A real-time KLT implementation for radio-SETI applications

Andrea Melis[a], Raimondo Concu[a], Pierpaolo Pari[c], Claudio Maccone[b], Stelio Montebugnoli[c], Andrea Possenti[a], Giuseppe Valente[a], Nicoló Antonietti[1], Delphine Perrodin[a], Carlo Migoni[a], Matteo Murgia[a], Alessio Trois[a], Massimo Barbaro[e], Alessandro Bocchinu[a], Silvia Casu[a], Maria Ilaria Lunesu[e], Jader Monari[c], Alessandro Navarrini[a], Tonino Pisanu[a], Francesco Schilliró[d], and Valentina Vacca[a]

[a]Osservatorio Astronomico di Cagliari, Via della Scienza 5, Selargius, Italy
[b]IASF Milano, Via Alfonso Corti 12, Milano, Italy
[c]IRA Bologna, Via Gobetti 12, Bologna, Italy
[d]Osservatorio Astrofisico di Catania, Via S. Sofia 78, Catania, Italy
[e]DIEE - University of Cagliari, Via Marengo 2, Cagliari, Italy
[1]Not affiliated to any institution

ABSTRACT

SETI, the Search for ExtraTerrestrial Intelligence, is the search for radio signals emitted by alien civilizations living in the Galaxy. Narrow-band FFT-based approaches have been preferred in SETI, since their computation time only grows like N*lnN, where N is the number of time samples. On the contrary, a wide-band approach based on the Kahrunen-Loève Transform (KLT) algorithm would be preferable, but it would scale like N*N. In this paper, we describe a hardware-software infrastructure based on FPGA boards and GPU-based PCs that circumvents this computation-time problem allowing for a real-time KLT.

Keywords: KLT, Sardinia Radio Telescope, FPGA, GPU

1. INTRODUCTION

The SETI project stands for "Search for Extra-Terrestrial Intelligence", i. e. the search for evidence of intelligence coming from a possible extraterrestrial civilization. As a consequence, the project is not concerned with finding evidence for the existence of animals or, more generally, life forms less advanced than humans. That also includes hypothetical alien creatures that have not reached a technological level similar to ours. Moreover, it is important to point out that the only goal of the SETI project is to receive and recognize the artificiality of the acquired signals; the issue of understanding the content of the message, as well as what type of response be sent back, could be handled by others (probably politicians). On the other hand, we need to think about the extraordinary variety of different languages on our planet: how can we "extract" the meaning of a message and what kind of protocol would be suitable in case a response would have to be sent back? Finally, the enormous distance between habitable planets suggests that the message would probably have been sent many years ago and, in the meantime, the technological progress of those civilizations would have reached a higher level; we can see how, for example, our world rapidly changed after the Second World War.

Up to now, no evidence of extraterrestrial life has yet been found, however it is statistically almost impossible that other life forms are not be present in the Universe. There are a hundred billions stars in the Milky Way, as well as hundreds of billions of galaxies - similar to the Milky Way - in the Universe: the idea that we are a unique life form is, statistically speaking, complete nonsense. In addition, we may have "listened" to the wrong frequency, with the wrong detection algorithm, in the wrong direction, at the wrong time and so forth. Therefore, SETI is a wide research field that humans are just now starting to explore.

Further author information: (Send correspondence to Andrea Melis)
Andrea Melis: E-mail: amelis@oa-cagliari.inaf.it, Telephone: +39 07071180220

Millimeter, Submillimeter, and Far-Infrared Detectors and Instrumentation for Astronomy VIII,
edited by Wayne S. Holland, Jonas Zmuidzinas, Proc. of SPIE Vol. 9914, 99143D
 · doi: 10.1117/12.2232347

2. MATHEMATICAL DESCRIPTION OF THE KLT

2.1 Introduction to the KLT

This chapter is a simple introduction regarding the use of the Karhunen-Loève Transform (KLT) to extract weak signals from any kind of noise. In general, the noise may be colored and be present over wide bandwidths, not just white and over narrow bandwidths. We show that the signal extraction can be achieved with the KLT more accurately than with the fast Fourier transform (FFT), especially if the signals buried into the noise are very weak, in which case the FFT fails. This superior performance of the KLT happens because the KLT of any stochastic process (both stationary and non-stationary) is defined from the start over a finite time span ranging between 0 and a final and finite instant T (contrary to the FFT, which is defined over an infinite time span). We have thus put, on a strong mathematical foundation, a set of important practical formulae that can be applied to improve SETI(([1]) and ([2])), the detection of exoplanets, asteroidal radars, as well as other fields of knowledge like economics, genetics, biomedicals, etc. to which the KLT can be equally applied with success. We believe that these improvements in the mathematical ways of handling the KLT will increase the interest of scientists in this algorithm, which may well replace the Fourier transform in the near future.

