Parametric adaptive filtering and data validation in bar gw detectors

A. Ortolan‡*, L. Baggio†, M. Cerdonio§, G. A. Prodi†, G. Vedovato‡ and S. Vitale†

‡INFN - Laboratori Nazionali di Legnaro, Via Romea, 4 I-35020 Legnaro (PD) Italy
†Department of Physics, University of Trento and I.N.F.N. Gruppo Collegato di Trento, I-38050, Povo, Trento, Italy
§Department of Physics, University of Padova and I.N.F.N. Sezione di Padova, Via Marzolo 8, I-35131 Padova, Italy

Abstract. We report on our experience gained in the signal processing of the resonant gw detector AURIGA. Signal amplitude and arrival time are estimated by means of a matched-adaptive Wiener filter. The detector noise, entering in the filter set up, is modelled as a parametric ARMA process; to account for slow non-stationariety of the noise, the ARMA parameters are estimated on a hourly basis. A remarkable aspect of the set up of a unbiased Wiener filter is the separation of time spans with "almost gaussian" noise from non-gaussian and/or strongly non-stationary ones. The separation algorithm consists basically of a variance estimate with the Chauvenet convergence method and a threshold on the Curtosis index. The subsequent validation of data is strictly connected with the separation procedure: in fact, by injecting a large number of artificial gw signals in the "almost gaussian" part of the AURIGA data stream, we have demonstrated that the effective probability distributions of signal-to-noise ratio, χ^2 and time of arrival are the expected ones.

PACS numbers: 04.80.Nn, 95.85.Sz

* To whom correspondence should be addressed: ortolan@lnl.infn.it

1. Introduction

Cryogenic and ultracryogenic gravitational waves (gw) detectors have been in operation in the last years and, from 1997, they joined in the IGEC to search for impulsive gw events [1], working as an observatory able to detect, in the present configuration of detectors, violent events in the Galaxy, as SN explosions or NS/BH mergers [2]. The operation of the AURIGA detector [3] in a world-wide community poses demanding requests to its daq and data analysis systems, namely: i) synchronization with the UTC within 1 μs; ii) fault tolerance to ensure continuous data taking; iii) robust and efficient data analysis to search for rare, impulsive gw events; vi) I/O of daq and data analysis in a common format for the data exchange (currently the agreed format for a complete exchange of data is the VIRGO/LIGO *frame format* [4], widely accepted by the gw community).

Signal processing plays a central role in the search for gw since the detector performances are limited by the ability to extract small signals buried into the detector noise. The detection of a gw burst at low signal to noise ratio (e.g. $SNR > 3 \div 4$ i.e. $\sim 10\ dB$) challeged us to face up to a twofold problem: 1) the correct description of noise properties (stationarity, gaussianity, correlation function, etc.) and 2) the choice of the template suitable for the incoming gw signal.

In fact, a signal can be fully reconstructed from its samples and its parameters correctly estimated without biases only if 1) and 2) are met; a remarkable example for the AURIGA data analysis is the reconstruction of the candidate gw signals by sample interpolation to allow the estimate of its arrival time up to μs accuracy [5].

The basic idea behind the set up of the correct filters is the separation of the almost gaussian noise from the (spurious) excitations, which continuously occur in the data stream. It is clear that these fast transients might lead to a misestimate of filter, inducing therefore systematic errors in amplitude and/or arrival time of gw signals. For the AURIGA detector, it turns out convenient to freeze completely the parameters values during transients.

On the other side, to cope with the problem of biases in the signal estimation, we can resort to the maximum likelihood criterion; it is equivalent, in the presence of gaussian noise, to the standard Wiener filtering together with the χ^2 test of the goodness of the fit [6]. The χ^2 value (which is statistically independent of the amplitude for signals which pass the test) is used, in the framework of the AURIGA data analysis, to test the consistency of our *a priori* hypothesis on the signal template and noise properties.

