

## Experimental neutron spectroscopy data visualization: Adaptive tessellation algorithm

I. Bustinduy and F. J. Bermejo<sup>a)</sup>

*C.S.I.C., Instituto de Estructura de la Materia, Serrano 123, Madrid 28006, Spain*

T. G. Perring

*Rutherford Appleton Laboratory, ISIS Facility, Chilton OX11 0QX, United Kingdom*

G. Bordel

*Department of Electricity and Electronics, University of the Basque Country, P.O. Box 644, Bilbao 48080, Spain*

(Received 19 January 2007; accepted 12 March 2007; published online 9 April 2007)

We report on an adaptive binning approach designed for data visualization within scientific disciplines where counting statistics are expected to follow Poisson distributions. We envisage a wide range of applications stemming from astrophysics to the condensed matter sciences. Our main focus of interest concerns, however, neutron spectroscopy data from single-crystal samples where signals span a four-dimensional space defined by three spatial coordinates plus time. This makes widely used equal-width binning schemes inadequate since physically relevant information is often concentrated within rather small regions of such a space. Our aim is thus to generate optimally binned data sets from one-dimensional to three-dimensional volumes to provide the experimentalist with enhanced ability to carry out searches within a four-dimensional space. Several binning algorithms are then scrutinized against experimental as well as simulated data. © 2007 American Institute of Physics. [DOI: 10.1063/1.2722398]

### I. INTRODUCTION

Since its very inception about six decades ago as an experimental tool for the investigation of structural and dynamic correlations on condensed matter samples (any solid or liquid specimen of physical, chemical, geo- or biological interest), the scattering of slow neutrons has established itself as an unrivaled technique for a wide variety of scientific problems. Experiments carried out using such techniques suffer, however, from severe intensity limitations inherent to present day neutron sources.

The present study focuses onto the most demanding kind of experiments such as those performed on single-crystalline samples using time-of-flight (TOF) spectrometers. Such measurements aim to explore space-time correlations in long-range-ordered matter that, by their very nature, subtend a four-dimensional space where a given point is defined in terms of three spatial coordinates ( $\mathbf{Q}_h, \mathbf{Q}_k, \mathbf{Q}_l$ ) plus a frequency. These constitute the Fourier space counterparts of those three spatial plus time coordinates needed to specify the oscillatory motion of a particle or a spin of electronic or nuclear origin, embedded within a three-dimensional crystal. The triplets  $h, k, l$  are known as Miller indices and are used to specify the locations in reciprocal space where Bragg peaks may occur. Last generation TOF machines have profited from improvements in neutron optics devices (i.e., neutron guides, choppers, and detectors), fast counting electronics, as well as from the increasing capacity for storage and

retrieval of massive data sets. An example of such an instrument is schematically shown in Fig. 1, and several others are being built at present for new neutron sources such as the Spallation Neutron Source project at Oak Ridge (Tennessee, USA) or Japan Proton Accelerator Research Complex (Tokai, Japan). Most of them employ position sensitive detectors (PSDs) that provide close to continuous coverage over large solid angles. In the particular case of the Time-of-Flight neutron spectrometer (MAPS) instrument hosted by the ISIS pulsed neutron source at The Rutherford Appleton Laboratory (United Kingdom), the 16 m<sup>2</sup> area is approximately divided into 36 864 square detector pixels. A total of  $\approx 160\,000$  detector pixels are used there in the forward direction. In addition, narrower strips of detectors provide additional angular coverage within the horizontal plane. The raw data each detector registers consists on histograms representing neutron counts versus arrival time taken with respect to an external reference. A data set thus comprises histograms of neutron counts versus time stored by each individual detector which sits at some given angular coordinates with respect to the incident beam.

Because of intensity limitations inherent to neutron scattering techniques, one usually needs to trade off instrumental resolution (the resolving power) versus counting statistics. The former, which concerns both momentum and energy transfer, is usually controlled by judicious choosing of different incident energies of the incoming neutron beam as well as proper instrument setups. Typical production runs from an instrument such as MAPS will deliver data sets from some 10<sup>8</sup> detector pixels. Such a large number of picture

<sup>a)</sup>Also at Department of Electricity and Electronics, University of the Basque Country, P.O. Box 644, Bilbao 48080, Spain.

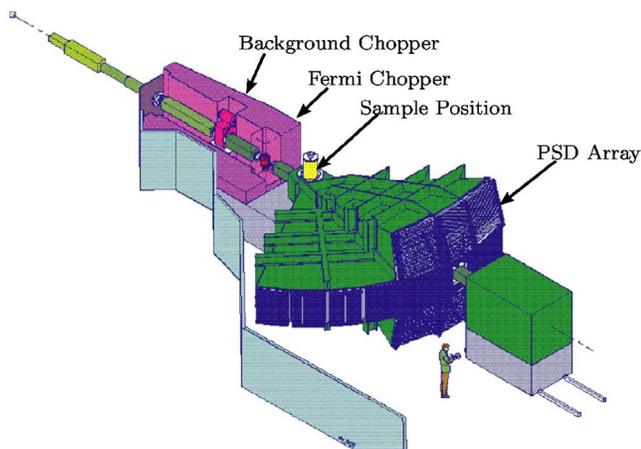


FIG. 1. (Color online). A schematic drawing of the MAPS TOF chopper spectrometer at the ISIS neutron source (sited at Rutherford Appleton Laboratory, UK). A beam of incoming neutrons is monochromatized by means of neutron velocity selector device previous to collision with the sample of interest. Scattered neutrons are then counted by detectors covering large solid angles. Such detectors forming PSD arrays count neutron events at a given time after collision with the sample specimen. The figure identifies the large PSD array as well as additional detectors (dark surfaces within the horizontal plane). The detector-pixel size in reciprocal space is significantly smaller than the resolution volume defined by other instrumental contributions.

elements thus offer the experimentalist a good chance for high-speed visualization and online data assessment, provided that adequate software visualization tools are available.

The data are represented as an intensity (i.e., neutron counts) measured along some trajectories within wave-vector energy space versus time of arrival at a given detector-pixel position. For a given detector element such trajectories are set by the neutron kinematics once the energy of the neutrons impinging upon the sample is selected. They define a two-dimensional (2D) grid and so a set of trajectories defines a data volume. The data format of interest for most scientific applications corresponds to projections within a four-dimensional space constituted by the three components of the momentum-transfer vector  $\mathbf{Q}$  as well as a scalar quantity that is the energy transfer ( $\epsilon$ ). Practical performance of such measurements by the TOF technique requires to set one of the variables to a fixed value, which reduces the dimensionality of the problem (see Ref. 1 for more details). In the sequel we will thus deal with three-dimensional problems. It is also worth noticing that the end user will most of the time carry out data analysis on 2D slices taken from the whole three-dimensional (3D) data volume.

