eye), while comparably weak sources are invariably detected if they are isolated; the reason for this is that the effect of strong sources on the computation of the local background level is difficult to screen out completely, and therefore the local detection threshold is raised with respect to its true value, causing the non-detection of weak sources. Since all strong sources are detected at this stage, we now correct the background maps eliminating their effect (by interpolation over their positions), and iterate a second time the whole procedure described above with the improved background (step 9). The background map used in the second iteration is therefore more like the "reference map" used by existing "map-detect" methods (cf. Harnden et al. 1984, Zimmermann et al. 1993), while the first iteration makes use of a reference background more like a "local-detect" method. One can iterate further times in the same way, until no more sources are detected.

Eventually, the algorithm produces a list of all sources in the input image detected above the chosen significance level (step 10). For each source we are able to estimate its position, count rate and size, with their errors. The accurate evaluation of errors on these quantities is not straightforward, and is discussed in section 4.5. Following a procedure entirely analogous to source detection, we estimate also upper limits to the count rate of undetected objects in the field of view. Finally, we apply a non-isotropic wavelet transform to extended sources (i.e. larger than the detector PSF), to derive their ellipticity and position angle.

We describe in detail all steps outlined above in the following subsections.

4.1. Computation of Background Maps

A necessary preliminary step to source detection is the evaluation of the reference background, that defines detection thresholds. This has to be computed *locally*, rather than as a global value in the image, since it may vary significantly across the image, both because of the detector vignetting and obscuring structures, and more importantly because of real variations of the sky background (extended sources, cosmic background gradients). The adoption of a uniform sky background value would lead to detection of false sources where the local background is higher than the average, as preliminary simulations confirm (Figure 7), and to lack of detection of real sources where it is lower than average.

The local background value should be computed with the least possible uncertainty, and should be ideally unaffected by point sources. In order to attempt to satisfy both requirements, we compute the reference background in two steps: the first, common to all scales a, is the generation of a smoothed background map, while in the second step a local median of this background map is computed, on a region of size increasing with the scale a at which the WT is being evaluated, in order to reduce to a minimum the influence of sources on background evaluation.

The construction of the initial background map is intended to follow spatial variations of background as closely as possible (including sources), and with large enough photon statistics to yield only a small error on the computed background. Since the smallest lengthscale on which sky background gradients can appear is set by the local PSF width $\sigma_{\rm PSF}$, we compute the background map by smoothing the original image $f(x, y)$ with a gaussian whose width is $\sigma_{smo} = 2\sigma_{\rm PSF}$ *in each point*, i.e. narrower at the center and increasingly wider farther off-axis. A larger smoothing gaussian width σ_{smo} is used in those points where there are too few photons entering smoothing, until a sufficient number of photons are available in an area σ^2_{smo} (25 photons in our implementation, that should ensure a minimum S/N ratio of about 5).

Where there are strong gradients in the exposure map, because of detector artifacts (e.g. PSPC "ribs"), their width in the image is independent of the local PSF, and so should not be included in the smoothing. We interpolate over detector artifacts, assigning relative gaussian weights to the smoothed points:

$$\bar{b}(x_0, y_0) = \frac{\displaystyle\int_{S'} f(x,y)\, e^{-\frac{r^2}{2\sigma_{smo}^2}}\, dx\, dy}{\displaystyle\int_{S'} e^{-\frac{r^2}{2\sigma_{smo}^2}}\, dx\, dy}, \tag{8}$$

where $\bar{b}(x_0, y_0)$ is the smoothed background in the point (x_0, y_0), $\sigma_{smo} = 2\sigma_{\rm PSF}$ is the width of the smoothing gaussian, and $r^2 = (x - x_0)^2 + (y - y_0)^2$; the integration domain S' excludes all points where the exposure time is less than 0.8 times the value outside ribs, affected only by vignetting. Exposure times for each pixel are deduced from properly computed exposure maps.

The final background value in a point, as mentioned above, is obtained at each scale a by computing the median of the background map over a square region of side $l = 4\sqrt{a^2 + \sigma_{\rm PSF}^2}$ centered on the given position. The median filter screens out the highest values (sources) of the background map, unless a source is very bright and affects a larger area of the background map. The scale dependence of the size of the region where the local background median is computed is in the spirit of the scale-invariant approach that characterizes wavelet transforms. We note also that this procedure may cause an extended source to be considered as background at some scales, smaller than its size, while the same source will be screened out from background evaluation at larger scales, comparable to the source size, as appropriate in order to detect it at these latter scales.

We stated that weak sources in the neighborhood of stronger ones are occasionally not detected by our algorithm. The reason for this lies in an overestimate of the background, not in the wavelet transform method itself, because residual influence from the strong nearby sources could not be eliminated.

Therefore, a correction is made after the first-stage detection process (see sections 4.2-4.5), excluding circles around point sources including 95% of source photons (radius $\sim 2.5\sigma_{\rm PSF}$ for the PSPC), and interpolating with equation (8). Sources found as extended ($\sigma_{src} > 2\sigma_{\rm PSF}$) are not interpolated over, since they effectively behave as background for overlying point sources. We then iterate the detection procedure after recomputing the background using the improved background maps.

It might be argued that a weak, isolated point source, just above the nominal threshold for the local "true" background, may remain undetected in the first iteration since it raises the local background estimate, derived as above; the same source will neither be detected in the second iteration, where the background is corrected only near detected sources. We have computed that the effect of such a barely detectable source on our local background evaluation is typically of order of 10%, and the source significance will thus be lowered by 5% (see below). However, this does not reduce the algorithm ability to detect weak sources, for two reasons. First, the same effect helps to reject positive background fluctuations as candidate sources. Second, the algorithm's threshold significance levels for detection have not been chosen from *a priori* criteria, but such as to yield an assigned number of spurious detections on the basis of the performances of the whole algorithm on simulated data (Paper II), and therefore the slight inaccuracy in our background evaluation does not reduce the detection efficiency, since it is compensated by a readjustment of the detection threshold. Namely, if the background computation method is self-shielding against detection of spurious fluctuations, the threshold can be lowered with respect to its "nominal" value, and weak (real) sources, that would otherwise be lost, can thus be detected.

We estimate the effect of a weak source on background median evaluation, as follows. The background is computed as the median over a square region of side l, defined above, of the smoothed background map. This latter is (locally) a function of the form

$$\bar{b}(r) = b + \frac{N_{src}}{2\pi\sigma_{tot}^2} e^{-\frac{r^2}{2\sigma_{tot}^2}} \tag{9}$$

arising from a local (uniform) background of density b, plus a point source with N_{src} counts, at the center of the square ($r = 0$). After smoothing (equation 8), this source has a width $\sigma_{tot} = \sqrt{\sigma_{smo}^2 + \sigma_{\rm PSF}^2} = \sqrt{5}\sigma_{\rm PSF}$. Since the function $\bar{b}(r)$ is a monotonically decreasing function of r, its median over the square is simply the value of $\bar{b}(r)$ at a radius R encircling half of the square area: $\pi R^2 = l^2/2$. Now we choose a detection threshold of 4σ, assume to be in the continuous-limit approximation, and set the detection scale $a = a_{max} (= \sqrt{3}\sigma_{\rm PSF})$, as appropriate for marginally detected sources (section 4.6). In this case, $N_{src} = (8/9) w(a_{max}) = (8/9) w_0 = (8/9) \cdot 4\sqrt{2\pi b a_{max}^2}$, since for a marginally detected source, the wavelet amplitude $w(a_{max})$ is equal to the threshold w_0 (equation 7). We have also $l = 4\sqrt{a_{max}^2 + \sigma_{\rm PSF}^2} = 8\sigma_{\rm PSF}$, and therefore

$$\bar{b}_{median} = \bar{b}(R) = b\left(1 + \frac{16}{45}\sqrt{\frac{6}{\pi}} \frac{e^{-\frac{16}{5\pi}}}{\sqrt{b\sigma_{\rm PSF}^2}}\right) = b\left(1 + \frac{0.177}{\sqrt{b\sigma_{\rm PSF}^2}}\right). \tag{10}$$

The quantity $b\sigma_{\rm PSF}^2$ is the number of background photons in the source (core) area, and is larger than one in most cases. As an example, in the ROSAT PSPC case, for a source $15'$ off-axis, exposure time of 5 ksec and average background, one obtains $\bar{b}_{median} = 1.137 \times b$, i.e. only a slightly inaccurate background estimate, as mentioned above. We also see that the background estimate will be worse for a very narrow PSF, and we discuss in Paper II that indeed this will be a major difficulty for detecting weak sources using low-background, very high-resolution detectors.

4.2. Wavelet Transform Computation and Source Detection

This step is at the heart of our detection method, yet it is maybe the simplest part of the whole procedure. The WT of the image is computed at a set of scales a, ranging from about $\sigma_{\rm PSF}$ at image center, to a few times $\sigma_{\rm PSF}$ at image edge, in (logarithmic) steps of $\sqrt{2}$, in order to cover the range of apparent sizes of most sources in the image (including moderately extended sources). Since real features in the image (apart from ribs) cannot be smaller than the PSF in size, the WT at a scale a is computed only on image regions where $\sigma_{\rm PSF} \leq a$. This choice of the region to be analyzed at each scale a, moreover, results in a reduction by an order of magnitude of the needed computing time, with respect to full-image analysis. The largest scale used is about one fifteenth of the field of view radius, since the approximation of uniform background is often not valid for larger scales. The spacing between points of the spatial grid where the WT is evaluated is chosen as $a/2$, since a grid spacing much smaller would only oversample the data with no gain of information. The WT is computed applying directly equation (1). We have found this computational method more convenient than computing the convolutions using FFT's, because the function $g(r/a)$ is localized and symmetric, and because the analyzed region radius and grid spacing change with scale a.

Both the image edges and detector artifacts cause the appearance of spurious ripples in the WT, which should be corrected for before source detection. The most important difference between image edge and artifacts is that exposure time drops to exactly zero beyond the edge, while it usually does not below artifacts; the obscuration due to artifacts is not total partly because of the telescope motion during the observation ("wobbling"), and also because they intrinsically turn out not to block all incident photons, even in the absence of telescope wobbling. This causes intense sources to appear even below artifacts, and simply ignoring image regions obscured by ribs would therefore waste useful data (e.g. about 10% of a PSPC image); details on how we deal with partially obscured sources are presented in Appendix B.

A correction is easily applied near the image edge. The edge is sharp compared to the local PSF, and therefore behaves like a step function (in the radial direction),

$$f_{step}(r) = \begin{cases} 1 & r \leq r_{edge} \\ 0 & r > r_{edge} \end{cases} \tag{11}$$

(r is a radial coordinate and the edge occurs at $r = r_{edge}$) whose wavelet transform, in the limit $a \ll r_{edge}$, is:

$$w_{step}(r - r_{edge}, a) = -\sqrt{2\pi}\; a\; (r - r_{edge})\; e^{-\frac{(r-r_{edge})^2}{2a^2}}. \tag{12}$$

This is positive for $r \leq r_{edge}$, meaning that a positive ripple in the WT appears inward of the edge over a lengthscale of order a, i.e. in zones where we look for sources. If the background at the edge is b, the WT ripple is $b \times w_{step}(r - r_{edge}, a)$. We therefore subtract this term from the previously computed WT of the image at each scale a, preventing the formation of WT peaks which would lead to detection of many spurious sources along edges.

Analogously, the WT of a (idealized) rectangular rib is the sum of two terms similar to equation (12), with two ripples along the rib border. As simulations show clearly (Paper II), a large number of spurious sources appear due to this effect, if no correction at all is applied.

Our solution which retains the information content of the whole image, is to apply the WT to the *count-rate image*, rather than to the *photon image*. The count-rate image is obtained dividing the photon image by the exposure map, pixel by pixel. Working on the rate (or flat-fielded) image, no appreciable background discontinuity nor WT ripples appear along ribs; the edge correction is still applied as before. Of course, the calculation of WT fluctuations has to be suitably scaled for the flat-fielded image, since in this latter the pixel statistics is no longer Poissonian. The amplitude of the count-rate WT is therefore converted into an "equivalent" photon WT, by a suitable "effective" exposure time t_{eff}. The computation of t_{eff}, which is critical for sources partially affected by detector artifacts, is discussed in Appendix B.

At each scale a, the WT of the rate image is then searched for the presence of peaks, or local maxima, defined as those pixels (x, y) where the transform $w(x, y, a)$ is larger than that in the eight neighboring pixels. A peak is retained as a positive detection if its amplitude is significant with respect to the expected fluctuations of the local background (computed as in the previous subsection), at a given confidence level $n\sigma$, as discussed in section 3. n is chosen to minimize the number of spurious detections while retaining good detection efficiency for real sources, and has to be tuned using extensive simulations for the particular detector under consideration (see Paper II). The threshold significance n cannot be determined from *a priori* considerations alone, since it is affected by many complex effects, as e.g. cross-correlation between background fluctuations at different scales, and the effect of sources on background evaluation, in both the first and second iteration. In this initial source search, the detection threshold is actually set to a slightly lower value $n_2\sigma$ ($n_2 = n - 0.2$), because the adopted spatial grid might incompletely sample some peaks of $w(x, y, a)$, yielding detections with apparent significance lower than the actual value.