A "near-optimal" data analysis is able to recognize gw signals and extract the signal parameters without distortion of their probability distributions (e.g. mass, orbit phase and orientation, gw amplitude, polarization, and arrival time for a coalescing binary system). To test the data analysis performances, *artificial* signals can be injected into the data stream either by a calibration procedure or, by an addition of signal samples (generated via software) to the detector noise. When the number of the inserted waveforms is very large ($> 10^4$), the latter procedure becomes a sort of Monte Carlo which characterizes the entire data analysis process with the properties of the real noise; with the help of the Monte Carlo, we are also able to estimate the detection efficiency ε, (required by any gw search performed having *a priori* no idea when signals are likely to arrive), and to assess the "effective probabilities" of time of arrival, amplitude, χ^2 and any other set of signal parameters, in the presence of the real noise, rather than the simple model of stationary and gaussian noise. The Monte Carlo can give also a quantitative definition of "almost gaussian" and "quasi-stationary" properties of the real noise as, for instance, we may allow the outliers of the expected distribution of signal parameters to be less than, say, few %. The Plan of the paper is as follows. In Sect. 2 we give some details of the AURIGA acquisition system. Sect. 3 is devoted to illustrate some features of the AURIGA data analysis, including the setup of the noise model, the setup of the signal template, and the data quality and data validation procedures to avoid biases in signal estimation. Some results of the AURIGA data analysis, with an implementation of the Monte Carlo for impulsive signals, are discussed in Sect 4. Finally, our conclusion are presented in Sect. 5.

2. Data acquisition

The AURIGA daq system, hardware and software architecture, is described in detail elsewhere [7]; here we report only its main features and current upgrades. An ADC with high resolution (18 equivalent bits, 110 $dBfs$) and low distortion ($< 110\ dBfs$) digitizes the detector output after the dc-SQUID amplifier with a sample rate of 5 kHz. A second multiplexed ADC (24 channels) acquires the data from the accessory instrumentation to monitor the detector environmental noise (e.g. seismic and electromagnetic noises). The sample rate and resolution of the second ADC are respectively 200 Hz and 16 bits. Both ADCs (HP1430A and HP1413A) are housed in the same VXI crate which is connected to a dedicated acquisition PC (Linux OS) through the MXI interface. The synchronization of the acquired data with the Universal Time (UTC) is achieved by the GPS100/S80 apparatus which provides the time stamps to date the triggers of the daq system. We gained an high time accuracy ($< 1\ \mu\ sec$) in tagging the data buffers using the hardware interrupts (IRQs) generated by the ADCs when a data buffer is ready for the acquisition. In the upgraded version of the daq, which will be ready for the second AURIGA run (late 2001), the acquired data are collected and formatted according to the VIRGO/LIGO *frame format* [4] and fed to removable 70 GB hard disks for definitive storage. To avoid unwanted losses of data, disks are also backed up in 35 GB DLT cassettes. A C++ library for the daq, the Process Control Library (PCL), has been developed for the control of processes interfaces and for interprocess comunication. It is worth noticing that, in case of malfunctions of ADCs or PCs, we are able to restart quickly the acquisition process as the acquisition chain (ADCs, VXI crate and PC) has been completely duplicated.

3. Data analysis

The AURIGA data analysis has been developed with the aim of recognizing characteristic gravitational waveforms in a noisy detector output. This complicated task can be successfully dealt with some simplifying assumptions on noise and transfer function of the detector: i) the dynamics of the system can be described (within the frequency band useful for the gw detection) by linear differential equations; ii) the noise can be represented by a zero-mean, gaussian, stochastic process. The stationariety assumption, which is implicitly contained in i) and ii), can be relaxed in the quasi-stationary one, in the sense that the time scale of variation of model parameters are much greater than the relaxation times (fixed by mechanical dissipations) of the systems. Within these quite general hypothesis, the whitening filter $L(i\omega)$ and the δ- matched filter $M_\delta(i\omega)$ (i.e. matched to the $\delta(t-t_a)$ gw template) have a simple representation as a pole-zero system; therefore, in the time discrete domain, they can be recast into ARMA processes with a significant decrease of computational costs [7]. The δ-matched filter provides a natural separation between the detector characterization (noise correlations and transfer function) and the search for physical waveforms, which can be conveniently performed off-line after this filter.