Studies on single crystals constitute extreme cases in difficulty because the regions where physical information is concentrated are usually small. The instrumental resolution achievable by the present generation of instruments is relatively high, typically hundredths of  $\text{\AA}^{-1}$  ( $10^{10} \text{ m}^{-1}$ ) for resolution in reciprocal space and fractions of meV ( $\approx 247 \text{ GHz}$ ) in frequency. However, in practical applications one needs to trade off signal to noise ratio (SNR) and resolution in order to increase the statistical significance of measured data. Because of the highly pixelated nature of current detector systems, a balance between resolution and

statistics usually involves some sort of averages taken over spatially close data. Amongst the commonly used averaging techniques, smoothing and equal-width binning (EWB) are the most popular. Single-crystal samples usually show large variations of SNR across the detector elements and, under such circumstances, smoothing algorithms are known to introduce some artifactual correlations affecting neighboring data which are difficult to quantify. This limits the usefulness of smoothing techniques for quantitative data analysis. On the other hand, EWB techniques average spatially close data by using an integration step that remains constant along the entire data set. By their very nature are nonadaptive techniques and very often high statistics areas are unnecessarily blurred in order to obtain a better view of the whole image. Such considerations become even more important if we consider that on the one hand some experiments require, at present, the combination of different runs, and on the other hand, next generation instruments like MERLIN (sited at the ISIS pulsed neutron and muon Source, United Kingdom) are expected to produce 20 times larger data sets.

Because of reasons given above, it becomes advisable to use adaptive binning algorithms where the bin sizes are adapted to local SNR regions: large bins result from low statistics areas, while smaller bins will mark high statistics regions. Several different adaptive binning schemes have been reported mostly dealing with studies on color histograms,<sup>2</sup> compound partitioning,<sup>3</sup> flow cytometry,<sup>4-6</sup> or Dalitz diagrams.<sup>7,8</sup> In this work we will build upon previous works that dealt with processing of astrophysical images. The work by Sanders and Fabian<sup>9</sup> reports on an analysis technique that involves subdividing a two-dimensional image into blocks that are more homogeneous than the image itself, thus revealing information about the structure of the image. In contrast, a more adaptive approach is described by Cappellari and Copin<sup>10</sup> that employs Voronoi tessellations to produce an optimal 2D binning algorithm, or a more recent work by Diehl and Statler,<sup>11</sup> where weighted Voronoi tessellations are used. The purpose of these approaches is to obtain a SNR target value as homogeneous as possible all over the image. As a consequence, high statistics areas are blurred in order to achieve the desired image homogeneity. In addition to these contributions, we have previously reported on a three-dimensional algorithm<sup>1</sup> in which high-SNR areas remain in the final image without being blurred. However this approach cannot guarantee that all bins that compose the image will be above the specified SNR target value.

The aim of the present paper is to present adaptive binning algorithms that preserve high-SNR features free from artifacts and are easy to implement in three or higher dimensions. This paper is organized as follows. Section II introduces the Voronoi diagrams, which are the basis of the algorithm presented in Sec. III; Sec. IV discusses the results obtained from different sorts of reconstructions. Finally, Sec. V discusses the conclusions.

## II. BACKGROUND

From now on, the term *polygon* or *bin* will stand for a collection of detector pixels we aim to analyze, and *pixel*,

unless explicitly stated otherwise, will be used as a shortcut for detector pixel.

Let us consider that each pixel has an intensity  $s_i$  and an associated uncertainty or error bar  $n_i$ . If we bin  $m$  pixels together, the resulting intensity  $S$  and associated error values  $N$  for the bin are calculated as follows:

$$S = \frac{1}{m} \sum_{i=1}^m s_i \quad \text{and} \quad N = \frac{1}{m} \sqrt{\sum_{i=1}^m n_i^2}. \quad (1)$$

### A. Previous work: Multiresolution algorithm

The *multiresolution algorithm* (MA), described in Ref. 1, creates an adaptive volume model from discrete data. It shares a common viewpoint with Sanders and Fabian<sup>9</sup> in the sense that for a given data set, the algorithm runs on a selected number of levels sequentially. Let  $L$  denote that number. At each level a *grid of bins* is defined. The grid resolution (that is, the bin size) for each direction  $u_1, u_2, u_3$  is established by the user.

Unlike the *quadtree method*,<sup>12</sup> which is performed from largest to smaller bin sizes, this multiresolution algorithm goes from smaller up to larger sizes, and removes already collected pixels in order to avoid possible correlation between neighbors. This algorithm has been easily extended to problems in higher dimensions. For a more in-depth discussion of MA, see Ref. 1.

### B. Voronoi tessellation

Although MA displays some interesting features (it is fast, robust, and easy to extend to higher dimensions), its usefulness is limited because the variation of bin size takes place in steps that are a power of two of the spatial dimension. This calls for more adaptive ways of image reconstruction. Searching for methods not limited to rectangular bins, one finds the *Voronoi tessellation* (VT) scheme, which has been applied to a wide variety of problems within fields as diverse as astronomy, biology, image and data analysis, resource optimization, sensor networks, geometric design, and so on.<sup>13-20</sup>

Given a set  $P$  of  $n \geq 2$  points  $p_1, p_2, \dots, p_n$  in a  $d$ -dimensional Euclidean space  $\mathbb{R}^d$ , the idea beneath VT is to divide the space into regions around each point in  $P$ , such that all the points in the region around  $p_i$  are closer to this one than to any other point in  $P$ . Consider now two points  $p_i$  and  $p_j$  in  $P$ , and write  $D(i, j)$  for the subset of points at least as close to  $p_i$  than to  $p_j$ ; that is,

$$D(i, j) = \{r \in \mathbb{R}^d : \|r - p_i\| \leq \|r - p_j\|\}, \quad (2)$$

where  $\|\cdot\|$  stands for the Euclidean norm. Note that  $D(i, j) \neq D(j, i)$ . We call the *Voronoi region* of  $p_i$  with respect to  $P$ ,

$$V(i, P) = \bigcap_{j \in [1, n] \text{ and } i \neq j} D(i, j). \quad (3)$$

The point  $p_i$  is called a *generator*. Finally, the *Voronoi diagram*  $V_P$  of the set  $P$  is defined as

$$V_P = \bigcup_{i, j \in [1, n] \text{ and } i \neq j} \bar{V}(i, P) \cap \bar{V}(j, P) \quad (4)$$

$\bar{V}$  denotes the closure of  $V$ .

