Mem. S.A.It. Vol. 89, 342

Memorie della

Serendip IV - SETI post processing pipeline

J. Monari and S. Montebugnoli

INAF- IRA Istituto di Radioastronomia, PO BOX 14, Via Fiorentina 3513, 40058 Villafontana (BO), Italy, e-mail: j.monari@ira.inaf.it

Abstract. New technologies under development aim to the realization of more sensitive and performing instruments. Designs of future radio telescope like SKA and back ends, will allow us to deepen our radio-signals analysis, increasing the probability to detect a signal emitted by Extra Terrestrial Intelligence (ETI). This means a huge amount of data to analyze and compute. The processing will be possible only in automatic mode, exploiting the computing power of future computers and a proper algorithm to perform the detection of "the type of signal we think it will be possible to receive".
It is reported the experience gained at the Medicina radio telescope by using the Italian Serendip-IV (SIV) module and the development of all the software needed to perform off-line, all the post processing operations necessary to detect the possible ETI candidate signal. An interesting algorithm based on Hough transform, has been used for data post processing, a collection of ad hoc procedures written for ET signals detection from the experience managed by our Institute after the installation of the multi-channel spectrum analyzer Serendip IV, is described inside this article.

Key words. Serendip IV – SETI – Medicina

1. Introduction

The Italian Medicina Radio astronomy Station is located in the Northern part of Italy, nearby Bologna, and is composed by two radio telescopes: the 32 mt (VLBI) dish and the Northern Cross, a large T-shaped parabolic/cylindrical antenna (30.000 sqm). The Medicina dish presents about 900 sqm of collecting area. It can work in the radio astronomical bands included in the 1.4 - 23 GHz range.

The Italian Serendip-IV (SIV) module, developed at the Berkeley University and installed in 1998, was a 24 million Channels (24 MCHs) at 15MHz input bandwidth (0.6 Hz of resolution) connected to the 32 mt dish (Montebugnoli 1996). The main characteristic of this back-end was to operate in piggy back mode (namely in parallel to the ongoing activities), so that SETI observations could be performed for 24 hours a day all the year long, at an extremely low cost.

SIV was the back-end used by the SETI-Italia program in the 1998-2008 period and, unfortunately due to the long time of operation, was dismissed and now no operational. Each day it produced a huge amount of data, to be post processed off line, in order to detect the possible ETI signals stored inside a proper archive. On the average, it generated 50 MB - 100 MB per day, depending on the interference levels.

The large quantity of data produced by SIV, was processed off-line by SALVE 2 (Software

Fig. 1. Italian Radio telescopes at Medicina station. Left: Northern Cross antenna Right: 32m Parabola dish.

Fig. 2. Old pictures showing the preliminary test on SIV and the group in charge to follow the program (from the left Jader Monari, Alessandro Orfei, Stelio Montebugnoli and Dan Werthimer)

Aimed at off Line Verification ETI signals, version 2.0)(Monari 1999).

This paper gives evidence of the different procedures created by SALVE 2 in order to increase both the reliability of the pattern recognition algorithms and the performance of the programs devoted to the data displaying. The experience gained in this field could be a good starting point for the development of future post processing pipeline data system, for SETI back-ends, if installed at the Noto (Sicily) and/or SRT (Sardinia) radio telescopes facilities.

2. SERENDIP

SERENDIP IV (Cobb 2000), (Werthimer 1995), (Cobb 1997) is composed by a bank of digital mixers and filters to separate the input band into sixty 2.4 MHz sub-bands. Each 2.4