4.3. Source Merging and Cross-Identification

The spatial grid where the WT at a given scale a is evaluated is chosen to avoid the oversampling of the wavelet transform peaks. Since photon noise may give rise to multiple peaks within $\sigma_{\rm PSF}$ we merge all detections within $d_{max} = 1.5\sigma_{\rm PSF}$ of the position of a given detection at each scale a (see Appendix D for the choice of d_{max}). The position and wavelet-transform amplitude $w(x, y, a)$ of the resulting merged source are those of the maximum value of $w(x, y, a)$ for that group of detections.

We then cross-identify sources found at different scales, in order to estimate the profile of the WT peak amplitude $w(a)$ for all scales, to compute source counts and extent with their associated errors. This cross-identification is done by matching the positions of sources detected at different scales. Two sources are matched if their distance is less than $d_{max} = \max(a, 1.5\,\sigma_{\rm PSF})$. The final position error for a given detected source is estimated at the scale where the detection is most significant (see section 4.4).

We cross-identify only sources detected at consecutive scales, since for a real source we do not expect "dips" in the detection significance as a function of a: such dips, yielding a double-peaked profile of $y(a)$, would be indicative of either a spurious source (which will be likely a low-significance detection), or the close superposition of two sources of very different size (e.g. a point- and an extended source), which therefore need to be defined as distinct objects. The ability to do this is peculiar to the multi-scale approach of a wavelet-transform analysis.

It is possible that a group of point sources, that are either too weak or too close together to be detected individually, are instead detected as a single extended source. Although the WT approach may allow us in some cases to resolve close sources, not resolvable by other methods (Appendix D), this difficulty is more closely related to the instrument or detector properties, and cannot be solved by a detection method, however refined.

4.4. Evaluation of Source Position

To derive the source position more accurately, we next compute the WT in a small neighborhood (square of side $2a$) of each source at the scale a_{max} where $y(a)$ is a maximum, over a finer grid of spatial positions (with spacing of 1/2 image pixel for the smallest scales, and 1 pixel for the larger ones) and find the WT peak yielding both a refined position and WT amplitude for the source. In the limit of infinite statistical significance, the positional accuracy attained is of order of half the refined grid spacing. For finite photon statistics positional accuracy is poorer; a study of it on the basis of both real and simulated data is presented in Paper II, for the specific case of the ROSAT PSPC. At this stage, the significance threshold for a detection is set exactly to $n\sigma$, since we are now accurately sampling the peak of $w(x, y, a)$; a few sources still below that significance level are now rejected.

4.5. Source Count Rate and Size

In order to determine the count rate and size of a gaussian source, we make use of the function $y(a) = w(a)/a$ introduced in section 2, that for a gaussian source is equal to

$$y(a) = \frac{1}{a} \frac{2I_{src}}{\left(1 + \frac{\sigma_{src}^2}{a^2}\right)^2}, \tag{13}$$

having a maximum for $a_{max} = \sqrt{3}\sigma_{src}$. Here $y(a)$ is proportional to the source intensity (count rate) I_{src}, not to the source total counts N_{src}, since we apply the WT to the count-rate image; σ_{src} is the source width. To make use of equation (13) it is necessary to evaluate the WT at the detected source position for various values of a, and then find the values of I_{src} and σ_{src} which best approximate the observed profile of $y(a)$.

While this problem seems at first sight to be soluble with ordinary (non-linear) χ^2 fitting, this is not straightforward to apply. First, many sources are weak, and not far above the detection threshold, so that $y(a)$ evaluated at scales much different than $a = a_{max}$ will be dominated by noise, and will not be usable in the fitting. Second, implicit in equation (13) is the assumption of an underlying flat background, which may be violated because of sky background nonuniformities. Third, and most important, equation (13) does not take into account the presence of other sources in the image, whose wavelet transforms, described by equation (4), start to interfere with each other as soon as the scale a approaches the separation between sources. In principle, it is possible to arrange all interference terms between N sources, at a fixed scale a, into a $N \times N$ square matrix, and invert this latter to recover the "original" wavelet-transform amplitudes (and source rates I_{src}), provided that *all* sources have been detected and that their size and positions are accurately known. However, such a matrix inversion would strongly amplify errors in the input data, yielding nonsense results when the number of sources N becomes fairly large.

However, even for images containing only sparse point sources on a flat background, without source interference, it is not rigorously correct to apply a χ^2 fitting to $y(a)$ data. Indeed, measurements of $y(a)$ made at different scales a are not independent, the source photons being reused at each a. Therefore, the errors on $y(a)$ measured at various a will not be mutually independent, and this will alter the derived parameter confidence intervals, based on the χ^2 probability distribution (although the best-fit parameters may be correct in the absence of source interference and appreciable noise). Therefore, there are a number of reasons that suggest us to *avoid the approach of a χ^2 fitting* to the $y(a)$ profile.

Looking for an alternative way of deriving count rates and sizes for detected sources we note that, for a given source, the values of $y(a)$ least affected by both large errors and interference with nearby sources are those close to $a = a_{max}$. The first assertion should be quite obvious, since the scale a_{max} yielding a maximum of $y(a)$ is close to the one yielding a maximum significance for that detection, because the significance curves for a given confidence level $k\sigma$, defined in section 3, are almost flat using the $y(a)$ representation. The wavelet transform $w(a)$ for $a = a_{max}$ is also not significantly affected by interference: in order for two close sources of width σ_{src} to be detected as distinct, it is necessary that their overall photon distribution makes two distinct peaks, although their wings may overlap, and therefore their separation is $d > 2\sigma_{src}$; such peaks of the photon distribution will produce distinct peaks (detections) in the WT for scales $a \leq 2\sigma_{src} < d$, and in particular for $a_{max} = \sqrt{3}\sigma_{src}$ (equation 6), while they interfere and eventually merge in the WT at larger scales. The relative interference term between the WT of two identical sources at a scale a_{max}, with the minimum separation $d = 2\sigma_{src}$, can be evaluated from equation (4) (section 2), and amounts to $e^{-1/2}/2 \sim 0.3$, decreasing rapidly for increasing d. Therefore, *the maximum of the function $y(a)$ provides the least biased measure of source properties* (counts and extent) for marginally detected or marginally resolved sources. We then estimate to zeroth order the source size as $\sigma_{src} = a_{max}/\sqrt{3}$, and the count rate as $I_{src} = (8/9)\, w(a_{max})$.

If a source has been detected with a high significance level, and is relatively free of interference with nearby sources, a more refined derivation of its properties may be made, based on two values of $y(a)$ close to maximum. Suppose to have detected the source at two scales close to a_{max}, say a_1 and a_2, and measured $w(a_1)/a_1 = g_1$ and $w(a_2)/a_2 = g_2$, respectively. From equation (13) we have:

$$\frac{a_2}{a_1}\left(\frac{1+\frac{\sigma_{src}^2}{a_2^2}}{1+\frac{\sigma_{src}^2}{a_1^2}}\right)^2 = \frac{g_1}{g_2}, \tag{14}$$

which, setting $\alpha \equiv \dfrac{a_2}{a_1}$ and $x \equiv \dfrac{\sigma_{src}^2}{a_1^2}$ transforms into:

$$\frac{\alpha^2 + x}{1 + x} = \sqrt{\frac{g_1 \alpha^3}{g_2}} \equiv k_0, \tag{15}$$

where k_0 is the only term dependent on the data (and affected by error). Thus,

$$\sigma_{src} = a_1 \sqrt{\frac{k_0 - \alpha^2}{1 - k_0}}. \tag{16}$$

Once σ_{src} is known, the count rate I_{src} is directly obtained from equation (13).

Although this procedure is not the same as a fit (there is just one curve of the family described by equation (13) having values g_1, g_2 in a_1, a_2, respectively), it shares with a χ^2 fit the mentioned problem that g_1 and g_2 are not statistically independent measures. This affects the determination of errors on σ_{src} and I_{src}. In order to overcome this difficulty, we have developed an orthogonalization procedure to derive such errors in the correct way, described in Appendix C. If the relative errors on I_{src} and σ_{src} derived in this way are not $\ll 1$, this procedure is not correct, and we must use the zeroth-order approximation for I_{src} and σ_{src}; the errors on rate and size in this case are computed in Appendix C as well.

For some purposes it may be useful to estimate the number of source counts, N_{src}. This may be derived from I_{src} already computed, using a suitably weighted exposure time t_{phot} (which is not the same as t_{eff} defined in section 4.2 and Appendix B). To derive t_{phot}, consider a source with known count rate I_{src} and size σ_{src}; if the i-th pixel in the given source region has exposure t_i, and a fraction e_i of the source photons fall in it because of the finite source size, then the total number of recorded source counts is:

$$N_{src} = \sum_i I_{src}\, t_i\, e_i. \tag{17}$$

If w_i are the coefficients of the WT at the source position at a scale $a = a_{max}$, the source count rate is derived as:

$$I_{src} = \frac{8}{9} w(a_{max}) = \frac{8}{9} \sum_i I_{src}\, w_i\, e_i. \tag{18}$$

Therefore, $t_{phot} = N_{src}/I_{src}$ will be:

$$t_{phot} = \frac{9}{8} \frac{\sum_i t_i e_i}{\sum_i w_i e_i}, \tag{19}$$

which allows us to derive N_{src} once I_{src} is known. Where $t_i \sim const. = T$ over the source region, we have $t_{phot} \sim t_{eff} \sim T$.

4.6. Upper Limits

The large areas of the image that do not contain sources above threshold still contain valuable information. Upper limits to fluxes (e.g. in X-rays) at the positions of objects known from other wavelenghts can be important (e.g. "dividing line" for stellar coronal activity, Linsky and Haisch 1979; Ayres et al. 1981). To estimate these upper limits we proceed in an entirely analogous way to source detection. A given undetected source will have a size determined by the local PSF, $\sigma_{src} = \sigma_{\rm PSF}$ if it is pointlike, or $\sigma_{src} = \sqrt{\sigma^2_{\rm PSF} + \sigma^2_{proj}}$ if it has a finite projected size σ_{proj}. Assigning σ_{src} fixes the shape of the wavelet-transform amplitude $w(a)$ of the source, defined by equation (5), letting free its normalization I_{src}, namely the source count rate. For increasing I_{src}, the source WT profile $w(a)$ will shift upwards, until for some scale a it will be equal to the threshold WT amplitude curve $w_0(n, q)/t_{eff}$ for $n\sigma$ source detection, defined by equation (7) in section 3, scaled by t_{eff} defined in Appendix B since we are working on the WT of a count-rate image. The threshold curve at a significance level $n\sigma$ depends on the photon density q (section 3), but may be defined as a function of a once the background density b is known, since $q = b\, a^2$. A source is marginally detected (at $n\sigma$ significance level) when its $w(a)$ profile becomes tangent to the threshold significance curve at $n\sigma$ (see Figure 8):

$$w(a) = \frac{2 I_{src}}{\left(1 + \frac{\sigma^2_{src}}{a^2}\right)^2} \leq \frac{w_0(n, b\, a^2)}{t_{eff}}. \tag{20}$$

This condition applies for a scale $a \sim a_{max} = \sqrt{3}\sigma_{src}$ ($a = a_{max}$ exactly in the continuous limit), which inserted in equation (20) allows us to derive the $n\sigma$ limiting count rate I_{src}. The background density b needed to define the proper significance curve should be defined on a scale close to a_{max}, as for source detection. If we want to compute the upper limit close to a detected source (within a few times $\sigma_{\rm PSF}$) a correction must be applied, that is derived in Appendix D.

The upper limit computed in this way for point sources in every position in the field of view makes a self-consistent limiting sensitivity map for point-source detection in a given observation. Depending on the particular application, such a sensitivity map may even be constructed *without* correcting for the presence of sources; for example, this would allow us to know how farther the detected sources themselves could have been located, while remaining detectable with our method.

4.7. Ellipticity of Extended Sources

If a source is found as extended, we may relax the assumption of isotropy that we made from the beginning, and study if the source has an elliptical shape, through a suitable change on the generating wavelet. Of course, such an analysis is mainly worthwhile if the source was detected with sufficiently high significance[1]. In this subsection we assume that our source has an elliptical shape, with a gaussian profile along each axis:

$$s(x,y) = \frac{I_{src}}{2\pi\sigma_1\sigma_2}\, e^{-\frac{(x\cos\theta_s + y\sin\theta_s)^2}{2\sigma_1^2} - \frac{(-x\sin\theta_s + y\cos\theta_s)^2}{2\sigma_2^2}}, \tag{21}$$

θ_s being the source position angle, and $\sigma_{1,2}$ its widths along two perpendicular directions. I_{src} is the total source count rate.