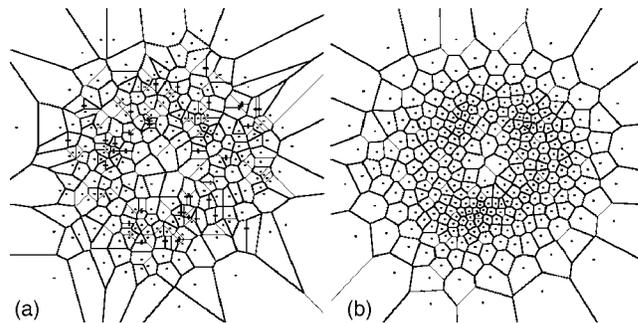


FIG. 2. (a) Voronoi tessellation. (b) Centroidal Voronoi tessellation. On the left, the Voronoi tessellation corresponding to 300 randomly points. The underlying density function is an analytical ring-shape synthetic model. The dots are the Voronoi generators. Note that the generators and the centroids do not need to coincide. On the right, a centroidal Voronoi tessellation; the dots are simultaneously the generators for the Voronoi tessellation and the centroids of the Voronoi regions.

Each Voronoi region  $V(i, P)$  is by definition the intersection of  $m-1$  open, convex, and in some cases unbounded polytopes. The set  $V_P$  of regions  $V(i, P)$ , also known as *Dirichlet domain*, tessellates the space surrounding  $P$ ; it forms a decomposition of the space. See Fig. 2 for an illustration.

Given a density function  $\rho = \rho(r)$  defined on  $V_P$ , one can calculate the *centroids* (defined in the sense of center of mass) for each Voronoi cell as

$$z_i = \frac{\int_{V(i, P)} r \rho(r) dv}{\int_{V(i, P)} \rho(r) dv}. \quad (5)$$

A Voronoi tessellation of a given set such that the associated generators are centroids (denoted by  $\hat{z}_i$ ) of the corresponding Voronoi regions is called a *centroidal Voronoi tessellation* (CVT). The most significant property of CVT is the *minimization property* of the following functional:

$$F(V_P, P) = \sum_{i=1}^n \int_{V(i, P)} \rho(r) \|r - p_i\|^2 dv. \quad (6)$$

This functional may represent whatever magnitude of interest such as energy, village population, cost, any sort of error, etc. If the points in  $P$  are centroids of its corresponding Voronoi cells, the functional  $F$  is minimized. This is the reason why CVTS have been applied to a wide number of fields. Several approaches to construct CVTS are reviewed in Ref. 13 and between those the deterministic method known as *Lloyd's algorithm* (see Ref. 21) has been perhaps the mostly used. Following this algorithm, a CVT is constructed by first determining the Voronoi regions of a given point set, second, moving the generator of each region to the centroid of that region, and third, iteratively repeating these two steps until generators and centroids of each region converge. In the following section we will see how we could profit from CVTS.

### III. NEW ADAPTIVE APPROACH

Now, let us return to the binning problem in neutron spectroscopy. One of the most interesting characteristics of CVTS is the ability to partition a region into bins whose sizes vary as a function of the underlying density distribution. Therefore, if this density function is related to SNR, bin sizes will be directly related to SNR. This question was originally addressed for optimal vector quantizers used in data compression and transmission by the so called *Gersho's conjecture*<sup>22</sup> and translated into the language of CVTS by Du et al.<sup>13</sup>

For one-dimensional problems, Gersho's conjecture states that considering that the density is bounded and strictly positive, the size of the Voronoi intervals are inversely proportional to one-third the power of the underlying density at the midpoints of the intervals. This conjecture was used and extended to  $n$ -dimensional problems by Capellari and Coppin<sup>10</sup> to use CVTS to partition a segment into equal SNR intervals. In an  $n$ -dimensional problem, if the CVT is performed on a density function  $\rho(r)=(S/N)^{(2+n/n)}(r)$ , this generates a VT with bins that asymptotically enclose a constant mass (constant SNR) according to the density  $S/N$ . However, the obtained VT is strongly dependent of the discrete nature of input data. Consequently a method to find the generators for the optimal VT was developed, the *bin-accretion* algorithm. Nevertheless, all these results are limited to the case where the underlying density adds in quadrature, i.e.,  $(S/N)_{1+2}^2=(S/N)_1^2+(S/N)_2^2$ .

In contrast to the cited work, where the sought objective is to obtain an equal SNR bins scheme, we want to bin low statistics areas in order to obtain a SNR above an established threshold while preserving high statistics areas from being unnecessarily blurred. In consequence, we need to find a cri-

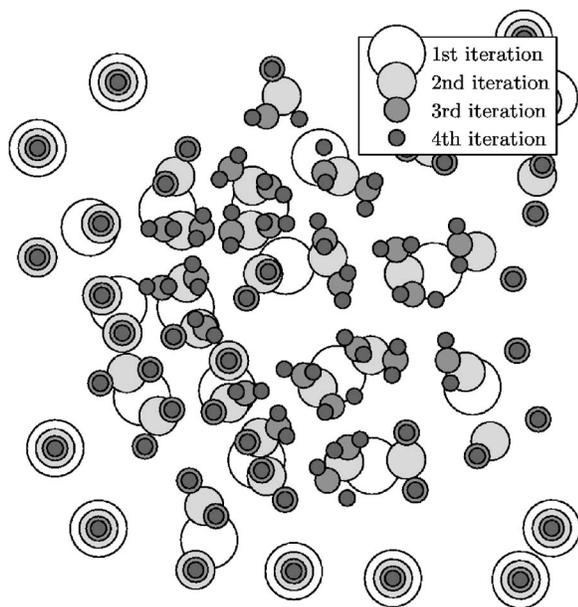


FIG. 3. ATA approach illustration. Consider an initial set of generators (white circles) placed by means of a Monte Carlo method. The number of iterations is equal to 4. Centers of the circles represent the location of the generator for each iteration, therefore concentric circles show that the generator does not move in the successive iterations.

terion to achieve the desired result. It is clear that, given a CVT where the underlying density is a function of SNR,  $\rho=(S/N)^m$ , the higher the value of  $m$ , the sharper the variations of the density associated to CVT will be. Hence, for a given number of generators, high density areas will be represented by more generators at the expense of lower density areas. In order to obtain a more general approach, this parameter has been left up to the potential user. According to our preliminary studies,  $m=2$  offers a good compromise. Thus, in order to perform the successive tests,  $\rho$  will be set up as  $\rho=(S/N)^2$ .

In the following subsection we will present a different approach that pursues the reconstruction of the observed data in a more adaptive way, where the total number of pixels is expected to be collected above an imposed threshold.