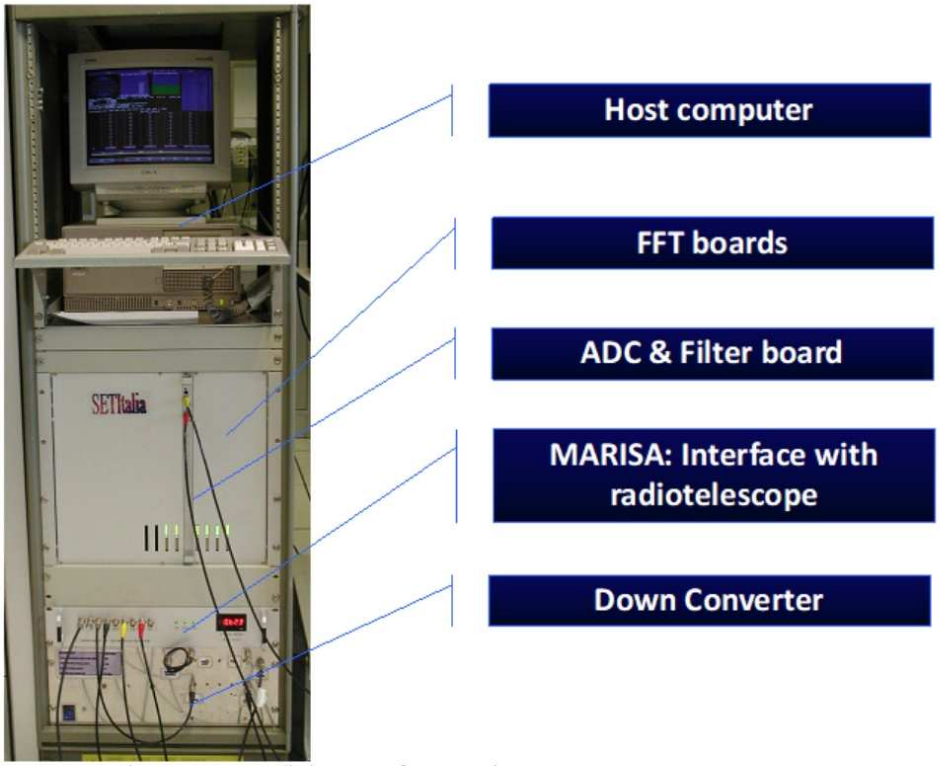

Fig. 3. Serendip IV system installed on 1998, for SETI Italia program.

MHz sub-band was further broken down into 0.6 Hz bins by means of a 4 million points Fast Fourier Transform board. The so calculated power spectra enter the post-processing portion of each board where baseline normalization, coarse resolution spectra computation and event thresholding occurred. Baseline normalization was achieved with a sliding eight thousand channel local mean boxcar. Thresholding was performed based on mean spectral power and signals that exceed the threshold were logged along with time, pointing coordinates, frequency, and power.

The SERENDIP IV observing cycle consisted of real-time data analysis by the SERENDIP IV instrument, near real-time data transfer across the Internet, off-line data reduction and candidate generation.

This process had to be as automatic as possible in filtering out any radio interference and detecting particular ETI paths stored in a dedicated data base. An operator was able to retrieve the stored data in order to accurately verify the alarm hits and the output in general.

The problem of automatic data elaboration and signal recognition is crucial, and it will be felt even more in the future when a new instrument as the newest Serendip generation will be ready. The heart of the whole process is the algorithm for ETI signal recognition. In fact, to do that, longer before decoding possible modulated signals, it is necessary to isolate them from thousands of human interferences and natural emissions originated on our planet or elsewhere in the universe. It is generally thought that a possible ETI emission could be a monochromatic signal, dopplered by the relative motion of the planet with respect to the Earth. As already mentioned, the big amount of data produced by SIV was processed off-line