3.1. Adaptive filter: setup of the noise model

Within the Reduced Bandwidth $RB \approx [800 \div 1000]\ Hz$ useful for gw detection, a suitable model of the power spectrum of the AURIGA noise, is the complex zeroes-poles function derived in ref. [7]

$$S(\omega) \equiv L(i\omega)L(-i\omega) = S_0 \prod_{k=1}^{N_P} \frac{(q_k + i\omega)(q_k - i\omega)(q_k^* + i\omega)(q_k^* - i\omega)}{(p_k + i\omega)(p_k - i\omega)(p_k^* + i\omega)(p_k^* - i\omega)}, \quad (1)$$

where S_0 is a constant representing the wideband noise level, N_P is the number of resonances and p_k and q_k are respectively the zeroes and the poles of $S(\omega)$. The physical meaning of poles and zeroes, the reason for their variations and the precision required in their estimate is reported in ref. [7]; here we would like to discuss the adaptive algorithms devised to estimate the q_k which are the most sensitive parameters to noise variations, in particular the ratio between the narrowband and wideband noise levels which enters in the SNR and arrival time [8]; the problem of p_k estimation, being common to the setup of δ matched filter, will be discussed in the next section. The adaptive algorithm must check the compliance of the data stream with the noise model, in order to select the appropriate time spans for the q_k estimation. In fact, the presence of environmental disturbances worsen both the power spectrum and the cumulative distribution of samples, clearly introducing biases, as the set up of matched filters depends on $L(\omega)$. For instance, the AURIGA output often contains clustered signals that mimic the effect of an increase of narrow band noise or electric spikes that jeopardize an increase of wideband noise[8]. The gaussianity and quasi stationariety of the AURIGA output is monitored over buffers of 4096 samples corresponding to about

90 *sec.* The gaussianity algorithm consists of a variance estimate with the Chauvenet convergence method (i.e. a recursive estimate of the variance by discharging at each step the data exceeding 3 times the variance of the previous iteration) and a threshold on the curtosis index (4-th connected moment). This algorithm is applied to whitened data buffers and a data buffer is considered gaussian if its curtosis does not exceed 0.15 and the Chauvenet convergence method has discharged less than 2 % of data; in addition, a whitened data buffer should have a correlation index not larger than 0.04. These figures are 3 times the values we found using the simulated output of the AURIGA detector, assuming the noise gaussian and stationary. We apply the test on the correlation of whitened data to be sure that our model for the noise spectral density is close enough to the real one, even if few spectral peaks (50 Hz harmonics or sinusoidal components arising from mechanical vibrations of the suspension wires) are present in the reduced bandwidth.

The buffers of data which fails the gaussianity tests (Chauvenet and threshold on curtosis index) are dropped out from the data stream before applying the q_k tracking algorithm that converges in one hour to the correct values of noise parameters [8]. This selection procedure allows the filter paramenters to be adjusted for drifts on a time scale longer than the mechanical relaxation time of the system (several seconds), while ignoring changes due to disturbances on smaller timescales. Of course, the incorrect modelling of the noise produces unpredictable effects on signal search and biases on estimated signal parameters: for instance, if the estimate of noise variance fails, the reduced χ^2 statistics has no longer unitary mean.

3.2. Matched filter: setup of the signal template

The set up of the δ matched filter $M_\delta(i\omega) \equiv H(-i\omega)/L(-i\omega)$ requires the accurate measurement of detector transfer function $H(\omega)$. Systematic errors on amplitude and phase part of $H(\omega)$ translate directly on biases of signal amplitude and arrival time. For a resonant detector, at least within the RB, we can write

$$H(i\omega) = H_0(i\omega) \;\; \frac{(-i\omega)^{N_P+2}}{\prod_{k=1}^{N_P}(p_k - i\omega)(p_k^* - i\omega)} \,, \tag{2}$$

where $H_0(i\omega)$ (calibration function) must be provided by the detector calibration procedures at the startup of a run and monitored during the data taking. The poles p_k entering in eq. (1) and (2) are subject to slow drifts mainly caused by discharges of the capacitive transducer or variations of the thermodynamic temperature (usually less than few mHz per month). We set up a simple pole tracking algorithm by measuring the phase shifts of N digital lock-ins tuned to the poles frequency.