#### A. The adaptive tessellation algorithm

The adaptive tessellation algorithm (ATA) is not a multilevel approach, like MA. In ATA, the user selects the num-

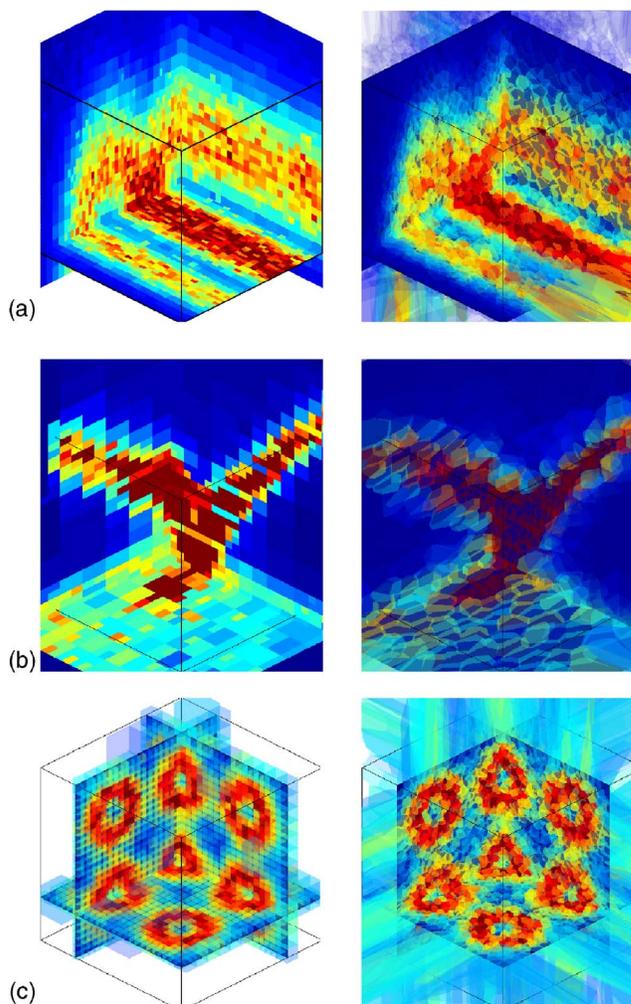


FIG. 4. (Color online) (a)  $u_2 \uparrow \langle u_3 \rangle$  1D dispersion along the chain direction (001) as a function of energy. (b)  $u_3 \uparrow \langle u_2 \rangle$  2D inelastic magnon scattering. (c)  $u_3 \uparrow \langle u_1 \rangle$  3D spin wave modes. 3D reconstructions using different inputs: 1D (KCuF<sub>3</sub> with  $u_1=[0,0,\mathbf{Q}_l,0]$ ,  $u_2=[0,0,0,\epsilon]$ ,  $u_3=[0,\mathbf{Q}_k,0,0]$ ), 2D (Rb<sub>2</sub>MnF<sub>4</sub> with  $u_1=[\mathbf{Q}_h,0,0,0]$ ,  $u_2=[0,\mathbf{Q}_j,0,0]$ ,  $u_3=[0,0,0,\epsilon]$ ), and 3D (simulated cobalt with  $u_1=[\mathbf{Q}_h,0,0,0]$ ,  $u_2=[0.5\mathbf{Q}_h,-1\mathbf{Q}_k,0,0]$ ,  $u_3=[0,0,\mathbf{Q}_l,0]$ ) magnetic system materials. Left figures show the reconstruction performed by means of MA approach; right figures unveil the reconstruction performed by ATA algorithm.

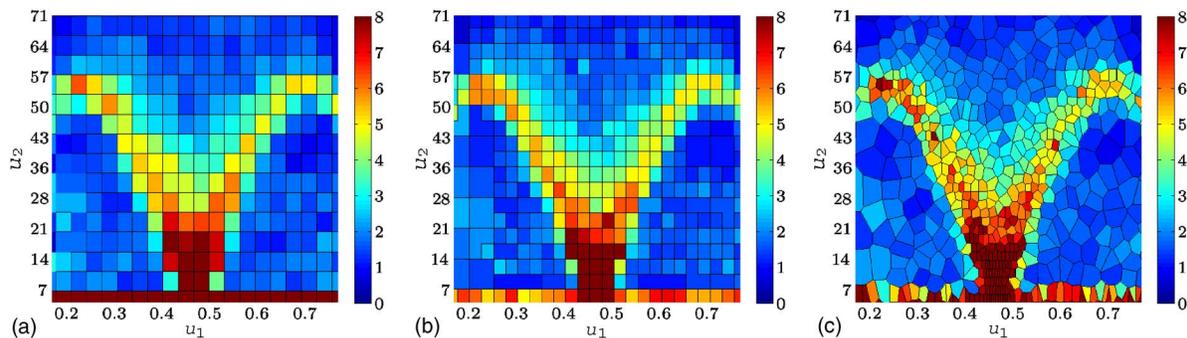


FIG. 5. (Color online) (a) EWB:  $\Delta u_1 = \Delta u_2 = 0.032$ . (b) MA: maximum bin size in both dimensions is 0.101. (c) ATA:  $k=100$ . Spin waves on the near-critical quantum antiferromagnet  $\text{KCuF}_3$ , sliced in the projection space defined by  $u_1 = [0, 0, \mathbf{Q}_j, 0]$  and  $u_2 = [0, 0, 0, \mathbf{e}]$ :  $T=5$  for all algorithms and  $L=3$  for all algorithms except ATA, which does not require this parameter. Color map represents intensity (neutron counts).

ber of initial generators, and a Monte Carlo method places them on the data set. The user also selects a SNR threshold. The first step is to calculate a CVT (with a density function related to SNR) for that set of generators. Thus, the starting point is a CVT, centroids with their associated centroidal Voronoi regions. Now, the algorithm iterates (see Fig. 3) the following four steps until convergence, which will be defined later.

- (1) Verify the SNR condition for all centroidal Voronoi regions. In those regions that satisfy the condition, add a new generator.
- (2) Calculate the CVT for the new set of generators (centroids of the previous CVT plus new generators, all together).
- (3) Verify the SNR condition for each new centroidal Voronoi region. Remove centroids corresponding to regions that do not satisfy the condition.
- (4) Calculate the CVT of the remaining set of generators (centroids of the previous CVT minus centroids of discarded regions).

The question now is when to stop the iterative process. Let us call  $n \in \mathbb{N}$  the number of iterations, and  $|\hat{z}_n|$  the number of centroids the algorithm returns after iteration  $n$ . This defines a sequence with an unknown limit. The Cauchy convergence criterion states that a sequence  $\hat{z}_n$ , with  $n$  in  $\mathbb{N}$  for all  $n$ , is convergent if and only if for every  $\epsilon > 0$ , there exists an  $N \in \mathbb{N}$  such that  $n, m > N$  implies  $\|\hat{z}_n - \hat{z}_m\| < \epsilon$ . That is, the algorithm will converge if there is a number of iterations

from which the number of centroids ( $|\hat{z}_i|$ ) get closer and closer to each other. By default,  $\epsilon = |\hat{z}_n|/100$  in ATA, but this value can be changed by the users to suite their needs.

Compared to MA, ATA constitutes a simpler approach. The users set the initial number of generators, the SNR condition, and possibly, the value of  $\epsilon$  that controls the convergence. As a result, this unsupervised classification of pixels into bins seems more suitable for a first-time exploration process.