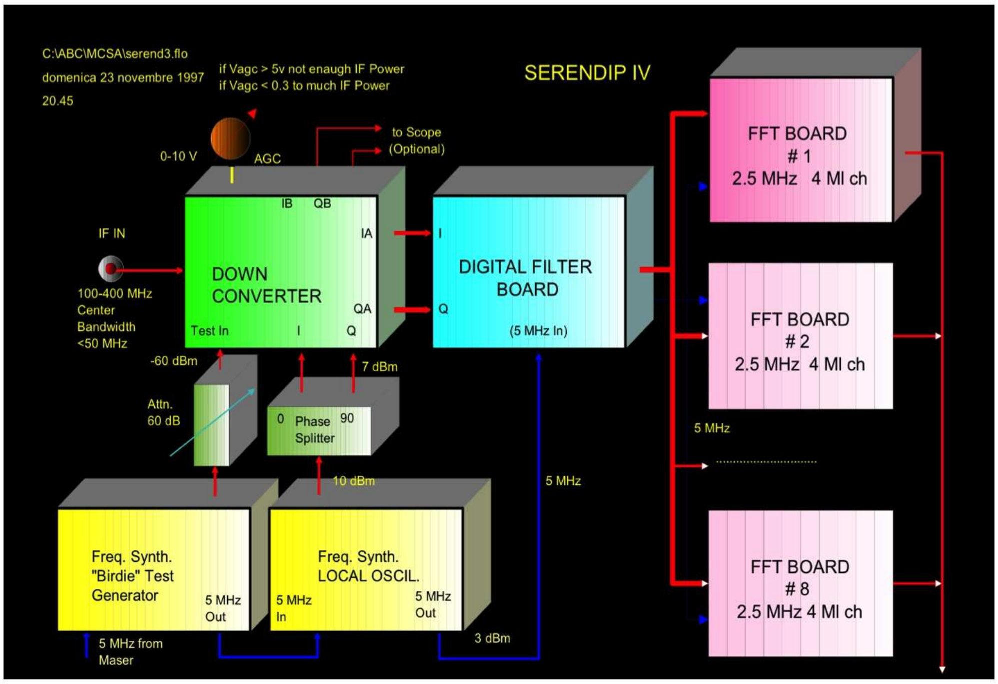

Fig. 4. Narrow band signals having power significantly above background noise levels were recorded, along with telescope coordinates, time, and frequency.

by SALVE 2 software package. In next chapters a collection of algorithms for SIV data processing and system scalability are described.

3. JTFM Matrix

It is very important to know the data set structure of the Join Time Frequency Matrix (JTFM) stored by SIV, in order to understand the way to recognize and detect a possible ETI signal. We could consider the stored data as a matrix where each column represents a frequency channel while the rows correspond to the time. Each entire row represents the bandwidth of SIV (15MHz for the Italian instrument). Each channel is "ON" only when a signal hits a particular user-defined threshold. The scheme is reported in Error: Reference source not found.

The matrix dimensions can be calculated taking in account the following considerations:

- the actual number of FFT boards installed inside the SIV is 6 and each one has 4106 channels;
- each board can store the first 100 points that hit the threshold;
- every SIV file it is approximately 1 MB corresponding almost to 200 computed FFT

Notice that the dimension of the matrix is $M \in M(200, 24\text{x}10^6)$, further we could consider it as a sparse matrix because the number of channels "ON" is really less than $24\text{x}10^6$. We intend as a good ETI candidate every Doppler shifted monochromatic signal which turns out not to be a false alarm due to a radio interference (RFI), rejected by dedicated filters software described later. From the literature such a type of signal could be considered one of several possible forms of an intentional alien transmission. In fact, we would like to point out that, thanks to the open structure of the SALVE 2 program, new pattern-recognition

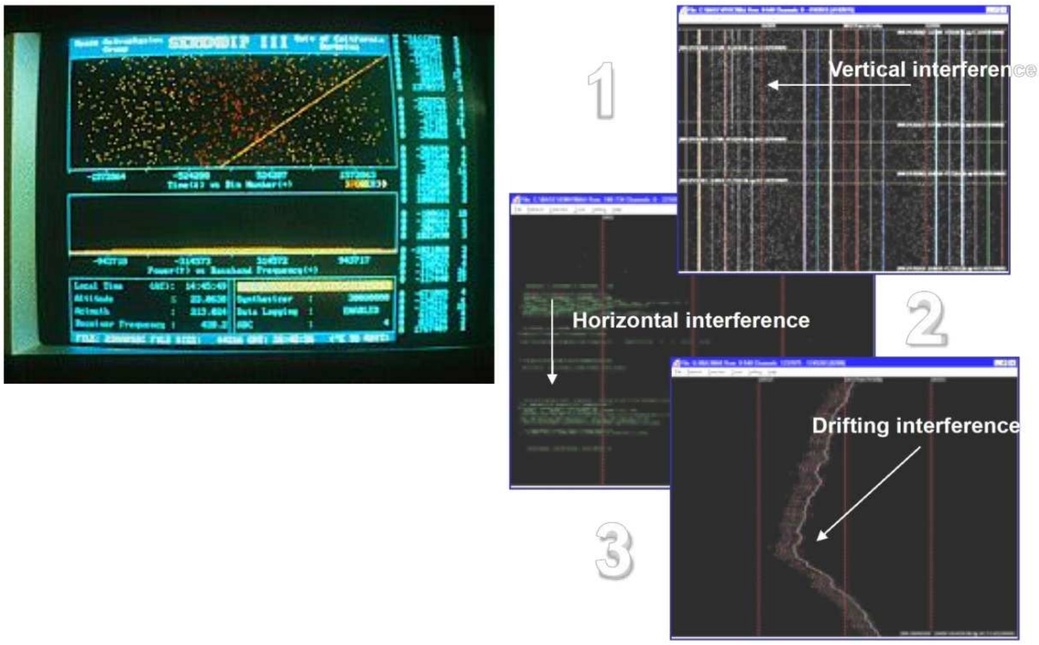

Fig. 5. Screenshot of Serendip III acquisition and some example of data contaminated by RFI.

algorithm, suitable for different types of signals detection, can be written and the results merged at the end of the parallel flows.