We can accordingly generalize the generating wavelet of equation (2) to a form $g(x, y, a_1, a_2, \theta_w)$, that includes a dependence on two scale parameters a_1, a_2 and one wavelet position angle θ_w. The convolution of this function with the elliptical source yields a function $w_{peak}(a_1, a_2, \sigma_1, \sigma_2, \theta_s - \theta_w, I_{src})$. The study of this latter function for variable a_1, a_2 and θ_w allows us to derive the source geometrical parameters σ_1, σ_2 and θ_s, and its intensity I_{src}. Details of these computations are described in Appendix E.

5. Summary and Conclusions

In this work, we have developed the general mathematical aspects of a wavelet-based source detection algorithm, to be applied to images obtained with current photon-counting detectors.

In the analysis of such images (e.g. X-ray images), we are faced with special difficulties, only rarely encountered for example in optical images: the image may be significantly vignetted, partially obscured by sharp ribs, there may be quantum efficiency variations, and the PSF may vary strongly with offaxis angle. While this latter property is easily handled using wavelets, the former three are not, and we had to devise a special technique, resorting to the observation exposure map to compute a count-rate image (flat-fielding), and then reconstructing the statistical properties of the transform *as if* it were done on the original photon image. We have paid particular attention to statistical issues, in order to derive accurate thresholds to discriminate background fluctuations from true sources at an assigned confidence level. To do so, the probability distribution of WT values obtained transforming a flat background noise has been studied in detail.

After a multiscale process of source detection and verification, properties such as source count rates and sizes (and possibly ellipticity) are derived, as accurately as the available photon statistics permits. Errors on these quantities are computed in a statistically accurate way as well. We evaluate upper limits on a list of positions, based on essentially the same principles which allow us to estimate count rates for detections; this enables us to build a self-consistent sensitivity map over the entire field of view. The whole detection process occurs in two steps, that differ in the way

[1]It is in principle possible that a weak elongated source be significant only when an anisotropic wavelet is used. We prefer anyhow to do the initial source search using an isotropic wavelet, since the use of anisotropic wavelets would enormously enlarge the parameter space to be searched (two wavelet scales plus position angle), and therefore the computational load, with only a very modest expected increase in the number of detections.

the background is computed: in the first one, a smoothed (local) background map is computed irrespective of the presence of sources, while in the second one a similar map is interpolated over detected source regions, thus more resembling a "map" method.

Of course, the application of the method to images obtained with a given detector requires a fine tuning to take into account the detailed properties of the detector. Among the most important differences between various detectors, from the point of view of implementation of our algorithm, are the shape and sharpness of image edge, the presence of underexposed regions (ribs or inter-chip gaps), and especially the average background density μ. The former two problems may be dealt with relatively easily through the use of the exposure map. A problem may instead arise if the background density *per resolution element* is very low, namely if $\mu\sigma_{\rm PSF}^2 \leq 1$, since in such a case it becomes difficult to evaluate accurately the local background, especially where the source density is high. Such a problem may occur with short exposures (as it was also the case for the *Einstein* Slew Survey; Elvis et al. 1992), and with high-resolution detectors, and will be extremely relevant for the AXAF HRC/ACIS detectors, having an extremely narrow PSF. Obviously, a compromise must be achieved between statistical significance and "locality" of the used background representation.

Notwithstanding such a difficulty of building an accurate background map, that is also shared with many other source detection schemes for photon-counting-detector images, with the wavelet approach we are able to deal with images consisting of sparse arrays of one-photon pixels, since we have calibrated WT detection thresholds down to low values of (scaled) background q, in the "discrete photon" limit.

Besides being appropriate to analyze low-exposure or high-resolution images, such a kind of situation also occurs if in place of single photons one has other objects, for example stars or galaxies, and one is interested to find local *clusters* of such objects ("sources"), in projection onto the sky plane (or any other plane). Wavelet transforms have already been applied to this problem by Slezak, Bijaoui and Mars (1990), Escalera and Mazure (1992) and Escalera, Slezak and Mazure (1992), but only using an approximate statistical treatment of the WT. To approach this problem with our method, only little modifications are required to the procedure described in this paper; among the most significant are the following: first, the position of such individual objects is usually known with high accuracy (with respect to their average apparent separation), so that the image "PSF" may be taken as zero, unlike the case dealt with here, where the PSF size sets the smallest length scales on which information is present in the data. This implies that the minimum and maximum scales for WT analysis may be difficult to define. Moreover, as is typical of the discrete photon limit, the local background level (density of objects outside clusters) may be difficult to determine, since this necessarily requires some average over large regions, possibly including clusters themselves.

Because of the form of the generating wavelet $g(r/a)$ chosen here, the WT is more sensitive to positive fluctuations (peaks) in the data than to negative ones (holes), which in the context of object cluster detection may occur e.g. due to local enhanced obscuration (dark clouds). To detect these "holes" in the data, some other generating wavelet $g(r/a)$ may be more appropriate, as for example simply the *negative* of $g(r/a)$, that we define as $g^{-}(r/a) \equiv -g(r/a)$, having a strong negative *minimum* for $r = 0$. The corresponding WT is of course $w^{-}(x, y, a) = -w(x, y, a)$, with thresholds $w_0^{-} = -w_0(k, q)$ (eq. [7]), that amounts to compute the WT exactly as before, find its *negative* peaks w_{min}, and select a source (i.e. hole) as significant if $-w_{min} > w_0(k, q)$.

A further generalization of WT analysis is its application to three-dimensional data sets, made e.g. of projected sky positions (x, y) plus a distance z (spectroscopic parallax, or redshift) for a list of objects. The "mexican hat" generating wavelet may then be generalized to three dimensions, as

$$g\left(\frac{r}{a}\right) = \left(3 - \frac{r^2}{a^2}\right)\, e^{-\frac{r^2}{2a^2}}, \qquad (r^2 = x^2 + y^2 + z^2) \tag{22}$$

in order to ensure its volume mean to be zero. It is then necessary to recompute all probability distributions $P_q(w)$, and use a proper three-dimensional background map. All other steps remain essentially the same, except for the increased computational load due to the extra dimension. The three-dimensional WT analysis would of course allow the detection in the real space of star- or galaxy clusters, which may not be evident when projected on the plane of the sky, because of the increased background of objects.

Finally, a generalization is possible to more dimensions, including e.g. the energy and arrival time data for detected photons in a X-ray data set, in addition to position. Such generalizations are however beyond the scope of this paper.

We acknowledge support from the italian ASI and MURST. We thank F.Reale, F.R.Harnden, G.Mackie, R.Rosner, V.Kashyap, T.Maccacaro, R.Della Ceca, and A.Wolter for fruitful discussions and suggestions. Special thanks are due to M.Elvis, G.Fabbiano, and S.Serio for their careful reading of the manuscript, that has led to many improvements in the content and presentation of this paper. We also thank the referee, A.Vikhlinin, for his useful comments and suggestions.

Appendices

A. Derivation of the Probability Distribution of the Wavelet Transform of Background Noise

In this Appendix we study the probability distribution of WT values resulting upon transforming a background noise image, with the purpose of deriving thresholds for source detection in the wavelet transform space. We examine both cases of very low and high background, whose statistics is markedly different.

One background photon: For one background photon in a circle of radius $5a$ around a fixed point (x_0, y_0) the probability that this photon falls in a surface element dA around a point (x_1, y_1) is:

$$P(x_1, y_1) dA = \frac{1}{25\pi a^2} dA. \tag{A1}$$

Then, the image to be transformed is $f(x, y) = \delta(x - x_1, y - y_1)$ and its WT in (x_0, y_0) is $w = g(r/a)$, where $r^2 = (x_1 - x_0)^2 + (y_1 - y_0)^2$. Defining $z \equiv \frac{r^2}{a^2}$, we have a WT of the form

$$w(z) = (2 - z)e^{-z/2}, \tag{A2}$$

and a probability distribution

$$P(z)\, dz = \frac{1}{25}\, dz. \tag{A3}$$

Now, the probability that the value of the WT falls in an interval $w_0 \pm dw_0$ is the sum of probabilities that z falls in any interval $z_i \pm dz_i$ such that $w(z_i) = w_0$ (inverse images of the interval $w_0 \pm dw_0$ on the z-axis); namely:

$$P(w_0)\, dw_0 = \sum_i P(z_i)\, dz_i \tag{A4}$$

(see Figure 9). For a fixed $w_0 \geq 0$, the equation $w(z) = w_0$ has only one solution $z_1 < 2$ (and the above sum contains only one term), while for $w_0 < 0$ it has two solutions $z_1 < 4$ and $z_2 > 4$ (two terms in the sum). Therefore the probability $P(w)$ is

$$P(w) = P(z_1) \left|\frac{dz}{dw}\right|_{z_1} = \frac{1}{25\,(2 - z_1/2)e^{-z_1/2}} \qquad (w \geq 0) \tag{A5}$$

$$P(w) = P(z_1) \left|\frac{dz}{dw}\right|_{z_1} + P(z_2) \left|\frac{dz}{dw}\right|_{z_2} = \frac{1}{25\,(2 - z_1/2)e^{-z_1/2}} + \frac{1}{25\,(z_2/2 - 2)e^{-z_2/2}} \qquad (w < 0).$$

Since the function $w(z)$ cannot be inverted analytically, we cannot write $P(w)dw$ explicitly as a function of w; rather, it has been computed numerically, and is shown in Figure 10. The distribution is highly asymmetrical (although it has zero mean), with tails up to $w = 2$ (for $z = 0$) and down to $w = -2e^{-2}$ (for $z = 4$); photons falling at larger z, for which $w(z)$ tends to zero from the negative side, build the strong peak of $P(w)$ just below $w = 0$; the non-monotonicity of $w(z)$ makes $P(w)$ discontinuous in $w = 0$. We shall refer to this distribution as $P_1(w)$, since it refers to one background photon.

Two or more background photons: If there are two background photons in the same area $A = 25\pi a^2$, the total WT will be $w = w(z_{\rm I}) + w(z_{\rm II})$, since the WT is a linear transform, where $z_{\rm I}$ and $z_{\rm II}$ refer to the first and second photon, respectively, and $w(z_{\rm I})$ and $w(z_{\rm II})$ are both distributed as $P_1(w)$. Under these circumstances, the total w will have a probability distribution $P_2(w) = P_1(w) \otimes P_1(w)$, namely the convolution of two one-photon distributions. Of course, this is generalized for the probability distribution $P_n(w)$ for n background photons, as $P_n = P_1 \otimes P_1 \otimes \cdots \otimes P_1$ (P_1 convolved n times with itself). As is known, convolution is best done in Fourier space, where it involves a simple product of functions. We need, therefore, to compute the Fourier transform of $P_1(w)$, say $\tilde{P}(p)$, raise it to the n-th power and transform it back to get $P_n(w)$.

The Fourier transform of $P_1(w)$ is

$$\tilde{P}(p) = \int_{-\infty}^{+\infty} e^{-ipw}\, P_1(w)\, dw = \int_0^{25} e^{-ip(2-z)e^{-z/2}}\, P(z)\, dz = \frac{1}{25} \int_0^{25} e^{-ip(2-z)e^{-z/2}}\, dz. \tag{A6}$$

This integral can be hardly evaluated analytically, so we have computed all Fourier transforms numerically, using FFT routines.

In general, we are interested to know in detail the positive tails of the distributions $P_n(w)$, say at probability levels equivalent to the gaussian 3σ to 5σ, to discriminate well true sources from background fluctuations, and obtain a very small number of false detections even after having tested a large number of positions ("trials"). For example, the threshold value $w = w_0$ corresponding to a probability level of 0.9999998 (5σ for a gaussian distribution) that $w \geq w_0$ is not due to a background fluctuation is given implicitly by the relation:

$$1 - \int_{-\infty}^{w_0} P_n(w)\, dw = 2 \cdot 10^{-7}. \tag{A7}$$

Such determination of w_0 for a given probability level is near or even beyond what can be computed with the finite numerical precision of the "standard" numerical approach using FFT's. This latter therefore should not be used to derive confidence levels for w, but simply to illustrate the shape of $P_n(w)$ for various n, and to show the existence of the two limiting behaviours of "discrete photon limit" for small n, and "continuous limit" for large n (section 3). As explained in section 3, it is also useful to define the dimensionless background parameter $q = ba^2$, on which the distributions $P_n(w)$ depend. In our case, since $b = \frac{n}{25\pi a^2}$, we have $q = \frac{n}{25\pi}$.