The ATA approach has been partly inspired by clustering techniques used in speech recognition and compression. For instance, the LGB algorithm<sup>23</sup> and the ATA approach share some common characteristics such as the idea of splitting the centroids to minimize the distortion. The ATA approach can then be described as a *partitional clustering* algorithm.

The next section will show some results to illustrate the behavior of the algorithm presented in this paper.

## IV. RESULTS

### A. Metrics-data preparation

We have chosen several metrics to evaluate the results obtained from the algorithms: the SNR mean value of bins forming the image; the number (percentage) of pixels collected into bins with a SNR above an imposed threshold, and for simulated data, distortion (difference between original and reconstructed images) was also evaluated by means of the mean square error (MSE):

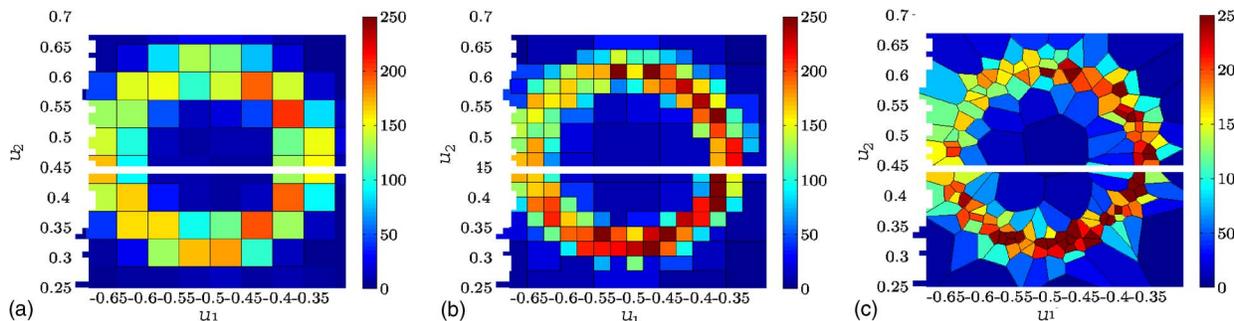


FIG. 6. (Color online) (a) EWB:  $\Delta u_1 = \Delta u_2 = 0.0461$ . (b) MA: maximum bin size in both directions is 0.098. (c) ATA:  $k=100$ . The spin wave spectrum of the 2D Heisenberg antiferromagnet  $\text{Rb}_2\text{MnF}_4$ , where the projection space is defined by  $u_1 = [\mathbf{Q}_h, 0, 0, 0]$  and  $u_2 = [0, 0, \mathbf{Q}_l, 0]$ :  $T=4$  for all algorithms and  $L=3$  for all algorithms except ATA, which does not require this parameter. Color map represents intensity (neutron counts).

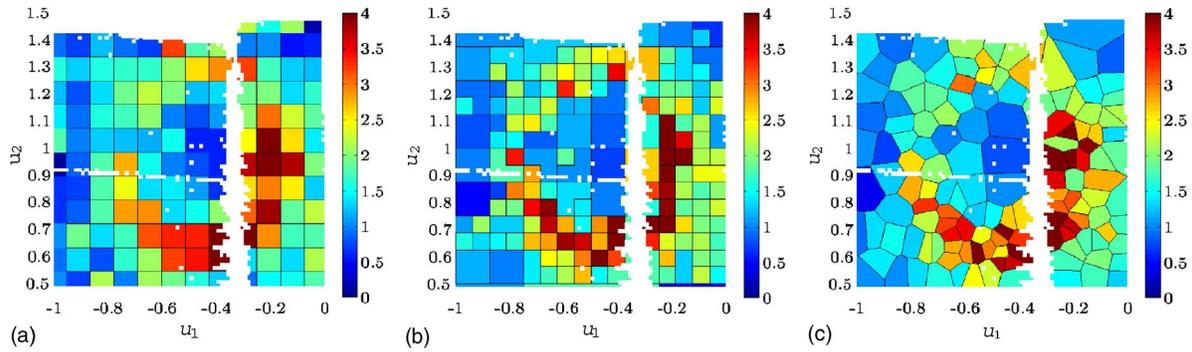


FIG. 7. (Color online) (a) EWB:  $\Delta u_1 = \Delta u_2 = 0.088$ . (b) MA: maximum bin size in both dimensions is 25.8. (c) ATA:  $k = 100$ . Collective motions of magnetic moments within cobalt; the spin dispersion volume is sliced in the projection space defined by  $u_1 = [\mathbf{Q}_h, 0, 0, 0]$  and  $u_2 = [0.5\mathbf{Q}_h, -\mathbf{Q}_k, 0, 0]$ ;  $T = 2.86$  for all algorithms and  $L = 3$  for all algorithms except ATA, which does not require this parameter. Color map represents intensity (neutron counts).

$$\text{MSE} = \frac{1}{m} \sum_{i=1}^m (s'_i - s_i)^2. \quad (7)$$

In this equation,  $s_i$  is the value of the  $i$ th pixel in the original image,  $s'_i$  is the value of the  $i$ th pixel in the reconstructed one, and  $m$  is the total number of pixels in the whole image. The lower the value of MSE, the smaller the difference between images and therefore a better result.

Notice that strict comparisons between the performance of the different approaches is specially difficult since the operation is accomplished in a very different way. The MA requires some level of knowledge of the problem, whereas ATA is an unsupervised process where the user only needs to specify a threshold value  $T$ . Furthermore, none of the presented approaches use the same input parameters, exception made of the threshold value  $T$ .

The *definition* of an image can be described as the degree of detail in the graphic image. If the bins that made the image tessellate the region properly, without holes or overlaps, then the number of bins will be intimately related to the image definition.

We next compare the performance of ATA and in order to provide a comparison in a known context, EWB and MA results are also given. The operation of all these algorithms is very different, although a comparison is still possible as long as the final image reconstructed by them has a similar number of bins. To achieve this, let us remember that ATA is the simplest algorithm from the user's point of view. By *simplest* we mean that once the SNR threshold is set, there is little we can do to control the number of final bins in the image since it is an unsupervised process. So, ATA will determine the number of bins within the images, and with that value in mind, the input parameters of the rest of the algorithms are tuned to achieve such a number, thus making a comparison between them possible.

## B. Type of data

To exemplify the performance of the algorithms reported here, we made use of two kinds of data, are simulated data blurred by statistical noise, and experimental data collected in real experiments carried out in the spectrometer sketched in Fig. 1. Simulated data consist of a pair of images blurred by random white noise as well as other characteristic uncer-

tainties associated with experiments (e.g., instrument background). The experimental data consist of different measurements of one-, two-, and three-dimensional magnetic systems: the thermal fluctuations and magnetic order of the near-critical quantum antiferromagnet  $\text{KCuF}_3$ ,<sup>24</sup> the spin wave spectrum of the 2D Heisenberg antiferromagnet  $\text{Rb}_2\text{MnF}_4$ ,<sup>25</sup> and the collective motions of magnetic moments within a metallic magnet such as cobalt.<sup>26</sup>

The simulated examples serve to evaluate the algorithms as far as the reliability of the reconstruction of an image that has been blurred by statistical Poisson noise is concerned. On the other hand, real data pose additional problems such as those derived from the finite efficiency of the detector sys-

TABLE I. Metrics from the 2D implementation. (A) Simulated data and (B) experimental data.