2.2 A bit of history

We argue that the Karhunen-Loève Transform (KLT) is the most advanced mathematical algorithm that is available in the year 2016 to achieve both noise filtering and data compression in processing signals of any kind. It took about two centuries (1800-2000) for mathematicians to create such a jewel, piece after piece, paper after paper. It is thus difficult to evaluate who did what in building up the KLT, and to be fair to each contributing author. In addition, both pure and applied mathematicians often speak in jargon so that even accomplished scientists sometimes find it hard to understand them. This unfortunate situation hides the aesthetic beauty of many mathematical discoveries, which were often historically made by their authors more for the joy of opening new lines of thought than for the sake of any immediate application to science and engineering. In essence, the KLT is a rather new mathematical tool to improve our understanding of physical phenomena, and is far superior to the classical Fourier Transform (FT). The KLT is named for two mathematicians, the Finnish actuary, Kari Karhunen (1915-1992) ([3]) and the French-American mathematician, Michel Loève (1907-1979) (([4]) and ([5])), who proved, independently and at about the same time (1946), that the series (2) hereafter is convergent. Put this way, the KLT looks like a purely mathematical topic, but in fact this is hardly the case. As early as 1933, the American statistician and economist Harold Hotelling (1895-1973) had used the KLT (for discrete time, rather than for continuous time), so the KLT is sometimes called the Hotelling Transform. Even much earlier than these three authors, as early as 1873, the Italian Eugenio Beltrami (1835-1899) had discovered the SVD (Singular Value Decomposition), which is closely related to the KLT in that area of applied mathematics nowadays called Principal Components Analysis (PCA). Unfortunately, a complete historical account about how these contributions developed since 1865 (when the English mathematician Arthur Cayley (1821-1895) invented matrices) simply does not exist. We only know about "fragments of thought" that give us an overall vision of both the PCA and the KLT. In the following sections, we will derive, heuristically and step-by-step, the many equations that make up the KLT. We think that this approach is much easier to understand for beginners than what is found in most pure mathematical textbooks, and hope that the readers will appreciate our efforts to explain the KLT as easily as possible to the non-mathematically trained people.

2.3 A heuristic derivation of the KLT

The KLT was born during the World War Two years out of the need to merge two different areas of classical mathematics: 1) The expansion of a deterministic periodic signal $x(t)$ onto a basis of orthonormal functions (sines and cosines, in this case), representeded by the classical Fourier series (first put forward by the French mathematician Jean Baptiste Joseph Fourier (1768-1830) around 1807):

$$x(t) = \frac{a_0}{2} + \sum_{n=1}^{\infty}\Big[a_n \cos(\omega_n t) + b_n \sin(\omega_n t)\Big] \tag{1}$$

2) The need to extend this too narrow and deterministic view to probability and statistics. The much larger variety of phenomena called noise by physicists and engineers will thus be encompassed by the new transform. This enlarged view considers a random function $X(t)$ (note that we denote random quantities by capitals, and that $X(t)$ is also called a "stochastic process of the time"). We now seek to expand this stochastic process onto a set of orthonormal functions ϕ_n according to the starting formula:

$$X(t) = \sum_{n=1}^{\infty} Z_n \phi_n(t) \tag{2}$$

which is called the Karhunen-Loève (KL) expansion over the finite time interval $0 \leq t \leq T$. What are the Z_n and the $\phi_n(t)$ in (2)? To find out, let us start by recalling what "orthonormality" means for Fourier series (1). Leonhard Euler (1707-1783) had already laid the first stone towards the Fourier series (1) by proving that, if $T = t_2 - t_1$ is the assumed period of $x(t)$ and one sets $\omega_n = n\frac{2\pi}{T}$, then the coefficients a_n and b_n in (1) are obtained from the known function (or "signal") $x(t)$ by virtue of the equations ("Euler formulae"):

$$a_n = \frac{2}{T}\int_{t_1}^{t_2} x(t)\cos(\omega_n t)dt \qquad b_n = \frac{2}{T}\int_{t_1}^{t_2} x(t)\sin(\omega_n t)dt \tag{3}$$

If the same result is going to be true for the Karhunen-Love expansion, the functions of the time, $\phi_n(t)$ in (2) must be orthornormal, i.e. both orthogonal and normalized to one, that is:

$$\int_0^T \phi_m(t)\phi_n(t)dt = \delta_{mn} \tag{4}$$

where $\delta_{mn} = 0$ for $m \neq n$ and $\delta_{mn} = 1$ for $m = n$. But what are the Z_n appearing in (2)? A random function $X(t)$ can be thought of as something made of two parts: its behavior in time, represented by the functions $\phi_n(t)$, and its behavior with respect to probability and statistics, which must therefore be represented by the Z_n. In other words, the Z_n must be random variables not changing in time, i.e. just random variables that are not stochastic processes. By doing so we have actually made one basic, new step ahead: we have found that the KLT separates the probabilistic behavior of the random function $X(t)$ from it behavior in time, a kind of untypical separation that is achieved nowhere else in mathematics! Having discovered that the Z_n are random variables, a few trivial consequences follow at once. Let us denote by $E\{\}$ the linear operator yielding the average of a random variable or stochastic process. If one takes the average of both sides of the KL expansion (2), one then gets here (we freely interchange the average operator $E\{\}$ with the infinite summation sign, bypassing the complaints of subtle mathematicians!):

$$E\{X(t)\} = \sum_{n=1}^{\infty} E\{Z_n\}\phi_n(t) \tag{5}$$

Now, it is not restrictive to suppose that the random function $X(t)$ has a zero mean value in time, namely that the following equation is identically true for all values of the time t within the interval $0 \leq t \leq T$:

$$E\{X(t)\} \equiv 0 \tag{6}$$

In fact, were this not true, one could replace $X(t)$ by the new random function $X(t) - E\{X(t)\}$ in all of the above calculations, thus reverting to the case of a new random function with zero mean value. Thus, the random variables Z_n must also have a zero mean value:

$$E\{Z_n\} \equiv 0 \tag{7}$$

This equation has a simple consequence: since the variance σ^2_{Zn} of the random variables Z_n is given by:

$$\sigma_{Zn}^2 = E\{Z_n^2\} - E^2\{Z_n\} \tag{8}$$

by inserting (7) into (8) we get

$$\sigma_{Zn}^2 = E\{Z_n^2\} \tag{9}$$

At this point, we can make a further step ahead, which has no counterpart in the classical Fourier series: we wish to introduce a new sequence of positive numbers λ_n such that for every λ_n:

$$\sigma_{Zn}^2 = \lambda_n = E\{Z_n^2\} > 0 \tag{10}$$

This equation provides the answer to the next natural question: do the random variables Z_n fulfill a new type of orthonormality somehow similar to what the classical orthonormality (4) is for the $\phi_n(t)$? Since we are talking about random variables, the orthogonality operator can only be understood in the sense of statistical independence. The integral in (4) must then be replaced by the average operator $E\{\}$ for the random variables Z_n. In conclusion, we find that the random variables Z_n must obey the important equation:

$$E\{Z_m Z_n\} = \lambda_n \delta_{mn} \tag{11}$$

In this equation, we were forced to introduce the positive λ_n in the right-hand side in order to let (11) reduce to (10) in the special case $m = n$. As for the KL equivalent of the Euler formulae (3) of the Fourier series, from the KL series (2) and the orthonormality (4) of the $\phi_n(t)$, one immediately finds that

$$Z_n = \int_0^T X(t)\phi_n(t)dt \tag{12}$$

In other words: the random variables Z_n are obtained from the given stochastic process $X(t)$ by projecting $X(t)$ over the corresponding eigenvector $\phi_n(t)$. If one likes the language of mathematicians and of quantum physics, then one may say that this projection of $X(t)$ onto $\phi_n(t)$ occurs in the Hilbert space, which is the infinitely-dimensional Euclidean space spanned by the eigenvectors $\phi_n(t)$ so that the square of $\phi_n(t)$ is integrable over the finite time span $0 \le t \le T$.

To sum up, we have achieved a remarkable generalization of the Fourier series by defining the Karhunen-Loève expansion (2) as the only possible statistical expansion in which all the expansion terms are uncorrelated from each other. The word "uncorrelated" comes from the fact that the autocorrelation of a random function of the time, $X(t)$, is defined as the mean value of the product of $X(t)$ at two different instants t_1 and t_2:

$$R_{XX}(t_1, t_2) \equiv R_X(t_1, t_2) = E\{X(t_1)X(t_2)\} \tag{13}$$

If we assume, according to (5), that the mean value of $X(t)$ vanishes identically in the interval $0 \le t \le T$, the autocorrelation (13) reduces to the variance of $X(t)$ when the two instants are the same

$$\sigma_{X(t)}^2 = E\{X^2(t)\} = E\{X(t)X(t)\} = R_{XX}(t, t) \tag{14}$$

Let us add one final remark about the basic notion of statistical independence of the random variables Z_n. It can be proven that, while the Z_n in (2) are always uncorrelated (by construction), they are also statistically independent if they are Gaussian-distributed random variables. This is fortunately the case for the Brownian motion and for the background noise we study in SETI. Therefore, we are not concerned about this subtle mathematical distinction between uncorrelated and statistically independent random variables.

2.4 The KLT finds the best basis (eigenbasis) in the Hilbert space spanned by the eigenfunctions of the autocorrelation of *X(t)*

Up to this point, we have not given any hint about how to find the orthonormal functions of the time, $\phi_n(t)$, and positive numbers λ_n i.e. the variances of the corresponding uncorrelated random variables Z_n. In this section, we solve this problem by showing that the $\phi_n(t)$ are the eigenfunctions of the autocorrelation $R_X(t_1, t_2) = E\{X(t_1)X(t_2)\}$ and that the λ_n are the corresponding eigenvalues. This is the correct mathematical phrasing of what we are going to prove. However, in order to ease the understanding of the further mathematics involved hereafter, a translation into common language of "common words" is now provided. Consider an object, for instance a book, and a three-axes rectangular reference frame, oriented in an arbitrary fashion with respect to the book. Then, the classical Newtonian mechanics shows that all the mechanical properties of the book are described by a 3x3 symmetric matrix called the "inertia matrix" (or, more correctly, "inertia tensor") whose elements are, in general, all different from zero. Handling a matrix whose elements are all nonzero is obviously more complicated than handling a matrix where all entries are zeros except for those on the main diagonal (i.e. a "diagonal matrix"). Thus, one may be led to wonder whether there exists a certain transformation of exists that changes the inertia matrix of the book into a diagonal matrix. Newtonian mechanics shows that there exists only one privileged orientation of the reference frame with respect to the book which yields a diagonal inertia matrix: the three axes must coincide with a set of three axes (parallel to the book edges) called "principal axes" of the book, or "eigenvectors" or "proper vectors" of the inertia matrix of the book. In other words, each body possesses an intrinsic set of three rectangular axes that best describes its dynamics, i.e. in the most concise form. This was proven again by Euler; one can always compute the position of the eigenvectors with respect to a generic reference frame by means of a certain mathematical procedure called "finding the eigenvectors of a square matrix". In a similar fashion, one can describe any stochastic process $X(t)$ by virtue of the statistical quantity called the autocorrelation (or simply the correlation), defined as the mean value of the product of the values of $X(t)$ at two different instants t_1 and t_2, and formally written $E\{X(t_1)X(t_2)\}$. The autocorrelation, obviously symmetric in t_1 and t_2, plays for the stochastic process $X(t)$ just the same role as the inertia matrix for the book example above. Thus, if one first seeks the eigenvectors of the correlation, and then changes the reference frame over to this new set of vectors, one achieves the simplest possible description of the whole (signal+noise) set. Let us now translate the aforementioned description into equations. First of all, we must express the autocorrelation $E\{X(t_1)X(t_2)\}$ by virtue of the KL expansion (2). This goal is achieved by writing down (2) for two different instants, t_1 and t_2, taking the average of their product, and then (freely) interchanging the average and the summations on the right hand side. The result is