Trigger search (i.e. the identification of candidate events) is performed in the time domain, by a max-hold algorithm, which identifies the time and the amplitude of the extremes of the filtered data separated by at least a time span about 3 times the reciprocal of the effective bandwidth of the system (i.e. of the order of 1 second) [7]. The actual timing accuracy depends on the SNR. For signals with $SNR > 20$ it is

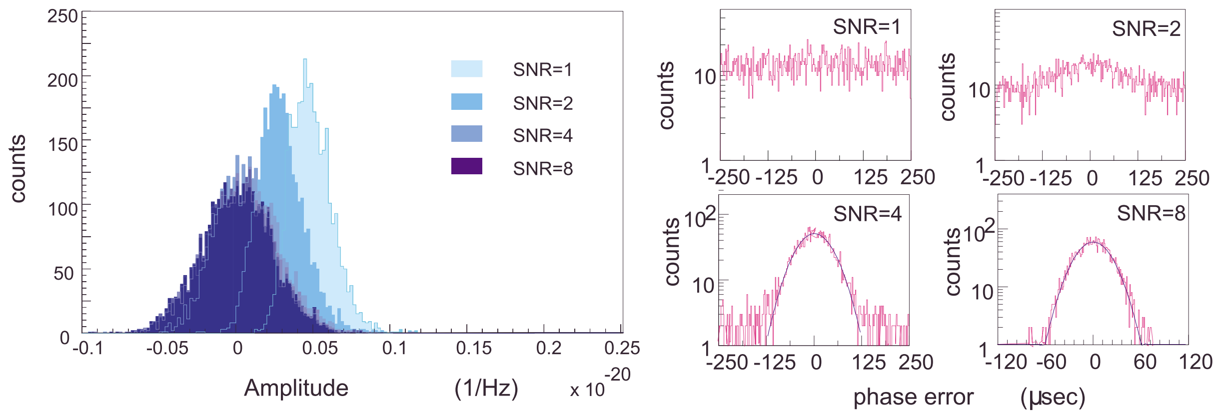

Figure 1. Monte Carlo of 3600 impulsive signals spread over 1 hour of AURIGA noise. On the left: Histograms of the detected deviates of event amplitudes, when the amplitude of injected events is set to $SNR = 1, 2, 4$ and 8. At high SNR ($SNR > 4$ is enough for the present bandwidth of the AURIGA detector) the histograms reproduce the zero-mean Normal density function of the underlying stochastic process, as predicted by linear estimate theory of signal amplitude. At $SNR < 4$ the max-hold algorithm is manifestly no more linear, and a bias toward greater amplitudes appears. On the right: Histograms of the phase error relative to the above example. The phase error is defined as $mod(t_d - t_a, T_0)$, where t_d and t_a are respectively the detected and true time of arrival and T_0 is the half period of oscillations in the filtered data [7]. The gaussian behaviour of deviates of phase error is recovered asymptotically at high SNR as expected from theory.

given approximately by 170 $\mu sec/SNR$. There is no amplitude threshold in the max-hold algorithm. An adaptive threshold $SNR_{thr} = 5$ is applied to the exchanged list of candidate events to perform the coincidence analysis with the other IGEC members. The reason for a threshold on exchanged candidate event is twofold: i) as one can easely recognize in Fig. 1 the trigger search algorithm has strong biases in amplitude and arrival time at least up to $SNR = 4 \div 5$; and ii) for $SNR > 5$ the false alarms rate falls down to acceptable levels for gw detections [2]. The bias at low SNR is originated by our lack of knowledge about the true time of arrival of the gw signal. The max-hold algorithm looks to the nearest local fluctuation of the noise without any phase relation with the injected waveform. As the SNR grows up, there is less chance for a noise fluctuation to reach such a SNR level and the max-hold locks to the real trigger. In this case, the estimated amplitude becomes unbiased and the time of arrival error strongly peaks around zero. It should be noticed that such a bias is unavoidable as the arrival time estimate is, in principle, a non linear algorithm, which can be linearized at high SNR around the true arrival time [9].