In the continuous limit, when the photon statistics per pixel becomes gaussian, $P_n(w)$ tends to a gaussian distribution, as it may also be derived from statistical arguments (central limit theorem). The integral

$$w = \int \left(2 - \frac{r^2}{a^2}\right) e^{-\frac{r^2}{2a^2}} f(r,\theta)\, d^2r, \tag{A8}$$

defines the WT of the background noise $f(r,\theta)$ (= number of photons per unit area, whose spatial average is a constant background density μ); it may be approximated as a discrete sum of terms (a finite number since the integrand is effectively non-zero only in a finite region): $w = \sum_i w_i n_i$, where $w_i = \left(2 - \frac{r_i^2}{a^2}\right) e^{-\frac{r_i^2}{2a^2}}$, and $n_i = f(r_i, \theta_i)\, A_i$ is the number of photons falling in the area A_i around (r_i, θ_i); we suppose that the r_i's are spaced closely enough that differences between w_i and w_{i+1} are so small that the summation approximates well the integral for w. If the n_i's have all a gaussian distribution with mean $< n_i >$ and variance σ_i^2 both equal to μA_i (the limiting form of a Poisson distribution for large n), their linear combination w just defined will have a gaussian distribution as well, with mean:

$$< w >= \sum_i w_i < n_i >= \mu \sum_i w_i A_i \rightarrow \mu \int \left(2 - \frac{r^2}{a^2}\right) e^{-\frac{r^2}{2a^2}} d^2r = 0 \tag{A9}$$

and variance:

$$\sigma_w^2 = \sum_i w_i^2 \sigma_i^2 = \mu \sum_i w_i^2 A_i \rightarrow \mu \int \left(2 - \frac{r^2}{a^2}\right)^2 e^{-\frac{r^2}{a^2}} d^2r = 2\pi\mu a^2 = 2\pi q \tag{A10}$$

on the basis of the convolution theorem. The variances σ_w^2 computed in this way are in excellent agreement with those of the distributions $P_n(w)$ (in the continuous limit) derived using FFT's. FFT's are a good test in this case since we are comparing variances and not higher quantiles.

In the discrete photon limit, where no analytic approximations are available, and FFT's are too noisy to reach the desired confidence levels for w, we have evaluated probability distributions $P_n(w)$ through MonteCarlo simulations, made for various values of q. As mentioned above, the value w_0 corresponding to a 5σ probability level (cumulative $P = 0.9999998$) is such that $w \geq w_0$ occurs 2 times in 10^7 trials. For each q, therefore, we have realized 10^7 simulations of a uniform background, computed their WT and built the distribution of w; we have then evaluated the upper quantiles of the distributions $P_q(w)$ computed in this way, up to the 5.5σ quantile, with increasing accuracy towards lower quantiles due to the better simulation statistics. The results of these simulations are shown in Figure 5 (section 3), as small dots. We have also verified that for $q \geq 300$, $P_q(w)$ becomes nearly indistinguishable from a gaussian with the expected variance $\sigma_w^2 = 2\pi q$ (continuous limit), in close agreement with the computations previously discussed.

B. Calculation of Effective Exposure Time

The effective exposure time t_{eff} defined in section 4.2 has the purpose of permitting a comparison between the positive fluctuations found in the rate WT with those expected from the neighboring background, to establish if they are real sources or not. Therefore, its meaning is that of a tool to convert detection thresholds between the case of a (ribs-free) photon WT and that of a rate WT under examination. Note that t_{eff} defined in this way *is not a source exposure time* to be used to convert a source rate to source total counts; this latter is introduced in section 4.5.

We have introduced the threshold $w_0(k,q)$ for the photon WT in section 3, and call now $w_{0r}(k,q)$ the corresponding threshold for the count-rate WT, such as $w_{0r}(k,q) \equiv w_0(k,q)/t_{eff}$. The parameter q, defined in section 3, is the same for $w_0(k,q)$ and $w_{0r}(k,q)$, since in both cases it is computed from the background map. An analytic expression for t_{eff} may be found in the "continuous limit" introduced in section 3; roughly speaking, this limit is applicable when there are many photons per pixel (which unfortunately is usually not the case below ribs). However, this way to compute t_{eff} is probably the best one before resorting to empirical methods. In the continuous limit we have simply $w_0(k,q) = k\sqrt{2\pi q}$.

Suppose to have a uniform cosmic background emission, yielding a count rate of R photons sec^{-1} pix^{-1} just before reaching the detector with its spatially-varying effective exposure time. After interaction with the detector, the number of photons in the i-th pixel (having effective exposure t_i) will be n_i, Poisson-distributed with mean μ_i and variance σ_i^2, $\mu_i = \sigma_i^2 = R\,t_i$. The WT in a point P of such a background count-rate image, modulated by spatially nonuniform exposure, is computed as (cf. eq. [1]):

$$w\left(\frac{n_i}{t_i}\right) = \sum_i w_i \frac{n_i}{t_i}, \tag{B1}$$

and its variance, computed as a quadrature sum, is:

$$\sigma_w^2\left(\frac{n_i}{t_i}\right) = \sum_i \frac{w_i^2}{t_i^2}\sigma_i^2 = \sum_i \frac{w_i^2}{t_i^2} R\,t_i = R\sum_i \frac{w_i^2}{t_i} \tag{B2}$$

assuming the t_i's are free of error. In these and the following equations, the writing $w(x_i)$ (or $\sigma^2(x_i)$) refers to a quantity w (σ^2) which is computed from a set of pixel values x_i. The coefficients w_i are defined as $w_i = g(r_{iP}/a)$, where g is the generating wavelet (eq. [2]), r_{iP} is the distance between the i-th pixel and the point P, and a is the WT scale; moreover all sums are to be intended as two-dimensional, running over all image pixels. Suppose now that the same background is observed with a similar detector, but free of spatial variation of exposure in the source region, so that $t_i = const. = T$. The variance of the WT of such a new background, yielding n_i' photons/pixel, is therefore:

$$\sigma_w^2\left(\frac{n_i'}{T}\right) = \frac{R}{T}\sum_i w_i^2, \tag{B3}$$

and the ratio between the two variances is:

$$\frac{\sigma_w^2\left(\frac{n_i}{t_i}\right)}{\sigma_w^2\left(\frac{n_i'}{T}\right)} = \frac{T\sum_i \frac{w_i^2}{t_i}}{\sum_i w_i^2}. \tag{B4}$$

Furthermore, $\sigma_w^2\left(\frac{n_i'}{T}\right) = \frac{1}{T^2}\sigma_w^2(n_i') = \frac{1}{T^2}2\pi q$. Since in the continuous limit the ratio w_0/w_{0r} $(= t_{eff})$ between detection thresholds for photon and rate WT, at the same confidence level $k\sigma$, is equal to the ratio between the standard deviations of photon and rate WT, respectively, we can write

$$\sigma_w^2\left(\frac{n_i}{t_i}\right) = \frac{1}{t_{eff}^2}\sigma_w^2(n_i') = \frac{2\pi q}{t_{eff}^2}, \tag{B5}$$

and therefore

$$t_{eff} = \sqrt{\frac{T\sum_i w_i^2}{\sum_i \frac{w_i^2}{t_i}}}, \tag{B6}$$

which solves our problem. In order to be applied in our case, T is the slowly-varying vignetted exposure time we would have in the absence of ribs, in the same position, and the reference background variance $\sigma_w^2(n_i') = 2\pi q$ is accordingly computed for a vignetted background, but unobscured by ribs. Such a background has been computed, as explained in section 4.1, by interpolation over ribs, even at positions actually covered by ribs, and needs not to be recomputed. Although derived in the continuous limit, this expression for t_{eff} works generally fine to define proper detection thresholds; we present in Paper II tests showing that there is no appreciable excess of spurious sources near ribs due to large count-rate fluctuations, and that at the same time a good detection efficiency is retained. If all t_i are nearly equal, $t_i \sim T$, then $t_{eff} \sim T$ as well, and equation (B6) may therefore be used far from ribs as well.

C. Derivation of Errors on Source Count Rate and Size

In this Appendix we present a statistically accurate derivation of errors on count rates and sizes, needed since these latter were evaluated in section 4.5 on the basis of non-independent measurements. We use here the same terminology used in that section. There, the size and count rate for a source were derived from two values g_1, g_2 of the function $y(a)$ at the source position, computed respectively at two scales a_1, a_2 close to a_{max}.

Since our derivation of σ_{src} depends only on the ratio $r \equiv g_1/g_2$, we should evaluate the error on such a ratio of non-independent measurements. To do this, we first note that g_1 and g_2 are computed as linear combinations of the count-rate image pixel values x_i, namely

$$g_1 = \sum_i g_1^i \, x_i \tag{C1}$$

(the sum is two-dimensional, over image pixels; the coefficients are defined in terms of the generating wavelet $g(r/a)$ (eq. [2]) as $g_1^i = g(r_i/a_1)/a_1$, where r_i is the distance between the i-th pixel and the source position), and analogously for g_2. We may then write g_1 as the scalar product $g_1 = \vec{g}_1 \cdot \vec{x}$, where $\vec{g}_1$ is the vector made of all g_1^i coefficients, and $\vec{x}$ is the vector of pixel data x_i. The lack of statistical independence of g_1 and g_2 then arises because (in general) they are not *orthogonal* linear combinations of the data x_i, namely $\vec{g}_1 \cdot \vec{g}_2 \neq 0$. In order to have two independent measurements, we should orthogonalize in a proper way the vectors $\vec{g}_1$ and $\vec{g}_2$. Before doing that, however, we have to normalize the data vector $\vec{x}$ by dividing each pixel x_i by its standard deviation σ_i, yielding a modified data vector $\vec{x}'$ whose elements are $x_i' = x_i/\sigma_i$. Failure to do so would never yield orthogonal $\vec{g}_1$ and $\vec{g}_2$: suppose to have a prominent pixel $x_k \gg x_{i\neq k}$ (and

$\sigma_k \gg \sigma_{i\neq k}$), which implies $g_1 \sim g_1^k x_k$ and $g_2 \sim g_2^k x_k$, causing g_1 and g_2 to be strongly correlated, no matter how they have been orthogonalized. Having changed $\vec{x}$ into $\vec{x}'$, we modify accordingly $\vec{g}_{1,2}$ into vectors $\vec{g}'_{1,2}$ with elements $g'^{i}_{1,2} = g^{i}_{1,2}\,\sigma_i$, in order that the scalar products $g_{1,2}$ remain unaltered: $g_{1,2} = \vec{g}_{1,2}\cdot\vec{x} = \vec{g}'_{1,2}\cdot\vec{x}'$. We then write r as:

$$r = \frac{g_1}{g_2} = \frac{\vec{g}'_1\cdot\vec{x}'}{\vec{g}'_2\cdot\vec{x}'}. \tag{C2}$$

To separate $\vec{g}'_1$ into two components, respectively normal and parallel to $\vec{g}'_2$, we apply a Gram-Schmidt orthogonalization procedure:

$$\vec{g}'_1 = \left[\vec{g}'_1 - \vec{g}'_2\left(\frac{\vec{g}'_1\cdot\vec{g}'_2}{\vec{g}'_2\cdot\vec{g}'_2}\right)\right] \;+\; \vec{g}'_2\left(\frac{\vec{g}'_1\cdot\vec{g}'_2}{\vec{g}'_2\cdot\vec{g}'_2}\right) \equiv \vec{g}'_3 \;+\; \vec{g}'_4, \tag{C3}$$

where $\vec{g}'_3$ ($\perp \vec{g}'_2$) is the term within square brackets in the middle expression, and $\vec{g}'_4$ ($\parallel \vec{g}'_2$) is the last term. Then we write:

$$r = \frac{\vec{g}'_3\cdot\vec{x}'}{\vec{g}'_2\cdot\vec{x}'} \;+\; \frac{\vec{g}'_4\cdot\vec{x}'}{\vec{g}'_2\cdot\vec{x}'} = \frac{g_3}{g_2} \;+\; \left(\frac{\vec{g}'_1\cdot\vec{g}'_2}{\vec{g}'_2\cdot\vec{g}'_2}\right) \;=\; \frac{g_3}{g_2} + k_1. \tag{C4}$$

The term k_1 (last term within brackets) depends only on the coefficients $g'^{i}_{1,2}$, known with no error if σ_i are so (as we will assume); therefore, it is a constant additive term to r, and does not contribute to the error on r. This term contains the "parallel" part of g_1, which varies in proportion to g_2. The error on r is thus entirely given by the error on the ratio between g_3 and g_2, which now *are* statistically independent (since $\vec{g}'_3 \perp \vec{g}'_2$), allowing to compute the error σ_r on r using classical methods: $\sigma_r^2/r^2 = \sigma_3^2/g_3^2 + \sigma_2^2/g_2^2$, where σ_2 (σ_3) is the error on g_2 (g_3), $\sigma_2^2 = \sum_i (g'^{i}_2)^2\,(\sigma'_i)^2 = \sum_i (g'^{i}_2)^2$, since the error on x'_i is $\sigma'_i = \sigma_i/\sigma_i = 1$ (and an analogous expression for σ_3^2). Having found the error on r, and therefore also the error σ_{k_0} on $k_0 = \sqrt{r\alpha^3}$ (cf. eq. [15]), we derive easily the error σ_x on $x = \sigma^2/a_1^2$ from equation (16), since we have $\sigma_x = \dfrac{|1-\alpha^2|}{(k_0-1)^2}\sigma_{k_0}$, from which the error $\Delta\sigma_{src}$ on source size σ_{src} is derived.