(A) Simulated data				
Data	Algorithm	MSE	$ \text{SNR}  \pm \text{std}(\text{SNR})$	Collected population (%)
Ring	MA	836.22	$3.8 \pm 0.6$	98.97
	ATA	844.87	$3.9 \pm 0.5$	100.00
	EWB <sup>a</sup>	802.69	$4.8 \pm 1.5$	91.09
$\text{KCuF}_3$	MA	92.62	$2.1 \pm 0.3$	99.83
	ATA	98.15	$2.1 \pm 0.2$	100.00
	EWB <sup>b</sup>	90.57	$2.3 \pm 0.6$	86.78
(B) Experimental data				
Data	Algorithm		$ \text{SNR}  \pm \text{std}(\text{SNR})$	Collected population (%)
Cobalt	MA		$3.7 \pm 0.6$	99.13
	ATA		$3.9 \pm 0.5$	100.00
	EWB <sup>c</sup>		$4.1 \pm 1.3$	84.02
$\text{Rb}_2\text{MnF}_4$	MA		$5.5 \pm 1.8$	88.50
	ATA		$5.7 \pm 0.2$	100.00
	EWB <sup>d</sup>		$7.6 \pm 4.4$	66.91
$\text{KCuF}_3$	MA		$7.6 \pm 3.1$	96.10
	ATA		$7.3 \pm 2.2$	100.00
	EWB <sup>e</sup>		$7.3 \pm 5.0$	59.35

<sup>a</sup> $\Delta u_1 = \Delta u_2 = 6.2$

<sup>b</sup> $\Delta u_1 = \Delta u_2 = 5.15$ .

<sup>c</sup> $\Delta u_1 = \Delta u_2 = 0.088$ .

<sup>d</sup> $\Delta u_1 = \Delta u_2 = 0.0461$ .

<sup>e</sup> $\Delta u_1 = \Delta u_2 = 0.0322$ .

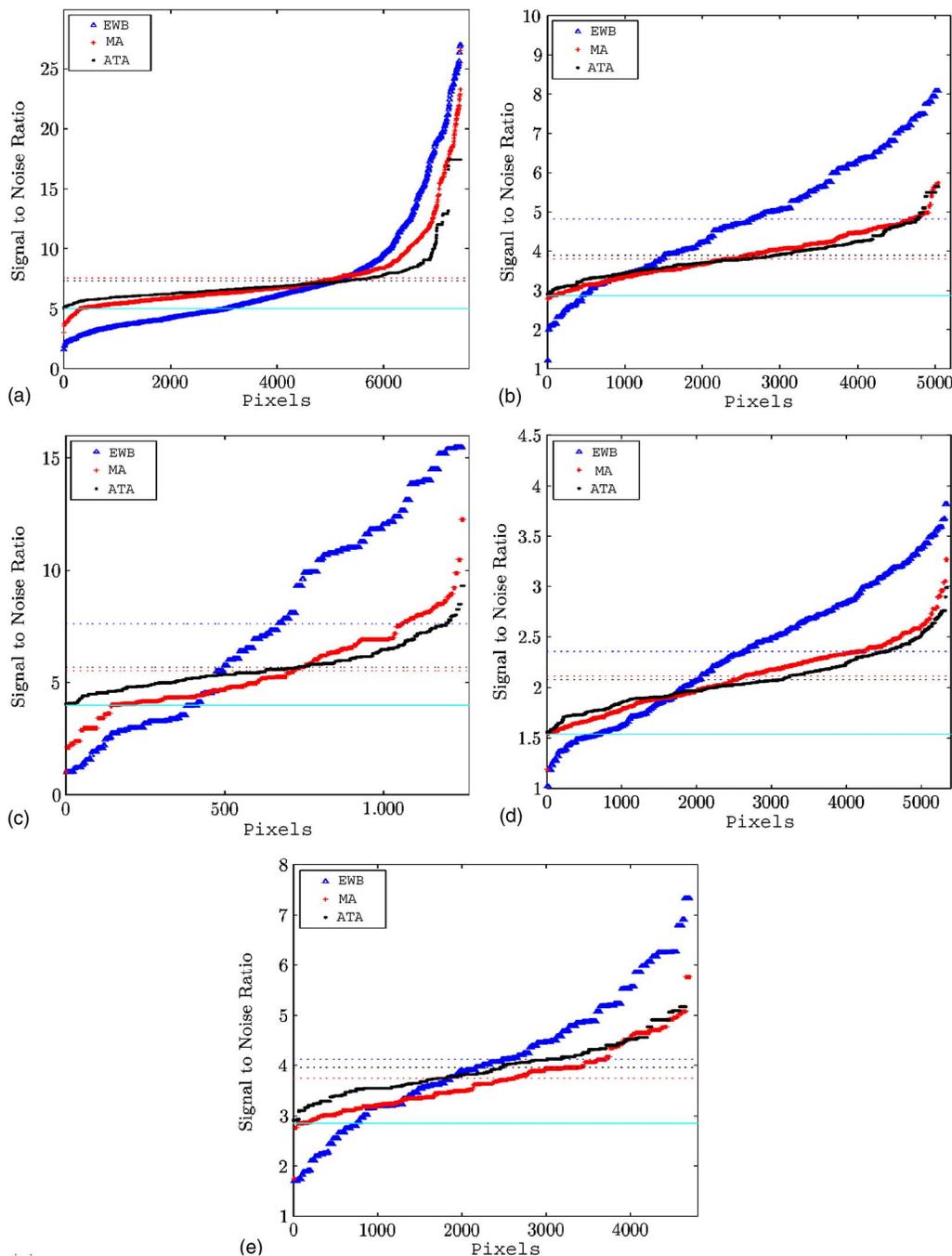


FIG. 8. (Color online) (a)  $\text{KCuF}_3$  experimental data. (b)  $\text{Rb}_2\text{MnF}_4$  experimental data. (c) Cobalt experimental data. (d) Ring-shaped simulated data. (e)  $\text{KCuF}_3$  simulated data. Collected SNR values (arranged in ascending order) by the different approaches for experimental and simulated data. Continuum line indicates the threshold value  $T$ . Dashed lines show each different SNR mean value. Input parameters had been chosen in order to obtain an analogous image definition.

tems, the presence of a nonuniform background or sample imperfections.