3.3. Data quality and data validation

The thorough study of the AURIGA noise with the help of refined releases of the data analysis lead us to a better knowledge of the behaviour of our detector. An important achievement for the noise estimation has been the separation of time spans with almost gaussian noise from the ones with non-gaussian and/or heavily non-stationary noise. Now we settle down to discuss the criterions for the *data quality*, i.e. how close are the data in given time span to the modelled noise, and for the *data validation*, i.e. which time spans are suitable for gw search. For what concern the data quality, we realized that the non modelled noise can be further separated into disturbances with known origin (e.g. detector maintenance activities) and other disturbances arising from sources that are beyond our control. In order to get off these excess noise sources and to form unbiased candidate event lists for the IGEC exchanges we set up a simple two-level tagging procedure which either accept or remove long standing time intervals (vetoes), where the noise is unmodellable, from the duty-cycle of the detector. We have devised a two level vetoing system: i) the first level vetoes are set by the experimentalist and represent time intervals, with duration larger than 1 minute, in which the detector is not operative due to maintenance activities or electronics malfunctioning; ii) the second level vetoes are automatically produced by the gaussianity tests on whitened and filtered data. To determine the time spans corresponding to the second level vetoes, we first assign the flag ON or OFF respectively to data buffers which satisfy or fail the tests and then form the sequence of the ON and OFF; if during 15 minutes (10 buffers) there are more than 4 OFF flags the period is considered not compliant with the modelled noise and therefore vetoed. The number of buffers was chosen to ensures enough ON buffers in a validated time span for tracking changes in detector noise (remind that the noise parameters are frozen within OFF buffers). In addition, we have imposed the validated time spans to last at least 20 minutes to avoid a fine grained structure in the AURIGA duty cycle.

4. Results and discussions

As already stated, to avoid biases in the estimate of signal parameters (e.g. amplitude, timing, χ^2, SNR for an impulsive waveform), the noise must be compliant with the adopted model and template must match the incoming signal waveform. For a gw detector, hypothesis testing, maximum likelihood, and χ^2 test are the basic statistical tools for gw detection or for assessment of upper limits and confidence levels; in fact, we can obtain indications about the correct estimation of noise (success of adaptive filtering procedures) and the matching of the candidate gw signal with the proper waveform (template matching).

The injection of artificial gw signals in the real noise of AURIGA is a powerful tool to study estimation biases in fuzzy condition for the noise or template choices. In fact, we are able to measure the real detection efficiency and the statistical significance of any signal observable (e.g. amplitude, arrival time) and, by means of the χ^2 statistics, any mismatch between the detector noise and/or transfer function with the injected waveform. In addition, we are able to study the effect of a template mismatching, which is crucial for the determination of tolerances on filter parameters in the case of

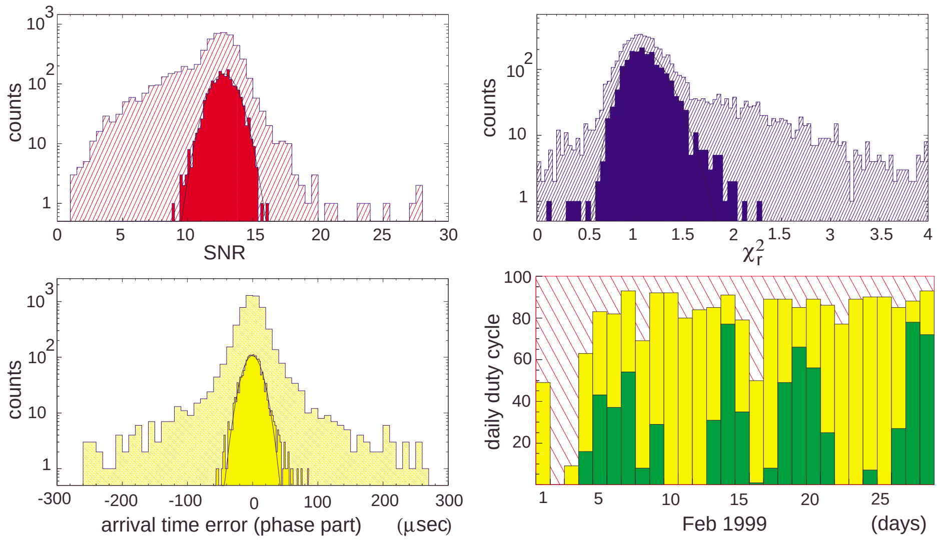

Figure 2. From left-bottom clockwise: Histograms of the detected deviates of time of arrivals (phase part), SNR and χ^2 obtained by superimposing, at random times, 5×10^3 impulsive events of $SNR = 12$ over 10 days of the AURIGA output. The dashed part of histograms represent all the measured deviates while the solid part represent the measured deviates within the validated time spans (see text). The fits of the solid histograms agree with the prediction of signal estimate in gaussian noise. The last figure shows the daily duty cycle of the AURIGA detector during the February 1999; dashed: vetoed time spans due to maintenance (first level vetoes); gray: vetoed time spans as the noise is non compliant with model (second level vetoes); dark gray: validated time span.