As for the error on the source count rate, we may write a suitable expression for I_{src}:

$$I_{src} = \frac{g_1 a_1}{2}\left(1 + \frac{\sigma_{src}^2}{a_1^2}\right)^2 = \frac{a_1(\alpha^2-1)^2}{2}\,\frac{1}{\left(\frac{\sqrt{\alpha^3}}{\sqrt{g_2}} - \frac{1}{\sqrt{g_1}}\right)^2}. \tag{C5}$$

Defining c as:

$$c = \sqrt{\frac{a_1(\alpha^2-1)^2}{2I_{src}}} = \frac{\sqrt{\alpha^3}}{\sqrt{g_2}} - \frac{1}{\sqrt{g_1}}, \tag{C6}$$

we have, with the same formalism as before:

$$dc = -\frac{1}{2}\frac{\sqrt{\alpha^3}}{\sqrt{g_2^3}}\,dg_2 + \frac{1}{2}\frac{1}{\sqrt{g_1^3}}\,dg_1 = \left(-\frac{1}{2}\frac{\sqrt{\alpha^3}}{\sqrt{g_2^3}} + \frac{k_1}{2}\frac{1}{\sqrt{g_1^3}}\right)\,dg_2 + \frac{1}{2}\frac{1}{\sqrt{g_1^3}}\,dg_3. \tag{C7}$$

Here, we have again decomposed the variation dg_1 of g_1 into two parts, $dg_1 = dg_3 + dg_4$, respectively orthogonal and parallel to dg_2 (namely, unrelated and proportional to dg_2, respectively). Note that dg_k (k=1-4) is the variation of g_k due to variations of the data x_i, not of the fixed coefficients g_k^i. The "parallel" part dg_4 is:

$$dg_4 = dg_2\,\left(\frac{\vec{g}'_1\cdot\vec{g}'_2}{\vec{g}'_2\cdot\vec{g}'_2}\right) \;=\; k_1 dg_2, \tag{C8}$$

allowing to write the last equality of equation (C7). Then,

$$\sigma_c^2 = \left(\frac{k_1}{2}\frac{1}{\sqrt{g_1^3}} - \frac{1}{2}\frac{\sqrt{\alpha^3}}{\sqrt{g_2^3}}\right)^2 \sigma_2^2 + \frac{1}{4g_1^3}\,\sigma_3^2, \tag{C9}$$

from which we derive easily the error on I_{src}, ΔI_{src}, using equation (C6).

If the errors $\Delta\sigma_{src}$ and ΔI_{src} thus derived are not small compared with σ_{src} and I_{src}, the assumptions of the procedure are violated, and we then use the zeroth-order approximation to evaluate σ_{src} and I_{src}. In this case, errors are computed as follows: since the spacing between adjacent scales where we evaluate $y(a)$ is a factor $\sqrt{2}$, the error in our derivation of a_{max} (and then σ_{src}) should be at most a factor $2^{0.25}$, namely $\Delta\sigma_{src}/\sigma_{src} = 2^{0.25} - 1 \sim 0.19$. To derive the error on I_{src}, we start from the expression given in equation (C5), to derive

$$\frac{dI_{src}}{I_{src}} = \frac{dg_1}{g_1} + \frac{4\sigma_{src}/a_1^2}{(1+\sigma_{src}^2/a_1^2)}\,d\sigma_{src} = \frac{dg_1}{g_1} + \frac{d\sigma_{src}}{\sigma_{src}}, \tag{C10}$$

the last passage being made since $a_1 = a_{max} = \sqrt{3}\sigma_{src}$, and therefore $(\Delta I_{src}/I_{src})^2 = \sigma_1^2/g_1^2 + (\Delta\sigma_{src}/\sigma_{src})^2$. Errors on rates and sizes evaluated directly from simulated data (Paper II) are in agreement with those estimated here.

D. Correction to Upper Limits near Detected Sources

If we have an estimate of the background at the position of a point source, which is virtually unaffected by the presence of that source (as for the background correction of section 4.1), the upper limit computed as in section 4.6 will be smaller than the detected source count rate. This is obviously incorrect, since having *actually* detected the source, we cannot possibly find a weaker one in the same position, nor very close to it (compared to $\sigma_{\rm PSF}$). We should therefore find a correct way to compute upper limits near detected sources. The detection of more sources close to an already detected source is essentially limited by the detector spatial resolution, rather than its sensitivity; to evaluate true upper limits near detected sources requires consideration of our ability to separately detect close sources.

Two sources of equal strength: Suppose there are two identical sources (pointlike, i.e. gaussian with width $\sigma_{src} = \sigma_{\rm PSF}$), separated by a distance d. It is simple to show that their overall spatial distribution of photons (assuming infinite photon statistics) makes two distinct peaks if $d > 2\sigma_{\rm PSF}$, while the two sources are blended together for smaller separation d. Our method is however based on the search for peaks in the WT of the image, not in the image itself, so that we need to compute the smallest separation d for which the two sources give rise to two distinct peaks in the WT. The WT of a point source with width σ_{src}, computed at a scale a is given by equation (4). The WT of both sources, computed along the direction of their separation, is proportional to:

$$I(x) = \left(2 - \frac{x^2}{b^2}\right)\, e^{-\frac{x^2}{2b^2}} + \alpha\left(2 - \frac{(x-d)^2}{b^2}\right)\, e^{-\frac{(x-d)^2}{2b^2}}, \tag{D1}$$

where $b^2 = a^2 + \sigma_{src}^2$, and $\alpha \le 1$ is the source intensity ratio; the stronger source lies in $x = 0$ and the fainter one in $x = d$ (Figure 11a, b, c). For two sources of equal intensity ($\alpha = 1$), a minimum of $I(x)$ occurs for $x = d/2$ by symmetry, if the sources make distinct peaks. If this is the case, we have $\left.\frac{d^2I}{dx^2}\right|_{x=d/2} > 0$; two sources are marginally resolved if

$$\left.\frac{d^2I}{dx^2}\right|_{x=d/2} = 0. \tag{D2}$$

This implies, after some algebra:

$$\frac{d^4}{b^4} - 28\frac{d^2}{b^2} + 64 = 0, \tag{D3}$$

and $d^2/b^2 = 14 - \sqrt{132} \sim 2.5$. Therefore, $d \geq \sqrt{2.5}\sqrt{a^2 + \sigma_{src}^2}$. If the scale $a \to 0$, we have the surprising result that the minimum distance d_{min} for resolving two sources is $d_{min} = \sqrt{2.5}\sigma_{src} \sim 1.6\sigma_{src} < 2\sigma_{src}$, namely *the wavelet transform may have two distinct peaks where the original image shows two blended sources, acting as a sort of deconvolution.* This is so because the WT in a point "sees" the photon distribution in the neighborhood of that point, not just in the point itself. However, this attractive property of the WT is of little use, except for the brightest sources, since the amplitude of $w(a)$ tends to zero as a^4 when $a \to 0$ (section 2); weak sources, indeed, are detected only at scales a close to $a_{max} = \sqrt{3}\sigma_{src}$, and in this case we would have $d_{min} = \sqrt{2.5} \cdot 2\sigma_{src} > 2\sigma_{src}$, larger than the minimum peak separation in the photon image.

Two sources of different strength: If the two sources have unequal intensities ($\alpha < 1$), d_{min} will be a function of the intensity ratio α which can be inverted to yield $\alpha = \alpha(d_{min})$. Then if I_{src} is the count rate of the stronger source in $x = 0$, the upper limit on the count rate of another source in $x = d$ is $I_{lim} = \max(I_{src}\alpha(d_{min}), I^0_{lim})$, where I^0_{lim} is the limiting count rate computed in the absence of the stronger source, simply because we would be unable (with the given detector and detection algorithm) to separate the two sources for a lower I_{lim} (although the sensitivity itself of the detector may allow the detection of the latter source down to a count rate I^0_{lim}, if the former source were absent).

To compute $\alpha(d_{min})$, we need to find out the parameter values α, d_{min} for which the two WT peaks (maxima) associated with the two sources start to merge together; in this case, the position of the minimum of $I(x)$ between the two sources and that of the maximum of $I(x)$ associated with the weaker source coincide (Figure 11d, e, f). In formulas:

$$\begin{cases} \frac{dI}{dx} = 0 = \left(-\frac{4x}{b^2} + \frac{x^3}{b^4}\right) e^{-\frac{x^2}{2b^2}} + \alpha\left(-\frac{4(x-d)}{b^2} + \frac{(x-d)^3}{b^4}\right) e^{-\frac{(x-d)^2}{2b^2}} \\ \frac{d^2I}{dx^2} = 0 = \left(-\frac{4}{b^2} + \frac{7x^2}{b^4} - \frac{x^4}{b^6}\right) e^{-\frac{x^2}{2b^2}} + \alpha\left(-\frac{4}{b^2} + \frac{7(x-d)^2}{b^4} - \frac{(x-d)^4}{b^6}\right) e^{-\frac{(x-d)^2}{2b^2}}. \end{cases} \tag{D4}$$

Combining these two relations, one obtains a relation between the position x of the minimum and the minimum separation d_{min}, namely (after long passages, and setting $y \equiv x/b$, $z \equiv d/b$):

$$zy^6 - 3z^2y^5 + (3z^3 - 5z)y^4 + (10z^2 - z^4)y^3 + (16z - 9z^3)y^2 + (4z^4 - 16z^2)y + (16z - 4z^3) = 0. \tag{D5}$$

This sixth-degree equation in y has among its solutions $y = y(z)$ one that for $\alpha \to 1$ tends to the previously derived value, namely approaching $y = d/2b$ for $z = d/b/\sqrt{2.5}$ (minimum separation for $\alpha = 1$) and disappearing for lower z, since for $d < \sqrt{2.5}b$ no source may be separated, whatever the value of α. Writing this solution as $y_0(z)$, we have $x = by_0(z) = by_0(d/b)$, and we then use equation (D4) to derive α as a function of d:

$$\alpha = e^{\frac{(x-d)^2 - x^2}{2b^2}} \frac{4x - \frac{x^3}{b^2}}{-4(x-d) + \frac{(x-d)^3}{b^2}}, \tag{D6}$$

for $d \geq \sqrt{2.5}b$, and $\alpha = 1$ for $d < \sqrt{2.5}b$, where the limit count rate concides with the count rate I_{src} of the source detected in $x = 0$. The above value of $\alpha(d)$ still depends on $b = \sqrt{a^2 + \sigma_{src}^2}$. For point sources, $\sigma_{src} = \sigma_{\rm PSF}$, while a is not assigned; however, we are interested to have the highest-significance detection, since the above approach is rigorous for infinite signal-to-noise ratio, so that we should choice $a = a_{max} = \sqrt{3}\sigma_{src}$, and therefore $b = 2\sigma_{\rm PSF}$ for point sources.

E. Derivation of Geometrical Parameters for Extended Sources

To derive the geometrical parameters of extended sources (semiaxes and position angle), we now generalize the generating wavelet of equation (2) to the form:

$$\begin{aligned} g(x, y, a_1, a_2, \theta_w) &= \left(2 - \frac{(x\cos\theta_w + y\sin\theta_w)^2}{a_1^2} - \frac{(-x\sin\theta_w + y\cos\theta_w)^2}{a_2^2}\right) \times \\ & \quad e^{-\frac{(x\cos\theta_w + y\sin\theta_w)^2}{2a_1^2} - \frac{(-x\sin\theta_w + y\cos\theta_w)^2}{2a_2^2}} . \end{aligned} \tag{E1}$$

The elliptical source $s(x, y)$ defined in equation (21) will still yield a peak in the WT at the position of its center. The value w_{peak} of the WT at the peak is computed as in equation (1), with $f(x, y) \equiv s(x, y)$, and after some lengthy calculations may be put in the form:

$$w_{peak}(a_1, a_2, \sigma_1, \sigma_2, \theta, I_{src}) = I_{src} \frac{2 + \frac{\sigma_1^2}{a_{\theta_1}^2} + \frac{\sigma_2^2}{a_{\theta_2}^2}}{\left(1 + \frac{\sigma_1^2}{a_{\theta_1}^2} + \frac{\sigma_2^2}{a_{\theta_2}^2} + \frac{\sigma_1^2\sigma_2^2}{a_1^2 a_2^2}\right)^{\frac{3}{2}}}, \tag{E2}$$

where $\theta = \theta_s - \theta_w$, and

$$\frac{1}{a_{\theta_1}^2} = \left(\frac{\cos^2\theta}{a_1^2} + \frac{\sin^2\theta}{a_2^2}\right), \quad \frac{1}{a_{\theta_2}^2} = \left(\frac{\cos^2\theta}{a_2^2} + \frac{\sin^2\theta}{a_1^2}\right). \tag{E3}$$

Of course, the individual position angles θ_s and θ_w are not important, and the WT amplitude depends only on the relative position angle $\theta = \theta_s - \theta_w$.