4. SALVE 2 programs

The structure of the post processing software (Monari 2001) is shown in Figure 7. A scheduler block controls and manages the entire process. ACQUIRE is the first application that takes care of the data transfer from the PC Host to the base directory inside the local PC SERENDIP PROC1 disc. At the end of this phase, always under the control of the scheduler, the next modules start for the interferences filtering and reduction via a particular dedicated software packages (VFILTER, DFILTER, INTERFO). To get used to the local interference, we developed a data display called SCAN 7. The study of the local RFIs allowed us to develop the first version of algorithms for their handling and rejection.

The MERGE file adds the possible candidates from the different pattern recognition algorithms (HOUGH, CORRELATION etc.) at the output level of the parallel flows already described.

– All the alarms are stored day by day in the folder ALARM; later an operator takes care of the analysis of the results.
– To save space on disc, the SQUEEZE program ”Zips” the binary data reducing the compressed file up to 60% with respect to the original one. It compresses all the SIV files.

In order to test the sensitivity of the whole Serendip IV system, a check has been performed by observing the Global Mars Surveyor orbiting around Mars, exploiting the on-board small power transmitter. The results of the test are reported in Figure 8. The signal affected by the Doppler shift (due to the Earth- satellite radial velocity component variation) is clearly visible (X axis → frequency, Y axis → time).

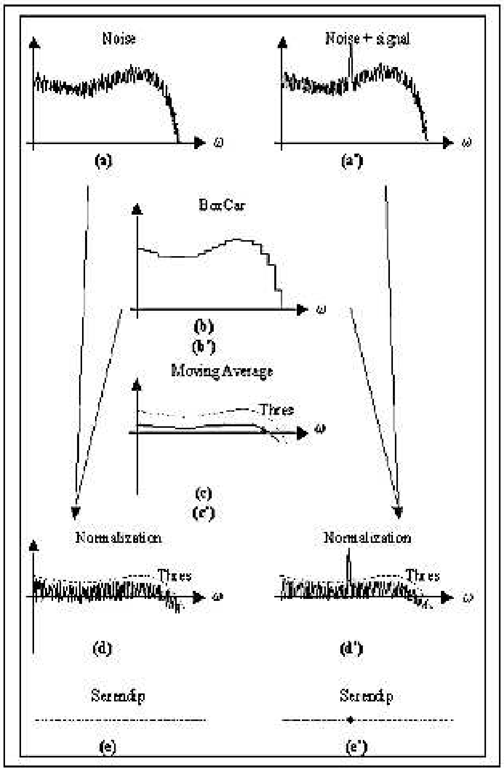

Fig. 6. Serendip on-line post-processing of the FFT data.

5. From Radon to Generalized Hough Transform

The Simple Hough Transform is a method used in literature to search for regular patterns (Cobb 2000), (Werthimer 1995) like lines in a two-dimensional set of data. For instance, every straight line in the Cartesian space (X,Y) is transformed by the Simple Hough transform in a peak (M,Q) in the bi-dimensional parameters space (Figure 9). Thus, the difficult problem of looking for the straight line patterns in a plane is transformed by the Hough transform into the simpler problem of finding the local maximum in the two-dimensional space of parameters (M,Q). By accumulating the contributions of every couple of points belonging to the coherent by virtue of the Hough transform, it is possible to look for the local maximum in the Hough space (M,Q) or in the equivalent (ρ,θ) polar space (Radon transform). In mathematical terms, a straight line can be represented in the polar domain (ρ,θ) by means of the polar equation:

$$\rho = x\cos\theta + y\sin\theta \tag{1}$$

In this case, the Radon transform is:

$$R(\rho,\theta) = \int_{-\infty}^{+\infty}\int_{-\infty}^{+\infty} g(x,y)\theta(\rho - x\cos\theta - y\sin\theta)dxdy \tag{2}$$

where g(x,y) is the general function to transform, (x,y) is the Dirac function, is the distance of the straight line from the origin of the axes and is the normal angle. This representation allows us to describe all the straight lines in the polar space within the range of $0 \leq \theta < 2\pi$ and $\rho \geq 0$, as in Figure 9.