$$E\{X(t_1)X(t_2)\} = \sum_{m=1}^{\infty} \sum_{n=1}^{\infty} \phi_m(t_1)\phi_n(t_2)E\{Z_m Z_n\} \tag{15}$$

Taking advantage of the statistical orthogonality of the Z_n, given by (10), (15) simplifies to

$$E\{X(t_1)X(t_2)\} = \sum_{m=1}^{\infty} \lambda_m \phi_m(t_1)\phi_m(t_2) \tag{16}$$

Finally, we now want the $\phi_n(t)$ to disappear from the right hand side of (16) by taking advantage of their orthonormality (4). To do so, we multiply both sides of (16) by $\phi_n(t_1)$ and then take the integral with respect to t_1 between 0 and T. One then gets:

$$\int_0^T E\{X(t_1)X(t_2)\}\phi_n(t_1)dt_1 = \sum_{m=1}^{\infty} \lambda_m \phi_m(t_2) \int_0^T \phi_m(t_1)\phi_n(t_1)dt_1 = \sum_{m=1}^{\infty} \lambda_m \phi_m(t_2)\delta_{nm} = \lambda_n \phi_n(t_2) \tag{17}$$

that is:

$$\int_0^T E\{X(t_1)X(t_2)\}\phi_n(t_1)dt_1 = \lambda_n\phi_n(t_2) \tag{18}$$

This basic result is an integral equation, called of the Fredholm type by mathematicians. Once the correlation $E\{X(t_1)X(t_2)\}$ of $X(t)$ is known, the integral equation (18) yields (upon its solution, that may not be easy at all to find analytically!) both the Karhunen-Loève eigenvalues λ_n and the corresponding eigenfunctions $\phi_n(t_2)$. Readers familiar with quantum mechanics will also recognize in (18) a typical eigenvalue equation with the kernel $E\{X(t_1)X(t_2)\}$. Let us finally summarize what we have proven so far in sections 2.3 and 2.4, and let us use the language of signal processing, which will lead us directly to SETI, the main theme of this paper. By adding random noise to a deterministic signal, one obtains what is called a "noisy signal" or, in case the signal power is much lower than the noise power, "a signal buried into the noise". The noise+signal is a random function of the time, denoted hereafter by $X(t)$. Karhunen and Loève proved that it is possible to represent $X(t)$ as the infinite series (called KL expansion) given by (2), and this series is convergent. Assuming that the (signal+noise) correlation $E\{X(t_1)X(t_2)\}$ is a known function of t_1 and t_2, then the orthonormal functions $\phi_n(t)(n = 1, 2...)$ turn out to be just the eigenfunctions of the correlation. These eigenfunctions $\phi_n(t)$ form an orthonormal basis in what physicists and mathematicians call the space of square-integrable functions, also called the Hilbert space. The eigenfunctions are actually the best possible basis to describe the (signal+noise), much better than any classical Fourier basis only made up by sines and cosines. One can conclude that the KLT automatically adapts itself to the shape of the (signal+noise), regardless of the behaviour in time it may have, by adopting the basis spanned by the eigenfunctions, $\phi_n(t)$, of the autocorrelation of the (signal+noise), $X(t)$, as new reference frame in the Hilbert space.

2.5 Continuous vs discrete time in the KLT

The KL expansion in continuous time, t, is what we have described so far. This may be more "palatable" to theoretical physicists and mathematicians inasmuch as it may be related to other branches of physics, or of science in general, in which time must obviously be a continuous variable. For instance, Claudio Maccone this author spent 15 years of his life (1980-1994) to mathematically investigate the connection between Special Relativity and KLT. The result was the mathematical theory of optimal telecommunications between the Earth and a relativistic spaceship either receding from the Earth or approaching it. Although this may sound like mathematical science fiction to some folks, the possibility that, in the future, humankind will send out relativistic automatic probes or even manned spaceships, is not unrealistic. Nor is it science fiction to imagine that an alien spaceship might approach the Earth, slowing down from relativistic speeds to zero speed. Thus, a mathematical physics book like[6] can make sense. There, the KLT is obtained for any acceleration profile of the relativistic probe or spaceship. The result is that the KL eigenfunctions are Bessel functions of the first kind (suitably modified) and the eigenvalues are determined by the zeros of linear combinations of these Bessel functions and their derivatives. Other continuous-time applications of the KLT are to be found in other branches of science, ranging, for instance, from genetics to economics. However, whatever the application may be, if the time is a continuous variable, then one must solve the integral equation (18), which may require considerable mathematical skills. In fact, (18) is, in general, an integral equation of the Fredholm type, and the usual iterated nuclei procedure used to solve Fredholm integral equations may be particularly painful to achieve. It is much easier to reduce the Fredholm integral equation to a Volterra integral equation, as shown in the book[6] for the time-rescaled Brownian motion in relation to Special Relativity. But let us go back to the time variable t in the KL expansion (2). If this variable is discrete, rather than continuous, then the picture changes completely. In fact, the integral equation (2) now becomes ... a system of simultaneous algebraic equations of the first degree, which can always be solved! The difficulty here is that this system of linear equations is huge, because the autocorrelation matrix is huge (hundreds or thousands of elements are the rule for autocorrelation matrices in SETI and in other applications, like image processing and the like). The characteristic equation is also huge, i.e. the algebraic equations whose roots are the KL eigenvalues. Can you imagine directly solving an algebraic equation of degree 1 million ? Thus, the KLT is practically impossible to find numerically, unless we resort to simplifying tricks of some kind, as has been done by the SETI-Italia team ([7]) since 2007.