matching with a family of templates (e.g. coalescing binaries) which depends on many parameters. In Fig. 2 are summarized the results we get by injecting 5×10^3 impulsive waveforms, at random time and fixed $SNR = 12$, on 10 days of AURIGA data; the data were taken during February 1999. It is clear that the estimated parameters of the injected signals are recovered (and the corresponding parameters correctly estimated) within the validated time spans while in the vetoed intervals the outcomes of the event search procedures are strongly biased. Of course, this result only demonstrates that the empirical vetoing procedures based on gaussianity tests are strong enough to check the compliance of the AURIGA noise with our model. Much work has still to be devoted to investigate the data quality problems and to implement robust procedures which both maximize the duty cycle and validate the estimated parameters of detected gw signals.

5. Conclusions

The relevant parts of gw data analysis, such as the estimate of the AURIGA duty cycle, the vetoing procedures, the detection efficiency, the estimate of the noise properties (gaussianity and stationariety), the filters set up and effective probability of signal observables, are tightly entangled. The AURIGA data analysis is a first but robust step toward the identification of candidate gw signals (triggers) and the assessment of the probability to the corresponding observables, and the false alarm and false dismissal probabilities for threshold crossing searches. The separation of time spans where the noise is compliant with specific models is possible but costly in terms of duty cycle, as only 1/3 of the data taking time in the AURIGA runs 1997-1999 has been left. The statistics of the injected artificial gw signals demonstrate that the vetoing procedure is a sufficient algorithm but an improved data analysis could be able to reduce the vetoed time spans for instance relaxing some requirements on gaussianity or stationariety, maybe at a cost of a lower detection efficiency and different false alarm and false dismissal probabilities. The current upgrades in the hardware (suspensions and trasduction chain) [10] and software (daq and data analysis) sectors of the AURIGA experiments and the choice of the *frame format* for daq and data analysis I/O are intended for setting up a new run of a detector with observatory capabilities, i.e. with high duty cycle and with bandwidth and sensitivity useful to join the network of interferometric gw detectors under construction.

6. References

[1] G. A. Prodi *et al.*, Int. J. Mod. Phys. **D 9** (2000) 237.

[2] Z.A. Allen *et al.*, Phys. Rev. Lett. **85** (2000) 5046-5050.

[3] G. Prodi *et al.*, *Proc. of the* 2^{nd} *E. Amaldi Conference On Gravitational Waves (1997)*, eds. E. Coccia, G. Veneziano, G. Pizzella, (World Scientific, Singapore, 1998) p. 148.

[4] B. Mours, *Gravitational Wave Detection, Proc. TAMA International Workshop on Gravitational Wave Detection*, K. Tsubono, M. K. Fujimoto, K. Kuroda, eds. (Universal Academic Press, Tokyo, 1997) p. 27; Joint LIGO/VIRGO note LIGO-T970130-B and VIRGO-SPE-LAP-5400-102.

[5] V. Crivelli Visconti *et al.*, Phys. Rev. **D 57** (1998) 1; M. Cerdonio *et al.*, Mod. Phys. Lett. **A 12** (1997) 2261.

[6] S. Vitale *et al.*, *Proc. of the International Conference on Gravitational Waves*, I. Ciufolini and F. Fidecaro Eds. (World Scientific, Singapore, 1997) p. 256; L. Baggio *et al.*, Phys. Rev. **D 61** (2000) 10200.

[7] A. Ortolan, *et al.*, in *Proc. of the* 2^{nd} *E. Amaldi Conf. On Grav. Waves (1997)*, eds. E. Coccia, G. Veneziano, G. Pizzella, (World Scientific, Singapore, 1998) p. 204; S. Vitale *et al*, *First E. Amaldi Conference on G.W. Experiments*, (World Scientific, Singapore, 1995) p. 220.

[8] L. Baggio *et al.*, Int. J. Mod. Phys. **D 9** (2000) 251. G. A. Prodi *et al.*, in *Gravitational waves, proceeding of the "Third Edoardo Amaldi Conference*, (CalTech - California, 1999), S.Meshkov Ed., AIP Conference Proceedings, New York (2000) p. 345.

[9] C. W. Helstrom, *Statistical Theory of Signal Detection*, (Pergamon, Oxford, 1968).

[10] J. P. Zendri *et al.*, these Proceedings.