If $\sigma_1 = \sigma_2$ (isotropic source), it is easy to see that w_{peak} will no longer depend on θ. This is also the case when $a_1 = a_2 = a$, since we then have $a_{\theta_1} = a_{\theta_2} = a$: an isotropic generating wavelet cannot of course yield information on position angles. For $\sigma_1 = \sigma_2 = \sigma_{src}$ and $a_1 = a_2 = a$, w_{peak} reduces to the form given in equation (5) of section 2. As a function of θ, w_{peak} has a maximum for $\theta = 0$ (if $\sigma_1 \geq \sigma_2$ and $a_1 \geq a_2$), and the general expression for $w_{peak}(\theta = 0)$ is:

$$w_{peak}(\theta = 0) = I_{src} \frac{\left(1 + \frac{\sigma_1^2}{a_1^2}\right) + \left(1 + \frac{\sigma_2^2}{a_2^2}\right)}{\left(\left(1 + \frac{\sigma_1^2}{a_1^2}\right)\left(1 + \frac{\sigma_2^2}{a_2^2}\right)\right)^{\frac{3}{2}}}, \tag{E4}$$

dependent only on the two ratios σ_1/a_1 and σ_2/a_2. For σ_1/a_1 and σ_2/a_2 going both to zero, $w_{peak} = 2I_{src}$; the function $y(a_1, a_2) \equiv w_{peak}/\sqrt{a_1 a_2}$, analogous to $y(a) = w(a)/a$ defined in section 2, will therefore have a maximum for finite σ_1/a_1 and σ_2/a_2, which occurs for

$\sigma_1/a_1 = \sigma_2/a_2 = 1/\sqrt{3}$, as in the isotropic case. Contours of equal values of the function $y(a_1, a_2)$ in the $(\sigma_1/a_1, \sigma_2/a_2)$ plane are shown in Figure 12.

The maximum of $y(a_1, a_2)$ as a function of a_1, a_2, and θ_w occurs therefore for $\theta_w = \theta_s$, $a_1 = \sqrt{3}\sigma_1$, $a_2 = \sqrt{3}\sigma_2$. Therefore, to find the semiaxes σ_1 and σ_2 and position angle θ_s of an elliptical source would in principle require the exploration of a three-dimensional parameter space (a_1, a_2, θ_w) to find the maximum of $y(a_1, a_2)$.

However, we know already that the source has been detected with a WT peak at a certain scale a_{max}, using an isotropic WT with $a_1 = a_2 = a$. In this case, the function $y(a, a) = w_{peak}/a$ has the form:

$$\frac{w_{peak}}{a} = \frac{I_{src}}{a} \frac{\left(1 + \frac{\sigma_1^2}{a^2}\right) + \left(1 + \frac{\sigma_2^2}{a^2}\right)}{\left(\left(1 + \frac{\sigma_1^2}{a^2}\right)\left(1 + \frac{\sigma_2^2}{a^2}\right)\right)^{\frac{3}{2}}}, \tag{E5}$$

whose maximum occurs for a scale $a = a_{max}$ satisfying the equation:

$$\frac{3(\sigma_1^2 + \sigma_2^2)\sigma_1^2\sigma_2^2}{a_{max}^6} + \frac{10\sigma_1^2\sigma_2^2}{a_{max}^4} + \frac{\sigma_1^2 + \sigma_2^2}{a_{max}^2} - 2 = 0. \tag{E6}$$

The solution a_{max} of this equation may be written in the form

$$\frac{a_{max}^2}{\sigma_1^2} = f\left(\frac{\sigma_2^2}{\sigma_1^2}\right), \tag{E7}$$

where f is a complicated analytic function. Then, we analyze the source with an anisotropic wavelet with (fixed) scales a_1 and a_2, respectively larger and smaller than a_{max} (for example $a_1 = \sqrt{2}a_{max}$, $a_2 = a_{max}/\sqrt{2}$), and with variable θ_w. Having fixed a_1 and a_2, w_{peak} is now only a function of $\theta = \theta_s - \theta_w$, whose maximum (occurring for $\theta = 0$) indicates the position angle θ_s of the source. The function $w_{peak}(\theta)$ has a period of π, a minimum occurring for $\theta = \pi/2$; the modulation amplitude $A_w = w_{peak}(\theta = 0)/w_{peak}(\theta = \pi/2)$ is a function of σ_1/σ_2, and is very sensitive to this ratio for values $\sigma_1/\sigma_2 \leq 10$, which includes most of the range observed for astronomical sources, with the above choice of a_1,a_2. The function $A_w(\sigma_1/\sigma_2)$ may be inverted numerically to derive σ_1/σ_2; using equation (E7) we then obtain the semiaxes σ_1 and σ_2 of the source, since a_{max} is known. Eventually, the source intensity I_{src} can be derived from equation (E5).

REFERENCES

Ayres,T.R., Linsky,J.L., Vaiana,G.S., Golub,L., Rosner,R. 1981, ApJ, 250, 293

Damiani,F., Maggio,A., Micela,G., Sciortino,S. 1996, ApJ, submitted (Paper II)

Daubechies,I. 1992, Ten Lectures on Wavelets (Philadelphia:SIAM)

Elvis,M., Plummer,D., Schachter,J., Fabbiano,G. 1992, ApJS, 80, 257

Escalera,E., Mazure,A. 1992, ApJ, 388, 23

Escalera,E., Slezak,E., Mazure,A. 1992, A&A, 264, 379

Grebenev,S.A., Forman,W., Jones,C., Murray,S. 1995, ApJ, 445, 607

Harnden,F.R.,Jr., Fabricant,D.G., Harris,D.E., Schwartz,J. 1984, SAO Special Report, n.393

Hasinger,G., Boese,G., Predehl,P., Turner,T.J., Yusaf,R., George,I.M., Rohrbach,G. 1993, MPE/OGIP Calibration Memo CAL/ROS/93-015

Kaiser,G. 1994, A Friendly Guide to Wavelets (Boston:Birkhäuser)

Linsky,J.L., Haisch,B.M. 1979, ApJ, 229, L27

Rosati,P., Burg,R., Giacconi,R. 1994, in The Soft X-ray Cosmos, eds. E.M.Schlegel and R.Petre (AIP Conf. Proc. 313), p.260

Slezak,E., Bijaoui,A., Mars,G. 1990, A&A, 227, 301

Starck,J.-L., Murtagh,F., Pirenne,B., Albrecht,M. 1996, PASP, in press

Vikhlinin,A., Forman,W., Jones,C., Murray,S. 1995, ApJ, 451, 542

Zimmermann,H.U., Belloni,T., Izzo,C., Kahabka,P., Schwendker,O. 1993, EXSAS User Guide, MPE Report 244, Third Edition, Garching:MPI Extraterr. Physik

This preprint was prepared with the AAS LaTeX macros v4.0.

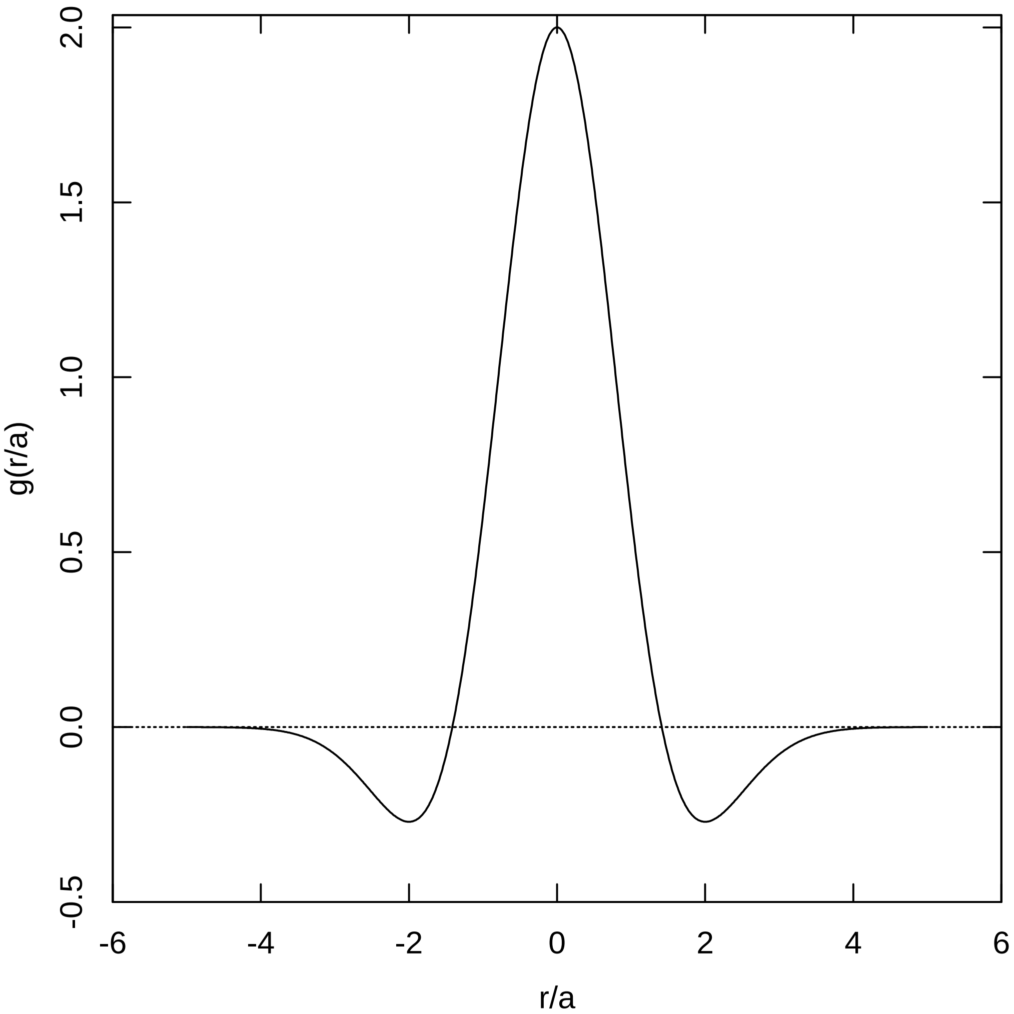

Fig. 1.— Mexican hat generating wavelet $g(r/a)$.

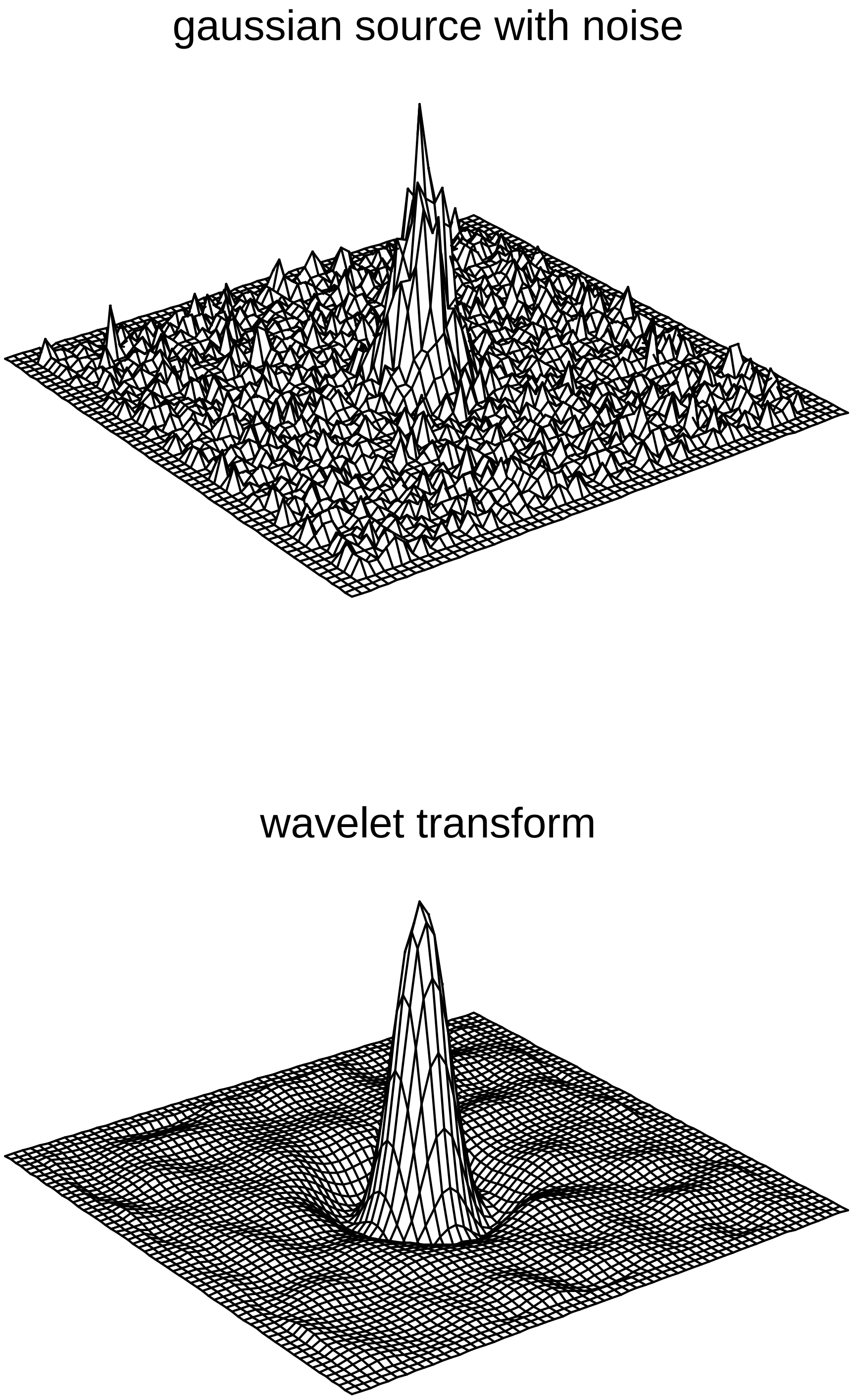

Fig. 2.— (a, upper): Image of a gaussian source over a flat background, with Poisson noise. The gaussian width is $\sigma_{src} = 2$ pixels. (b, lower): Wavelet transform WT of the image in panel a, at scale $a = \sqrt{2}\sigma_{src}$. The negative annulus around the main peak is evident. In both panels the image border is set to zero to provide a reference level, and comparison between a and b shows that the WT of a flat background is zero on average.