2.6 The KLT: just a linear transformation in the Hilbert space

We explained the KL expansion (2), but did not explain what the KL Transform is yet. We do so in this section. The next step is the rearrangement of the eigenvalues λ_n in decreasing order of magnitude. Let us uppose we have done this. We also rearrange the eigenfunctions $\phi_n(t)$ so that each eigenfunction corresponds to its own eigenvalue. It can be proved that no mismatch can possibly arise in doing so, inasmuch as each eigenfunction corresponds to one eigenvalue only, namely it can be proved that there is no degeneracy (contrary to what happens in quantum physics, where, for instance, there is a lot of degeneracy in the eigenfunctions of even the simplest atom of all, the hydrogen atom). Furthermore, all eigenvalues are positive, and so, once rearranged in decreasing order of magnitude, they form a decreasing sequence where the first eigenvalue is the largest one, and is called the "dominant" eigenvalue by mathematicians. We are now ready to compute the Direct KLT of the (signal+noise). We use the new set of eigen-axes to describe the (signal+noise). Then, in the new representation, the (signal+noise) is simply the Direct KLT of the old (signal+noise). In other words, the KLT transform is a linear transformation of axes (Incidentally, this accounts for the title of Karhunen's first paper Ueber Lineare Methoden in der Wahrscheinlichkeitsrechnung = On the Linear Methods in the Calculus of Probabilities, ref. [1], which obviously refers to the linear character of the transformation of axes in the Hilbert space).

2.7 Numerical KLT

In a general way, our goal is to find a set of vectors picked with a telescope like the Sardinia Radio Telescope[8](SRT), and describe the autocorrelation matrix of the input random process $X(t)$. Let us suppose that the data frame has N elements; as a consequence, the autocorrelation matrix A is of the order N, i.e. $A \in R^{NxN}$. We can rewrite the equation (18) in the discrete form:

$$A\upsilon = \lambda\upsilon \tag{19}$$

where λ is the eigenvalue and υ is the eigenvector. The autocorrelation and the eigenvalues are linked by the following formula:

$$NA(0,0) = \sum_{i=1}^{N} \lambda_i \tag{20}$$

where $A(0,0)$ represents the energy of the acquired signal.

If we consider a matrix of order N, we can extract N eigenvalues and N eigenvectors. Each eigenvector has an associated eigenvalue; we can define, as eigenpair, the pair of elements (υ_i, λ_i), where υ_i is the eigenvector or i-th axis in the Hilbert space, while λ_i is the eigenvalue indicating the energy associated with the eigenvector.

The matrix A has N axes representing the eigenvectors, thus we can decompose the acquired data frame as the sum of eigenvectors weighted by the values of an appropriate vector p. The equation (2) can be rewritten as follows:

$$x = V\vec{p} \tag{21}$$

where V is the following matrix containing the eigenvectors, while $\vec{p} = [p1, p2, ...pN]$

$$V = \begin{pmatrix} \upsilon_0(1) & \upsilon_1(1) & ... & \upsilon_{N-1}(1) \\ \upsilon_0(2) & \upsilon_1(2) & ... & \upsilon_{N-1}(2) \\ . & . & & . \\ . & . & & . \\ \upsilon_0(N) & \upsilon_1(N) & ... & \upsilon_N(N) \end{pmatrix}$$

Consequently, the equation (2), in the discrete case, can be written as:

$$x(t) = \sum_{i=1}^{N} V_i(t) p_i \tag{22}$$

where the weights p, in analogy with the equation (12), can be obtained by the following formula:

$$p_i = \sum_{k=1}^{N} V_i^T(k) x(k) \tag{23}$$

In other words, weights are obtained via a projection of the acquired data on the transformation matrix given by the eigenvectors. In the context of this paper, we are using the use of the KLT to reveal possible alien signals, therefore we are not interested the reconstruction of the original signal. Hence, the KLT algorithm that we have used is partially truncated to facilitate the heavy digital signal processing that is required.

3. IMPLEMENTED KLT ALGORITHM

The FFT algorithm divides the overall energy contained in the acquired signal into a certain number of pieces; therefore, we can establish a threshold to determine whether, in a certain piece of energy (called channel or bin), a signal stronger than the threshold is present. Our adopted KLT algorithm ([9]) uses a different approach: it considers all of the signal's energy instead of the energy of a particular channel. Figure 1 shows our flow chart.