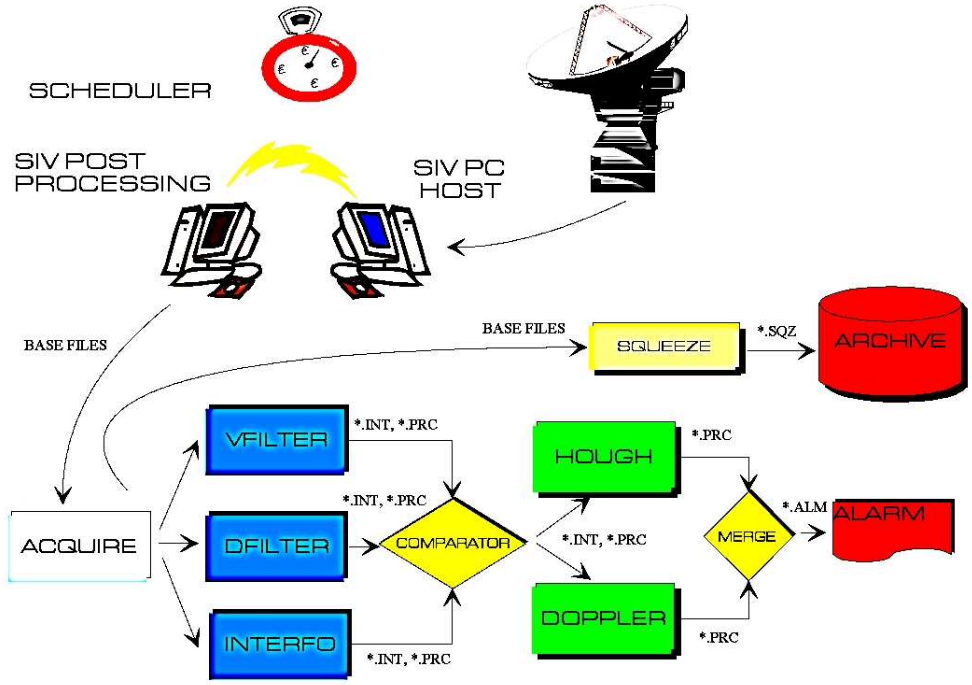

Fig. 7. Scheme of SALVE2 post processing pipeline for ETI detection.

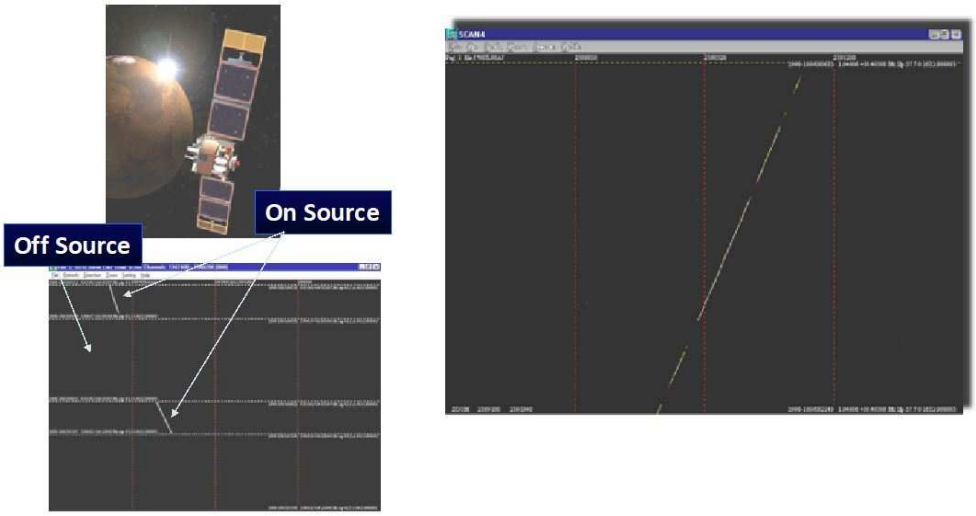

Fig. 8. Detection of the Mars Global Surveyor, used as a weak test signal.