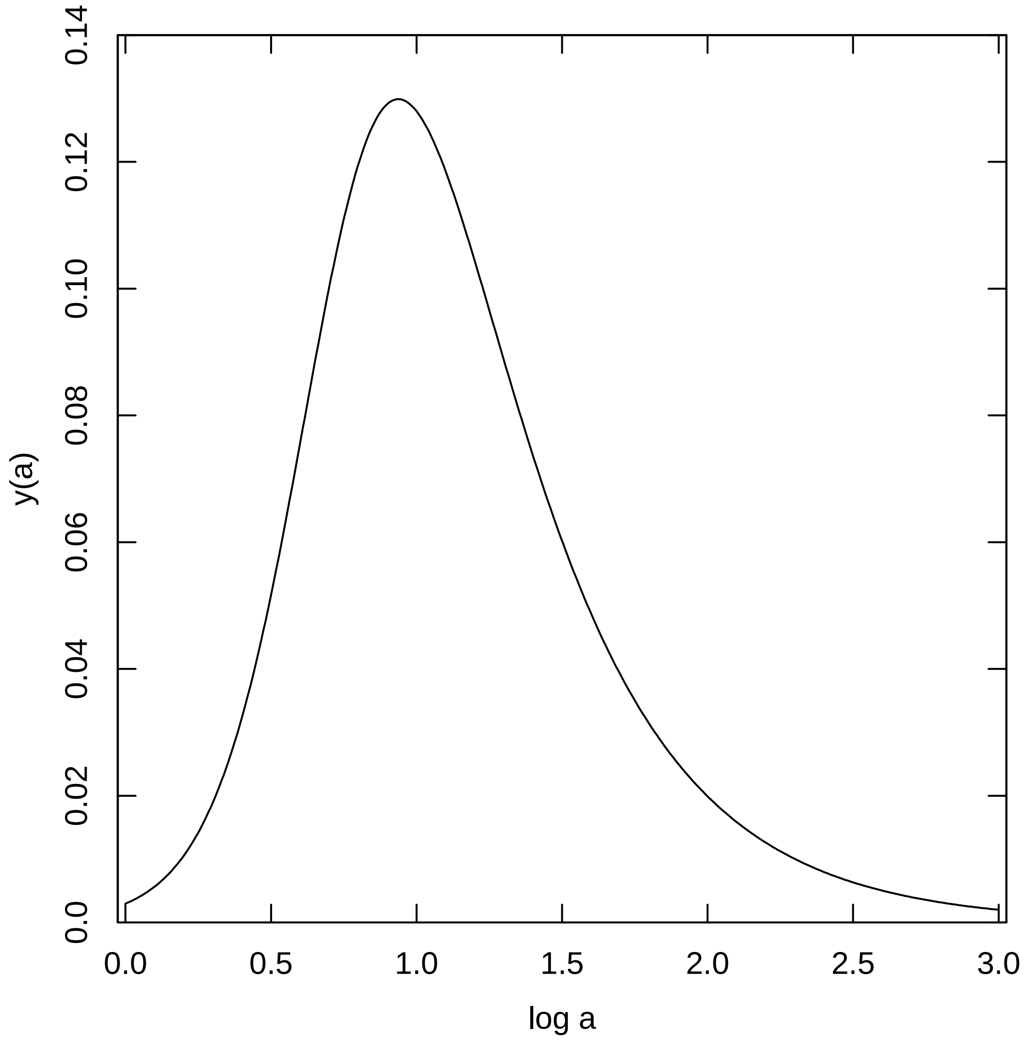

Fig. 3.— Profile of the function $y(a) = w(a)/a$ vs. $\log a$ for a gaussian source.

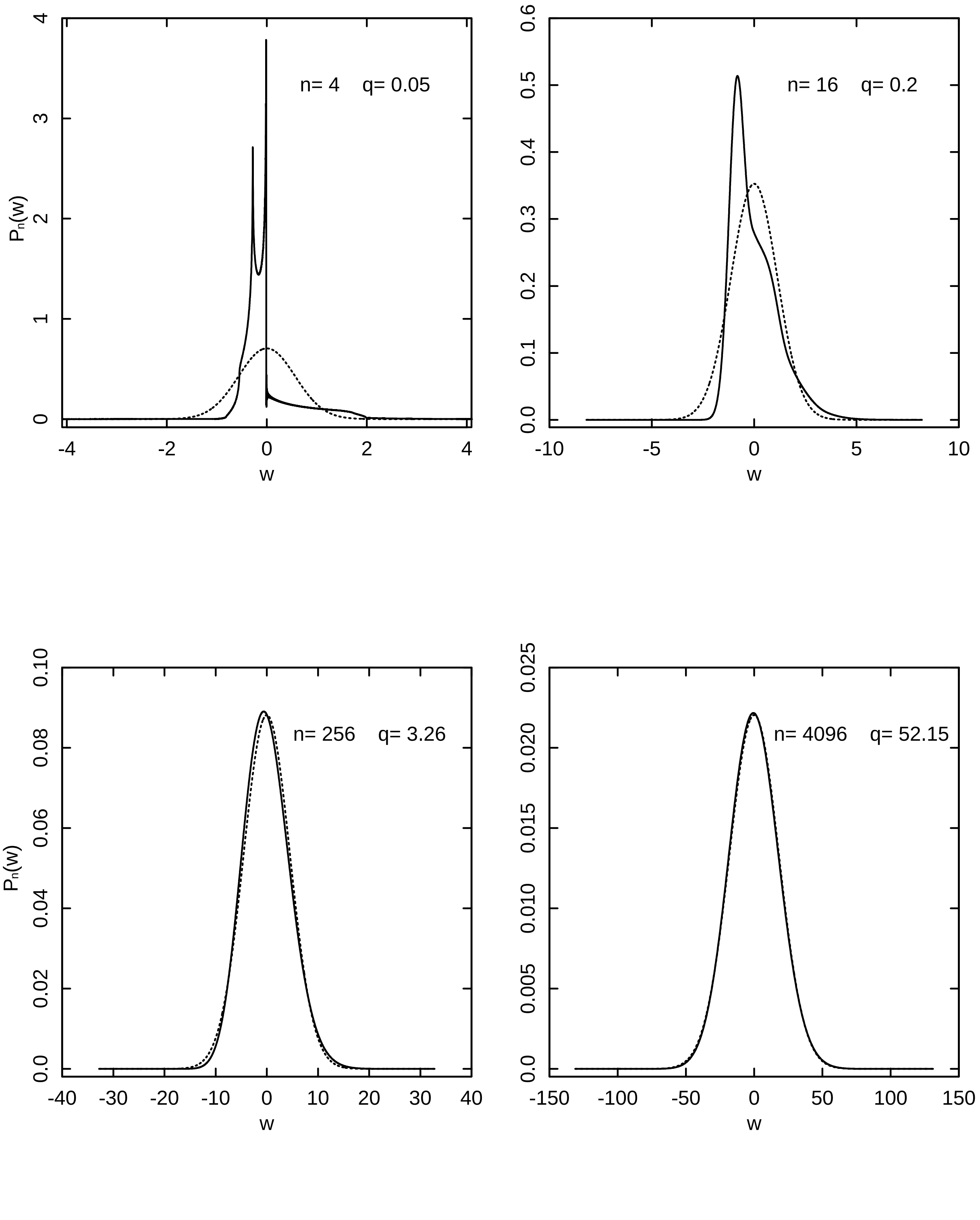

Fig. 4.— Probability distribution $P_n(w)$ (solid line) for n background photons within $25\pi a^2$, for $n = 4, 16, 256, 4096$. Dotted lines show gaussian distributions with variance $\sigma = 2\pi q$, that approach $P_n(w)$ (where $n = 25\pi q$) for large n.

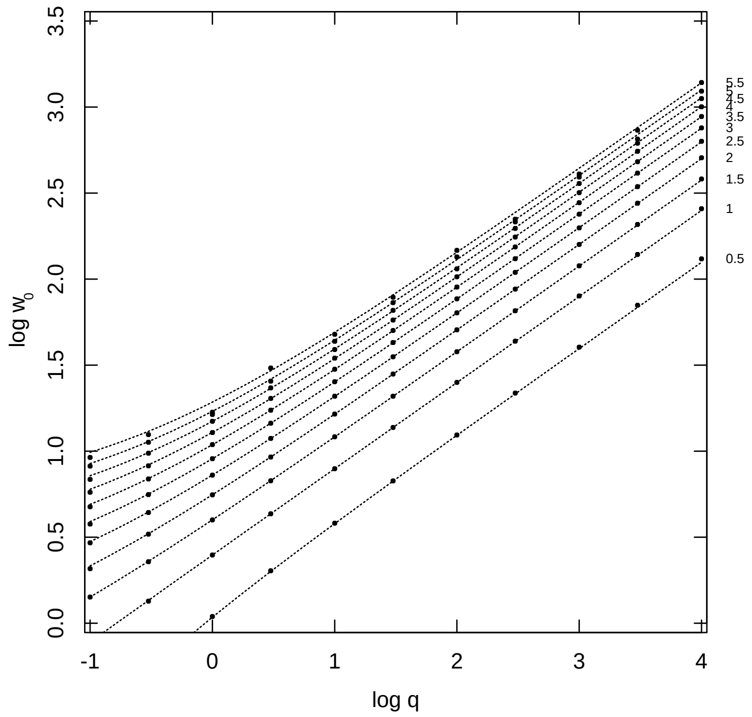

Fig. 5.— Detection thresholds $w_0(k,q)$ (quantiles of $P_q(w)$), for k in the range $0.5-5.5$. $k\sigma$ is the significance threshold, and $q = b \times a^2$, where b is the local background density. Small circles indicate the results of Montecarlo simulations, while dotted curves are the analytic approximations to the simulation data discussed in the text. Each curve refers to a different value of k, indicated on the right-hand side of the Figure. For $q \geq 300$, the dotted curves are close to straight lines, implying that $P_q(w)$ is close to a gaussian.

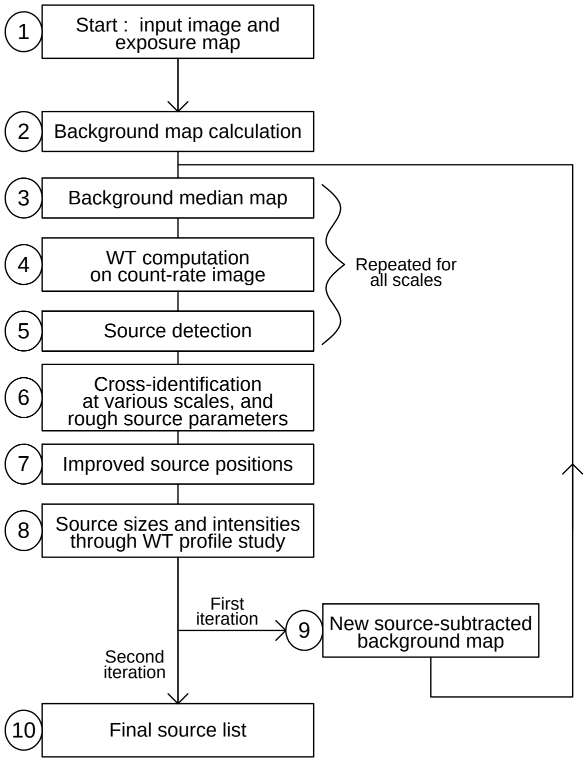

Fig. 6.— Block diagram of the algorithm.