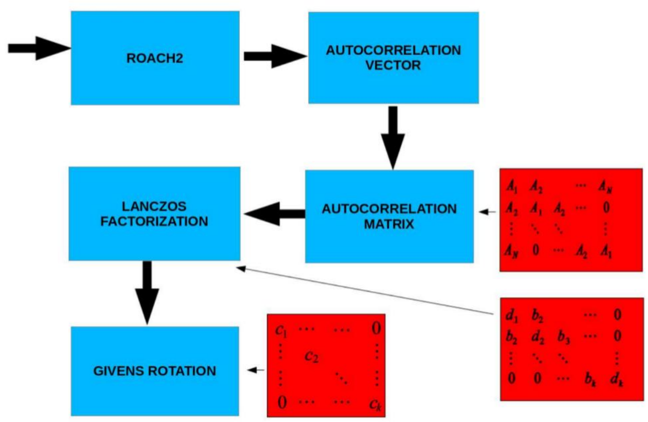

Figure 1. *Block diagram of the adopted algorithm.*

The "autocorrelation vector" block computes the autocorrelation of the data provided by the ROACH2 board and outputs a vector "[0, 1, 2, ...N-1,..., 2N-1]". This vector allows the calculation of the autocorrelation matrix in the following block. The matrix has a particular structure: it is symmetric and with constant elements on each diagonal; usually, such a kind of matrix is named "Toeplitz matrix". For instance, with a vector $[edcbabcde]$, the matrix A has the following structure:

$$A = \begin{pmatrix} a & b & c & d & e \\ b & a & b & c & d \\ c & b & a & b & c \\ d & c & b & a & b \\ e & d & c & b & a \end{pmatrix}$$

The "Lanczos factorization" block reduces the matrix A - which is of order N - so as to make the calculation of the eigenvalues less complex. The output of this block is a square matrix (tri-diagonal) of order $k << N$.

The last block, named "Givens rotation", extracts the eigenvalues from the aforementioned tri-diagonal matrix by exploiting the algorithm known as "Givens rotation". This algorithm provides a diagonal matrix - of order k - that contains the k approximated eigenvalues of the matrix A.

Essentially, the goal is to get the axes that describe the acquired signals in the best possible way; this corresponds to the calculation of eigenvectors/eigenvalues of the autocorrelation matrix of these signals. Once we obtain the eigenvalues and the eigenvectors, a specific energy evaluation can be done by SETI experts.

4. ADOPTED DIGITAL PLATFORM

Since SETI research is often conducted in piggyback mode only, a dedicated conditioning module is employed at the level of the intermediate frequencies. In particular, it acts as a frequency compensation mechanism necessary for keeping the chosen bandwidth stable even when a Doppler tracking system is activated, or in case the chosen bandwidth changes completely during the observation. Figure 2 shows a block diagram of this scenario.

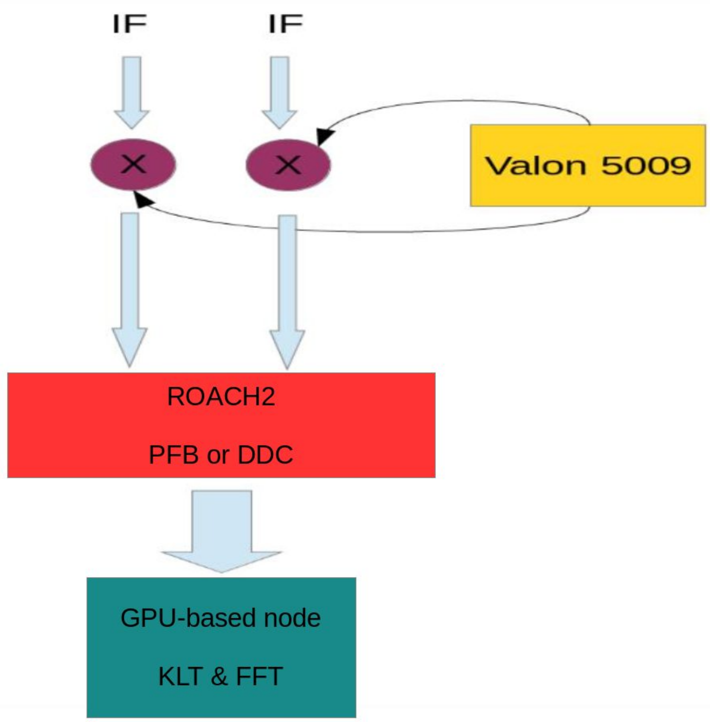

Figure 2. *Block diagram for piggyback SETI.*

The intermediate frequency signals are beaten by a tunable tone generated by the synthesizer "Valon 5009" ([10]) so as to compensate possible Doppler tracking programs, as said earlier. Astronomers, in fact, can observe a particular known source, therefore the Doppler correction is needed to keep a particular narrow-band emission in a certain channel, usually the central one of the entire bandwidth of interest. Moreover, in case the observer moves the band to observe another portion of the spectrum, the synthesizer can be used to keep the preceding sub-band. Clearly, if the band is moved to a totally different area, the ongoing SETI search ought to be stopped and a new one - starkly different - must begin.