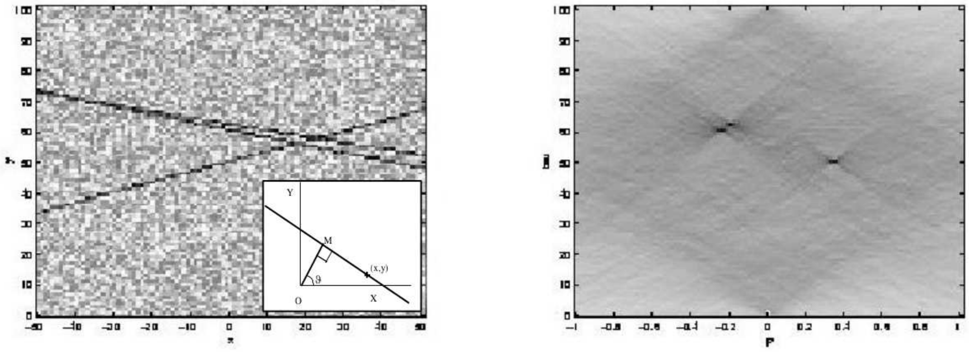

Fig. 9. Radon transform for straight line, left original Cartesian space and right picture Radon polar space

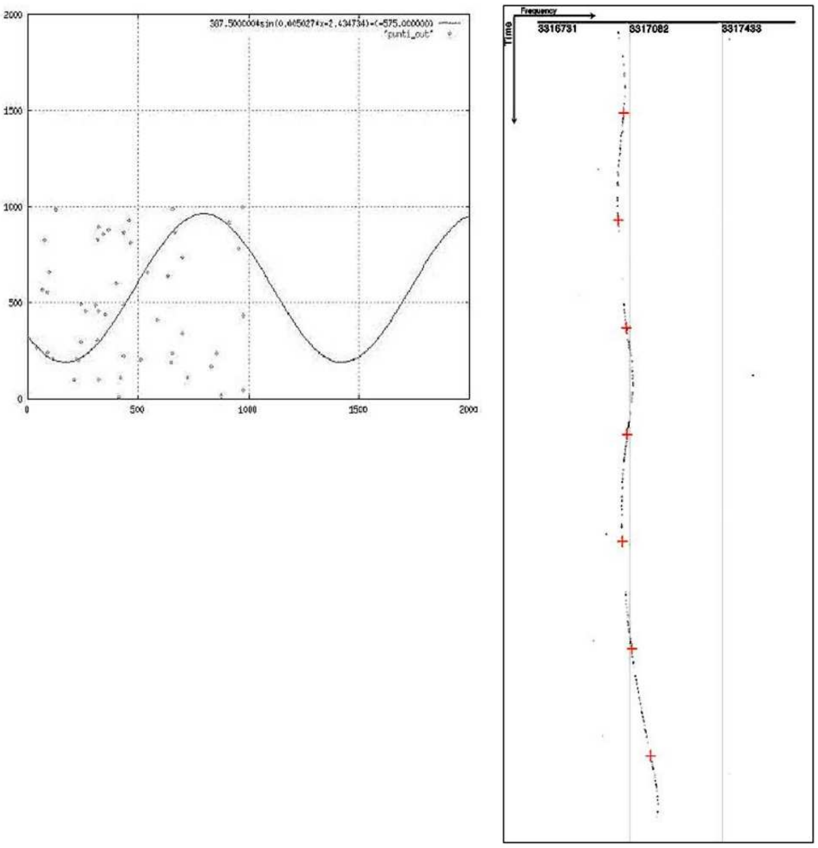

Fig. 10. Left: only one good candidate has been found by GHT in a sparse matrix, it perfectly matches with the simulated sinusoid set by the user. Right: Italian WOW candidate 1420MHz, never confirmed. It is interesting the sinusoidal behavior.

Table 1. Main parameters for Hough Transform

Parameter	Range
A amplitude	[0, 1000]
P angular velocity	$2\pi/\tau$, where τ=[35min, 14.7h] $\rightarrow$ [1250, 31250]
F phase	[0, π]
q offset	[-1000, 1000]

The Radon Transform is the simplest algorithm implemented for the actual post-processing in SALVE2 programs; the initial Cartesian space (x,y) or "space of image" is the JTFM, then each point is transformed in the corresponding Radon Polar (RP) space R(ρ,θ). In practice a real dopplered signal can be considered a sinusoidal line in the Cartesian space (Figure 8) a straight line can be approximated only for short time periods. In fact, supposing the JTFM to be compensated with respect to its own reference frame, the hypothetical alien transmitter, in comparison to the system geometric earthling, would only show the motions of rotation and revolution around its star. The second version of the algorithm (Generalized Hough Transform), that is a generalization of the previous one, differs from it for the simple but expensive fact which, instead of seeking straight lines in the space image, sinusoids or sinusoidal arcs are sought. In general, that can be of the type:

$$f(x) = A\sin(Px - F) - q \tag{3}$$

where (A, P, F, q) are the main vectors of Generalized Hough (GH) space and correspond to:

- A: amplitude of sinusoid
- P: angular velocity $2\pi/\tau$ where τ is the period corresponding to the Doppler shift of the signal.
- F: Initial Phase
- q: Axis Offset of sinusoid

As it is easy to understand, the newest GH space has four dimensions, increasing the computational complexity with respect to the standard Radon algorithm. To physically avoid impossible solutions and therefore to decrease both the computational time and the required memory, a subset of the quadric-dimensional dominion is chosen (the express integer in the here following table has been considered as the number of columns or lines of the JTFM):

The transformation could be considered as:

$$H(A, P, F, q) = \int_{-\infty}^{+\infty} g(x, A\sin(Px - F) - q)dx = \tag{4}$$

$$\int_{-\infty}^{+\infty}\int_{-\infty}^{+\infty} g(x, y)\rho(y - A\sin(Px - F) - q)dxdy \tag{5}$$

Such function is not null when the (x,y) fits the function $y - A\sin(Px - F) - q$ or when, once fixed the parameters (A,P,F,q), the couple (x,y) is a solution of the equation:

$$y - A\sin(Px - F) - q = 0 \tag{6}$$

that represents a sinusoid in the image space while, in the GH space, is a hypersurface of dimension four. The intersection, for instance, of two of these hypersurfaces behaves as a double value in the points of intersection, and this is a proof of a good candidate.

At this point it is possible to analyze the GH space and, with appropriate filters and algorithms, to find out the global maximum. To improve the search of the maximum and to avoid that the discretization of the space image could introduce false alarms, a clusterization K-MEANS is used. Defining at least a radius of tolerance of the hyperspheres of the transformed points, all the transformed points of the space image whose hyperspheres intersect it can be considered as a single cluster. To better define the parameters (A, P, F, q) of the cluster, an interactive process can define a common barycenter and associate an average vote to it.

6. Conclusions

SIV has been the back-end used SETI-Italia program during the decade following the 1998 year and, due to troubles of technology obsolescence faults, it was switched off in 2008 BTW the algorithm developed to post process the data and described in this paper can be very useful and attractive also for future SETI back-ends. In fact, also the newest generation of data processing for the modern ATA telescopes called SonATA (SETI on ATA) uses the same ideas to find a Dopplered drift narrow signal in matrix that are the representation of the power as a function of frequency and time are called spectrograms. The used mathematical tool is the four-dimensional Generalized Hough transform with the due adaptations to more powerful computers, in a near future, it could be used for the most advanced SETI searches.

References

Cobb, J., et al. 1997, The Serendip IV Intereference Rejection and signal detection system, in Astronomical and Biochemical origins and the search for life in the universe, C. B. Cosmovici, S. Bowyer and D. Werthimer eds. (Editrice Compositori, Bologna), IAU Colloquium, 161 , 677

Cobb, J., et al. 2000, SERENDIP IV: Data Acquisition, Reduction, and Analysis, in Bioastronomy '99: A New Era in Bioastronomy, ed. G. Lemerchand, K. Meech (ASP, San Francisco), ASP Conf. Ser., 213, 485

Monari, J., et al. 1999, SALVE (Software Aimed at off Line Verification Eti), Proceeding of Bioastronomy Conference in Kona Hawaii August 1999

Monari, J. et al. 2001, SALVE 2 (Software Aimed at off Line Verification ETI signals ver 2.0, International Astronautical Proceeding of 52ns IAF Toulouse, France ref. IAA-01-IAA.9.1.10

Montebugnoli, S. et al. 1996, The SETI facilities at the Medicina/Noto sites, Proceeding of IAF96 Congress in Beijing ref. IAA-96-IAA.9.1.11

Werthimer, D., et al. 1995, The Berkeley SETI program: Serendip III and IV instrumentation, in Progress in the Search for Extraterrestrial Life: 1993 Bioastronomy Symposium, G. Seth Shostak ed. (ASP, San Francisco), ASP Conf. Ser., 74, 293