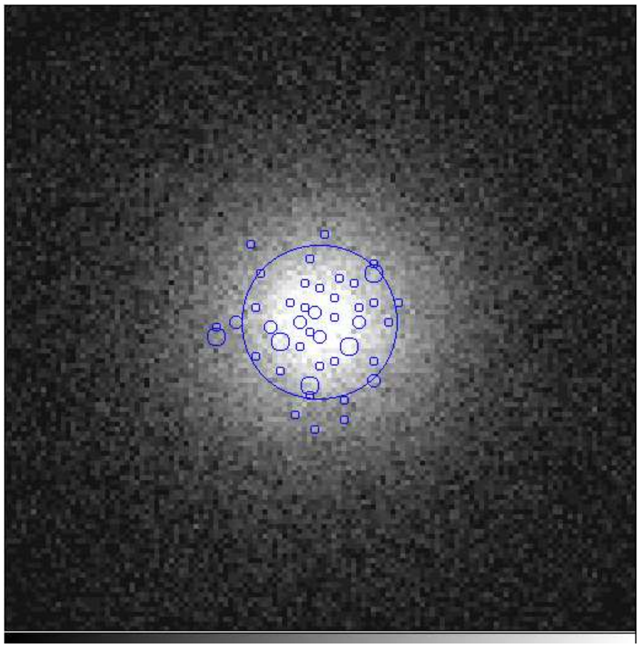

Fig. 7.— Simulation of an extended source, namely much larger than $\sigma_{\rm PSF}$, with indicated positions of detections (circles) obtained by assuming for the whole image the same average background level. The circle radius is equal to the scale of each detection. All sources but the one indicated by the large circle are spurious, and arise because of the bad adopted background estimate.

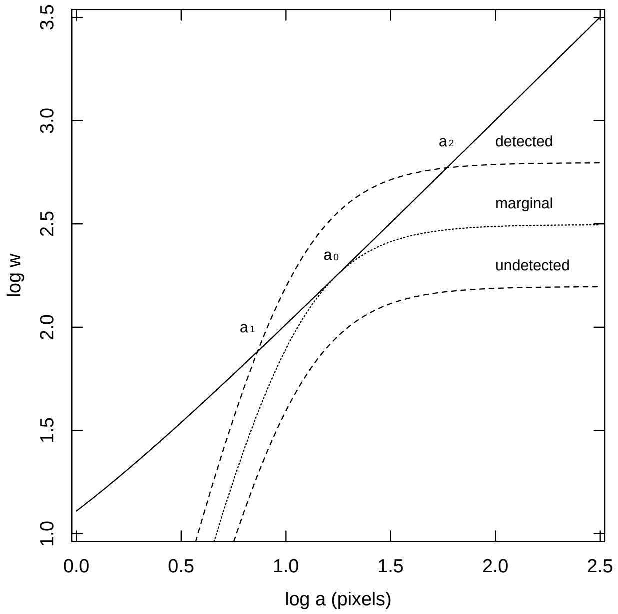

Fig. 8.— Comparison between the function $w(a)$ (dotted and dashed) and detection threshold curve $w_0(k, q)$ (solid), to evaluate upper limits. The threshold curve is computed assuming a background density $b = 1.0$ counts/pixel^2, and the $w(a)$ curves are computed for a source size $\sigma_{src} = 10$ pixels, and three different count rates I_{src} (normalization). The curve $w(a)$ labeled as "detected" is higher than the threshold between the scales a_1 and a_2, and such a source is therefore detected at scales in this range. The curve labeled "undetected" never crosses the threshold, and the source is not detected at any scale a. The "marginal" case (dotted) corresponds to a source which is barely detected at one single scale a_0: the normalization corresponding to such a case yields the upper limit I_{src} on the count rate of an undetected source, with the given size σ_{src} and local background.

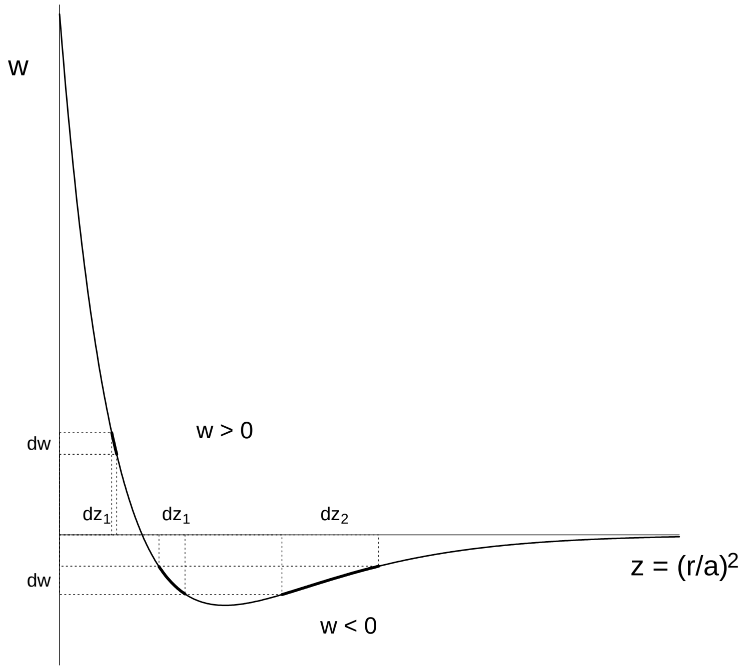

Fig. 9.— The function $w(z)$, and the mapping of an element dw on elements dz_i on the z-axis.

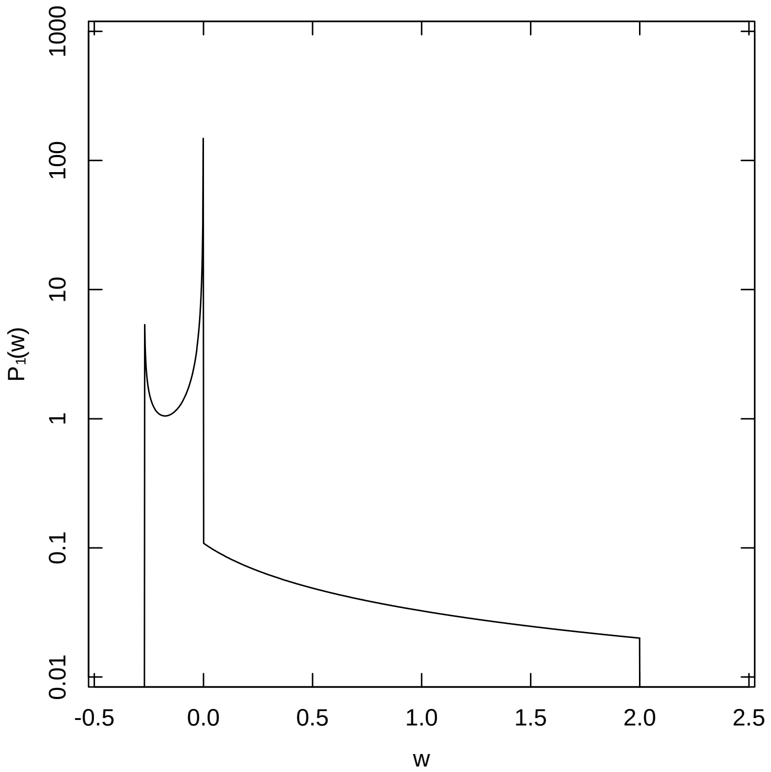

Fig. 10.— Probability distribution $P_1(w)$ for one background photon falling in an area $25\pi a^2$. This distribution is normalized such that $\int P_1(w)dw = 1$, despite $P_1(w) \gg 1$ for some values of w.

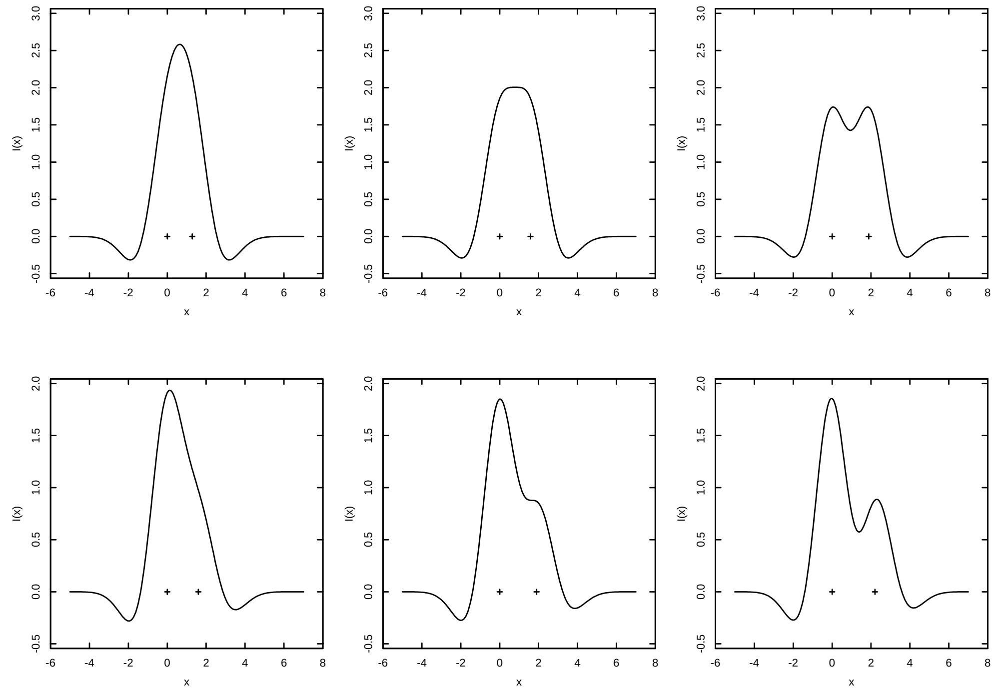

Fig. 11.— Profiles $I(x)$ of the WT of two close sources, along the direction x of their separation; the two crosses at $I(x) = 0$ in each panel mark the position of the two sources, in $x = 0$ and $x = d$ respectively. All sources have a width $b = \sqrt{a^2 + \sigma_{src}^2} = 1$. The three upper panels refer to two sources of equal intensity, while the three lower panels refer to two sources with intensity ratio $\alpha = 0.57$. a (upper left): The two sources are unresolved, since $d < d_{min}(\alpha = 1) = 1.585$. b (upper middle): The two sources are marginally resolved ($d = d_{min}(\alpha = 1)$). c (upper right): The sources are well resolved ($d > d_{min}(\alpha = 1)$). d (lower left): The two unequal-strength sources are unresolved, since $d < d_{min}(\alpha = 0.57) = 1.9$. e (lower middle): The two unequal sources are only marginally resolved ($d = d_{min}(\alpha = 0.57)$), despite their distance is the same of the two resolved sources of panel c, showing the dependence of d_{min} on α. f (lower right): Two unequal sources resolved ($d > d_{min}(\alpha = 0.57)$).

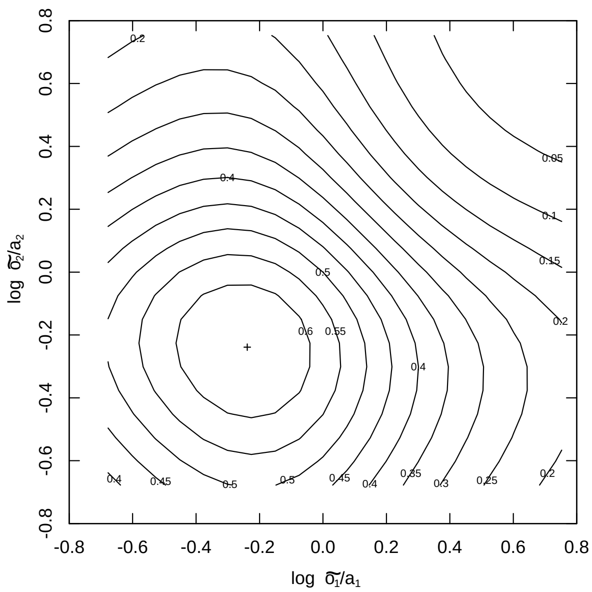

Fig. 12.— Contour plot of the function $y(a_1, a_2) = w_{peak}/\sqrt{a_1 a_2}$ in the $(\sigma_1/a_1, \sigma_2/a_2)$ plane, for a source with total intensity $I_{src} = 1$. The cross indicates the position of the maximum, for $\sigma_1/a_1 = \sigma_2/a_2 = 1/\sqrt{3}$. Contours are labeled with the value of $y(a_1, a_2)$, assuming $\sigma_1 = \sigma_2 = 1$, and at maximum $y(a_1, a_2) = \sqrt{27}/8 = 0.6495$.