The KLT algorithm is very heavy from a computational point of view, and in the next chapter we discuss all of these aspects numerically. In any case, the acquired bandwidth with the ROACH2 ([11]) (up to 2.5 GHz) must be reduced in order only to store a portion in base-band mode. We have two possibilities: we can either implement the polyphase filter bank or, alternately, the digital down conversion; we have decided to use the latter. In particular, the acquired signal is beaten with a complex mixer turning it into a complex form - i.e. with a real and an imaginary part - and a decimating filter is applied to the signal. As a consequence, the complex signal is base-band converted with the center of the chosen bandwidth at zero. Figure 3 shows these steps graphically:

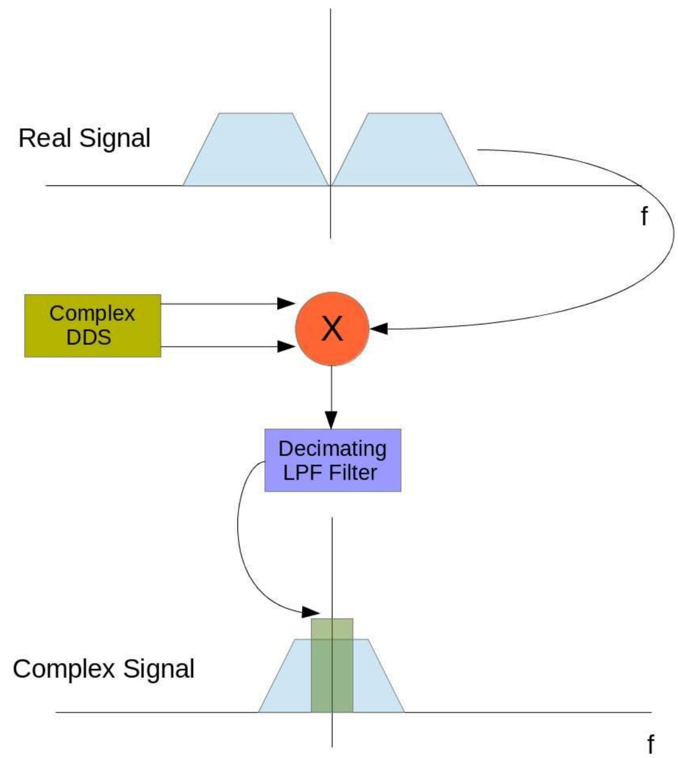

Figure 3. *Digital Down Conversion of the acquired signal.*

Finally, the signal is sent to a GPU node within which the KLT and the FFT are performed. The signal's format is already ready for the FFT engine, because there are several architectures that able to perform the FFT for both real and complex signals. On the contrary, the KLT algorithm uses real signals, therefore a choice must be made: we can use the absolute value of the complex number or, alternately, we keep the real part of the signal. However, in the latter case, a filter is mandatory for avoiding aliasing phenomena.

However, the best way to do this is to implement a complex-to-real conversion ([12]), by translating the output of the filter by 1/4 of the sample frequency. Essentially, if C_n is the output of the filter at the sample n, the real signal is given by $S_n = Re(C_n exp(2\Pi in/4))$. The exponential assumes, cyclically, the values (1, -i, -1, +i), i.e. it has only integer real or imaginary components, and can be simply implemented by the scheme in fig. 4.

5. REAL-TIME KLT IMPLEMENTATION

As a starting point, in order to estimate the computational time required, we implemented the "original" KLT code written in C language ([9]) and running in the conventional CPU mode. Of course, the KLT takes a time closely associated with the number of samples employed. Moreover, the quality of the eigenvalues improves by increasing the number of samples, therefore we should use as many samples as possible, in agreement with the available computing resource: 81920 samples was the final choice. The number of eigenvalues can be set via software: for our purpose we have chosen to calculate the first 70.

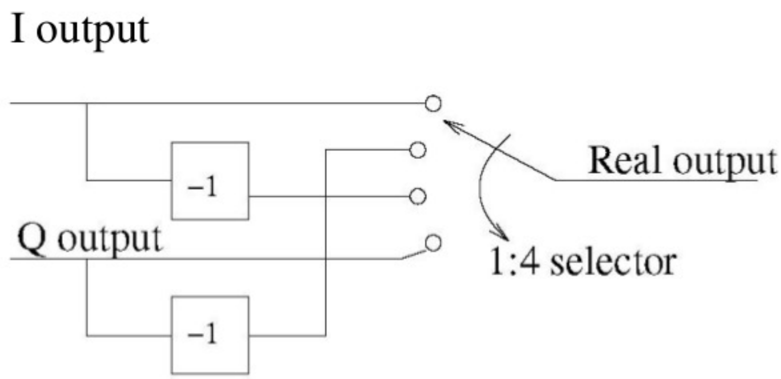

Figure 4. *Complex to Real conversion.*

The CPU version has been tested in a node with the motherboard SuperMicro X9DRI-F DUALXEON LGA2011 ([13]) equipped with two CPU INTEL XEON 8 CORE E5-2640V2 2,0/2,5ghz 20MB LGA2011: the elaboration took approximately 9 seconds, clearly very far from our goal.

The next step was to determine which parts of our program are taking most of the execution time; we employed the well-known $gprof$ profiler ([14]). Profiling allows us to learn where the program spent its time and which functions called which other functions.

The "flat" profiler for the KLT algorithm gave us the following results: the functions "Correlation2" and the "dotprx" took more than half of the time necessary to execute the program.

dotprx is used to multiply a vector by a matrix. If we indicate with A a NxN matrix and with v a vector of order N, the function *dotprx* performs the calculation Av.