In order to present the results concerning the different reconstructions, we will follow the procedure that a user would possibly consider. The first step consists on a three-dimensional exploration, which is usually characterized by scarce information. A qualitative analysis is thus necessary to bound the spatial regions where information is concentrated. Once this is done, a quantitative examination is required by means of inspection of sets of two-dimensional projections. This is first done to assess the statistical data quality, the importance of spurious contributions, etc. The last steps are

usually taken within one-dimensional projections and involve testing the theoretically predicted models versus experiment.

### C. 3D implementation

The ATA algorithm deserves a special attention when implemented in three and higher dimensions. It is the one that, by construction, collects all the input pixels above a given threshold value. Moreover, its behavior and the ensuing results are rather uncoupled from the *a priori* knowledge on a potential user.

Figure 4 shows the qualitative results of the three-dimensional ATA implementation. These reconstructions are accompanied by others performed by MA. A variety of magnetic systems are used as input to show that similar results are obtained in both cases, except that ATA performs this reconstruction without any expertise assistance.

## D. 2D implementation

Results from the two-dimensional implementation are presented in Figs. 5–7. Again, different types of experiments are used to demonstrate the flexibility of the presented algorithm.

Table I provides information concerning the reliability of the reconstructed images in terms of the MSE metric. Here we see that the best reconstruction corresponds to MA whereas ATA yields to a rather distorted image. This is because ATA only controls one parameter that is the position of generators, which determine the size and shape of the respective polygons. These polygons are built by interdependency between generators, and therefore the modification of a single generator affects the construction of the surrounding polygons. Therefore, the method is to be considered as having limited flexibility. In a practical vein, we have to trade off between easiness of use and accuracy. The algorithm MA constitutes a significant improvement over EWB. It is able to adapt to the local details by means of a simple method, although it might not be capable of ensuring that the image has the minimum required precision.

Figure 8 shows the number of pixels collected through the binning process and the SNR value of the bins that enclosed them for simulated and experimental data. For a given image definition, the tendency is repeated for each approach for the different input trials. The proposed ATA collects all the pixels above the threshold, as expected by its very design. A minimum level of precision is thus preserved in the represented image.

On the other hand, the naive EWB algorithm obtains some remarkable SNR-value pixels at the expense of leaving a significant amount of information below the required precision level (see Table I). Low-SNR pixels are indistinguishable from the high-SNR ones; consequently, the bins from which the image is constructed might contain important inaccuracies.

## E. 1D implementation

Although MA does not ensure that all pixels are collected above the requested precision level, it can be conveniently adjusted to represent the peaks with the maximum definition, as can be seen in Fig. 9. In contrast, little knowledge from instrument users is required by ATA to obtain a rather good reconstruction. In addition, it is the only one that has been designed to collect all the input pixels above the corresponding threshold value  $T$ , although the level of detail achieved is lower. Moreover, as opposed to previous attempts, the obtained scheme is the most straightforward to use in comparisons between experimental data and theoretically predicted models due to the absence of levels. Notice that in Fig. 9 the separation between detector

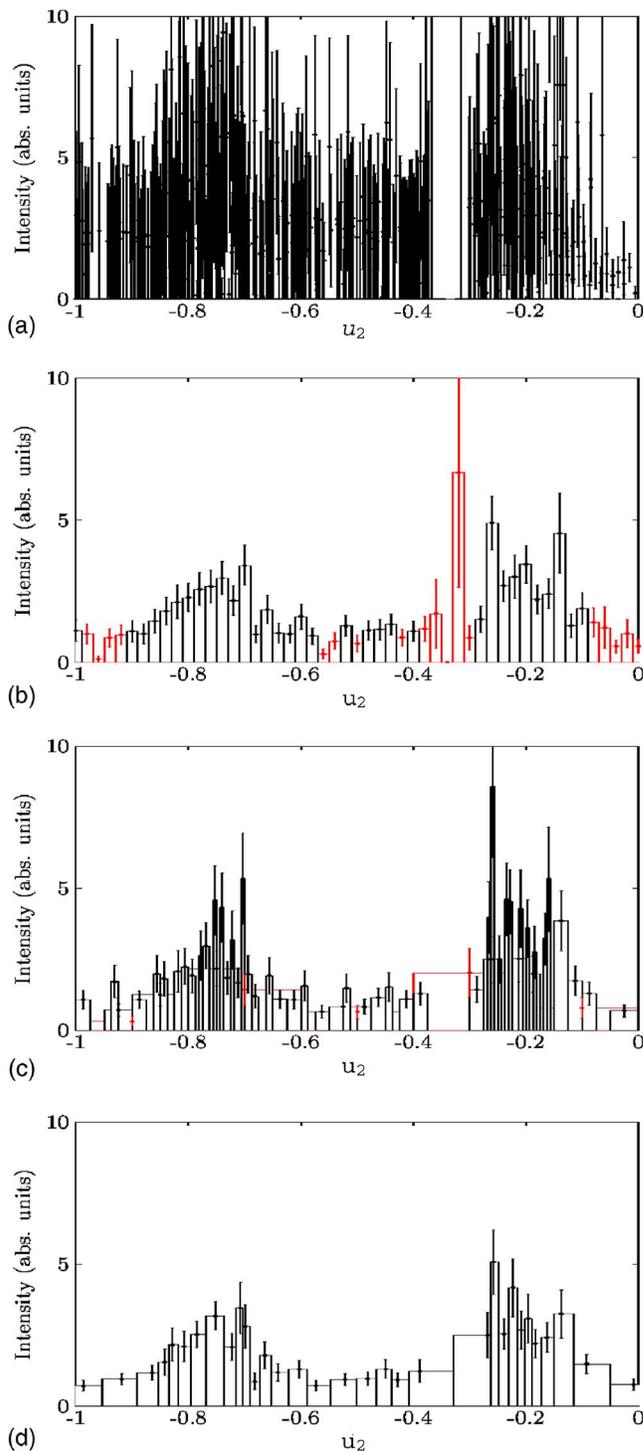


FIG. 9. (Color online) (a) Input data. (b) EWB:  $\Delta u_1 = 0.02$ . (c) MA: maximum bin size is 0.2. (d) ATM. 1D data reconstruction performed by the different approaches over a cut along the  $u_2 = [0.5\mathbf{Q}_h, -\mathbf{Q}_k, 0, 0]$  direction on cobalt sample. Thickness range in  $u_3 = [0, 0, \mathbf{Q}_l, 0]$  goes from 2.9 to 3.1, and in  $u_1 = [\mathbf{Q}_h, 0, 0, 0]$  it goes from  $-0.4$  to  $-0.2$ . The threshold value  $T$  of 2.86 was the same in all algorithms. Red marks indicate the bins below the threshold value. The pattern shows an evolution close to the dispersion for a Heisenberg ferromagnet.

banks—affecting at the area close to  $-0.34$  in  $[0.5\mathbf{Q}_h, -\mathbf{Q}_k, 0, 0]$ —has been conveniently handled by MA and ATA.