$Correlation2$ is used to get the autocorrelation from a vector containing the base-band data provided by the ROACH2 board. The autocorrelation vector is a symmetric vector with $2 * N$ elements.

The aforementioned node is also equipped with a NVidia GTX 980 Ti ([15]), i.e. a GPU board with 2816 cores and 6 GB of RAM memory. We implemented the functions $Correlation2$ and *dotprx* in CUDA, in order to make a parallel implementation of those pieces of code for which this is convenient. Unfortunately, we had no time to verify the results of the profiler on this new "parallelized" version, however the overall time required to perform the KLT on a slot of 81920 was 2 seconds, dramatically reduced in comparison with the original version. This means that, in accordance with the Nyquist theorem, the maximum instantaneous bandwidth that we are able to process in a real-time mode is roughly 20 KHz, not enough for massive SETI but certainly a very good starting point. At this stage of the development, if a large bandwidth is desired, we need to store base-band data and then post-process them. However, we are going to do a further optimization of the algorithm so as to increase the number of simultaneous samples and, at the same time, to reduce the time required as much as possible.

6. CONCLUSIONS

The KLT is a powerful mathematical tool for all branches of scientific research because it allows the extraction of very weak signals out of background noise of any kind, not just out of white noise, as the FFT does. It also goes under the name of Principal Axis Transformation since it works in the Hilbert space spanned by the eigenfunctions of the autocorrelation of the (noise+signal) input of every antenna and, in particular, every radio telescope. It is also used in space missions as the best lossy compression algorithm keeping the radio link, since it approximates a large amount of data by transmitting only the most important part of the data (i.e. the first few eigenvectors) while discarding the rest as merely noise. However, given an autocorrelation matrix of size N, the KLT scales like N*N rather than like N*ln(N), as the FFT does. This is precisely why scientists preferred to use the FFT instead of the KLT over the last 50 years: they preferred the FFT's shorter computing time to the KLT far better filtering capabilities. But now the situation has changed: the advent of extremely fast computing boards (GPUs) makes the KLT computationally feasible, and that is particularly important in SETI. In SETI, in fact, we know little or nothing about the ET signals that we would like to detect, or about the noise in which they are embedded. Thus, at the Sardinia Radio Telescope[16] we are now implementing a KLT-based software to do SETI searches (with all three available receivers: L-P[17] band,, C-band and K-band) that might be able to detect signals otherwise beyond our reach. We regard Italy as one of the leading countries in SETI, at least in this regard.

REFERENCES

[1] Maccone, C., [*Deep Space Flight and Communications*], Praxis-Springer, first ed. (2009).

[2] Maccone, C., [*Mathematical SETI*], Praxis-Springer, first ed. (2012).

[3] K.Kahrunen, "Uber lineare methoden in der wahrscheinlichkeitsrechnung," *Ann. Acad. Sci. Fennicae, Series A 1, Math. Phys* **37**, 3–79 (1946).

[4] Loève, M., "Fonctions alatoires de second ordre," *Rev. Sci.* **84, n.4**, 195–204 (1946).

[5] Loève, M., "Probability theory: Foundations, random sequencies," *Princeton, NJ: Van Nostrand* (1955).

[6] Maccone, C., [*Telecommunications, KLT and Relativity Volume 1*], IPI press, Colorado Springs, Colorado, USA (1994).

[7] Montebugnoli, S., Bortolotti, C., Caliendo, D., Cattani, A., D'Amico, N., Maccaferri, A., Maccone, C., Monari, J., Orlati, A., Pari, P., Poloni, M., Poppi, S., Righini, S., Roma, M., and Teodorani, M., "Seti-italia 2003 status report and first results of a kl transform algorithm for eti signal detection," *paper IAC-03-IAA.9.1.02 presented at the 2003 International Astronautical Congress held in Bremen* (2003).

[8] http://www.srt.inaf.it/.

[9] Pari, P., "Rilevazione di segnali tramite trasformata klt nel progetto seti," *Thesis at Universita' degli Studi di Bologna, Facolta' di Matematica, Fisica e Scienze Naturali* (2003).

[10] http://valontechnology.com/5009%20Product%20Info.pdf.

[11] https://casper.berkeley.edu/wiki/ROACH-2_Revision_2.

[12] Comoretto, G., "Design of a fir filter using a fpga," *Arcetri Technical Report n.5/2002* (2002).

[13] http://www.supermicro.com/products/motherboard/Xeon/C600/X9DRi-F.cfm.

[14] https://sourceware.org/binutils/docs/gprof/.

[15] http://www.nvidia.it/object/geforce-gtx-980-ti-it.html.

[16] Melis, A., Concu, R., Maccone, C., Migoni, C., Montebugnoli, S., Valente, G., Trois, A., Perrodin, D., Possenti, A., Schilliró, F., Barbaro, M., Lunesu, M., Monari, J., Navarrini, A., Pari, P., Pisanu, T., Vacca, V., Cobb, J., MacMahon, D., Siemion, A., and Werthimer, D., "Perspective for the sardinia radio telescope as a seti facility," *OAC Internal Report* (2015).

[17] Valente, G., Pisanu, T., Bolli, P., Mariotti, S., Marongiu, P., Navarrini, A., Nesti, R., Orfei, A., and Roda, J., "The dual-band l-p feed system for the sardinia radio telescope prime focus," *Proc. SPIE Millimeter, Submillimeter, and Far-Infrared Detectors and Instrumentation for Astronomy V* **7741** (2010).