## F. Time consumption

Time consumption is a fundamental issue to ensure a rapid visual feedback. The formulation of queries by direct

TABLE II. Time consumption (seconds) for the different approaches, using random nonzero size data. ATA has been also implemented in FORTRAN, using a 1.5 GHz IBM G5 (XL FORTRAN compiler).

Implementation		Random data						
		1000	10 000	20 000	30 000	40 000	50 000	100 000
MATLAB	MA	0.14	0.22	0.36	0.97	2.70	5.62	14.97
	ATA	2.50	110.22	472.73	895.91	1 606.00	2 825.00	11 221.00
XL	ATA	0.04	2.89	14.13	30.96	63.16	90.03	369.59

manipulation and the immediate display of the results has many advantages for both new and expert users. This enables the experimentalist to gain a deeper understanding of the data set, providing information on the nature of the problem that cannot be grasped using a static representation.

As we can see in Table II, the best performance in terms of time consumption is MA. While this algorithm constructs grids along the different levels in a deterministic way, ATA requires additional iteration processes, resulting in a reduction of the time efficiency. The time complexity of MA is  $O(n^3)$ , and  $O(n^2)$  for ATA.

Taking into account the importance of time consumption, we implemented ATA that offers a better balance for a first-time exploration process in FORTRAN. Such a lower-level programming language comes to be much more efficient than MATLAB. We tested it on a IBM G5 with IBM XLF compiler. Time consumption was dramatically reduced if compared to the same algorithm using MATLAB, and the results are shown in Table II. Although MA still gives the best results, it is clear that a better performance can be obtained by means of code parallelization or hardware customization.

## V. DISCUSSION

The algorithm herein described constitutes a significant improvement over traditional equal-width binning procedures as well as over the multiresolution algorithm previously reported on Ref. 1 It provides simultaneous displays of both intensity data up to a given level of precision, and their statistical significance.

The *adaptive tessellation algorithm* (ATA) uses pixelated data above a given threshold to provide a full tessellation of the available data space. That is, it provides image reconstructions without leaving any empty patches and avoiding overlapping regions. Furthermore, because of its grounding on the use of *centroidal Voronoi Tessellations* (CVTs), the algorithm becomes highly portable since the entire image can be stored into two arrays containing the intensities and the polygon generators, easy to transport into most visualization schemes.

A drawback of the present implementation constitutes the significant amount of CPU time involved in the computation presented here, specially compared to MA—although MA algorithm does not ensure the reconstructed image complies with the accuracy requirements. As a corollary of general validity, the data presented here illustrate the need of

extending grid-computing procedures in order to provide the end user with near real-time access to the results.

Finally, and on a broader scope, the work reported here provides a contribution of interest to the general field of scientific visualization where, as mentioned not long ago<sup>27,28</sup> not only do the data need to be visualized but also their associated uncertainties must be displayed.

The MATLAB and FORTRAN implementations of the algorithm described in this paper is available from Ref. 29.

## ACKNOWLEDGMENT

One of the authors (I.B.) was supported in part by Grant No. MAT2002-04540-C05-03 (Spain).

- <sup>1</sup>I. Bustinduy, F. J. Bermejo, T. G. Perring, and G. Bordel, Nucl. Instrum. Methods Phys. Res. A **546**, 498 (2005).
- <sup>2</sup>W. K. Leow and R. Li, Comput. Vis. Image Underst. **94**, 67 (2004).
- <sup>3</sup>M. J. Bayley and P. Willett, J. Mol. Graphics Modell. **17**, 10 (1999).
- <sup>4</sup>M. Roederer, A. Treister, W. Moore, and L. A. Herzenberg, Cytometry **45**, 37 (2001).
- <sup>5</sup>M. Roederer, W. Moore, A. Treister, R. R. Hardy, and L. A. Herzenberg, Cytometry **45**, 47 (2001b).
- <sup>6</sup>M. Roederer and R. R. Hardy, Cytometry **45**, 56 (2001).
- <sup>7</sup>P. Rubin *et al.* (CLEO Collaboration), Phys. Rev. Lett. **93**, 111801 (2004).
- <sup>8</sup>B. Aubert *et al.*, Phys. Rev. D **72**, 052008 (2005).
- <sup>9</sup>J. S. Sanders and A. C. Fabian, Mon. Not. R. Astron. Soc. **325**, 178 (2001).
- <sup>10</sup>M. Cappellari and Y. Copin, Mon. Not. R. Astron. Soc. **342**, 345 (2003).
- <sup>11</sup>S. Diehl and T. S. Statler, Mon. Not. R. Astron. Soc. **368**, 497 (2006).
- <sup>12</sup>H. Samet, ACM Comput. Surv. **16**, 187 (1984).
- <sup>13</sup>Q. Du, M. Faber, and M. Gunzburger, SIAM Rev. **41**, 637 (1999).
- <sup>14</sup>S. Valette and J. M. Chassery, Comput. Graph. Forum **23**, 381 (2004).
- <sup>15</sup>S. Hiller, H. Hellwig, and O. Deussen, Comput. Graph. Forum **22**, 515 (2003).
- <sup>16</sup>C. G. Wager, B. A. Coull, and N. Lange, J. R. Stat. Soc. Ser. B (Stat. Methodol.) **66**, 429 (2004).
- <sup>17</sup>J. Cortes, S. Martinez, T. Karatas, and F. Bullo, IEEE Trans. Rob. Autom. **20**, 243 (2004).
- <sup>18</sup>Q. Du and M. Gunzburger, Appl. Math. Comput. **133**, 591 (2002).
- <sup>19</sup>Q. Du and D. Wang, Int. J. Numer. Methods Eng. **56**, 1355 (2003).
- <sup>20</sup>A. B. Mendes and I. H. Themido, Int. Trans. Oper. Res. **11**, 1 (2004).
- <sup>21</sup>S. P. Lloyd, IEEE Trans. Inf. Theory **28**, 129 (1982).
- <sup>22</sup>A. Gersho, IEEE Trans. Inf. Theory **25**, 373 (1979).
- <sup>23</sup>Y. Linde, A. Buzo, and R. M. Gray, IEEE Trans. Commun. **28**, 84 (1980).
- <sup>24</sup>B. Lake, D. A. Tennant, C. D. Frost, and S. D. Nagler, Nat. Mater. **4**, 329 (2005).
- <sup>25</sup>T. Huberman, R. Coldea, and R. A. Cowley, Phys. Rev. B **72**, 014413 (2005).
- <sup>26</sup>T. G. Perring, ISIS Experimental Report, 2001 (unpublished).
- <sup>27</sup>C. R. Johnson, IEEE Comput. Graphics Appl. **24**, 13 (2004).
- <sup>28</sup>C. R. Johnson and A. R. Sanderson, IEEE Comput. Graphics Appl. **23**, 6 (2003).
- <sup>29</sup><http://www.isis.rl.ac.uk/excitations/software.htm>.