A simple and efficient line detection algorithm applied to Virgo data

F Acernese[6], P Amico[10], M Al-Shourbagy[11], S Aoudia[7], S Avino[6], D Babusci[4], G Ballardin[2], R Barillé[2], F Barone[6], L Barsotti[11], M Barsuglia[8], F Beauville[1], M A Bizouard[8], C Boccara[9], F Bondu[7], L Bosi[10], C Bradaschia[11], S Braccini[11], A Brillet[7], V Brisson[8], L Brocco[12], D Buskulic[1], G Calamai[3], E Calloni[6], E Campagna[3], F Cavalier[8], R Cavalieri[2], G Cella[11], E Chassande-Mottin[7], C Corda[11], A C Clapson[8], F Cleva[7], J P Coulon[7], E Cuoco[2], V Dattilo[2], M Davier[8], R De Rosa[6], L Di Fiore[6], A Di Virgilio[11], B Dujardin[7], A Eleuteri[6], D Enard[2], I Ferrante[11], F Fidecaro[11], I. Fiori[11], R Flaminio[1,2], J D Fournier[7], S Frasca[12], F Frasconi[2,11], A Freise[2], L Gammaitoni[10], A Gennai[11], A Giazotto[11], G Giordano[4], L Giordano[6], R Gouaty[1], D Grosjean[1], G Guidi[3], S Hebri[2], H Heitmann[7], P Hello[8], L Holloway[2], S Kreckelbergh[8], P La Penna[2], V Loriette[9], M Loupias[2], G Losurdo[3], J M Mackowski[5], E Majorana[12], C N Man[7], M Mantovani[11], F Marchesoni[10], E Marchetti[3], F Marion[1], J Marque[2], F Martelli[3], A Masserot[1], M Mazzoni[3], L Milano[6], C Moins[2], J Moreau[9], N Morgado[5], B Mours[1], A Pai[12], C Palomba[12], F Paoletti[2,11], S Pardi[6], A Pasqualetti[2], R Passaquieti[11], D Passuello[11], B Perniola[3], F Piergiovanni[3] L Pinard[5], R Poggiani[11], M Punturo[10], P Puppo[12], K Qipiani[6], P Rapagnani[12], V Reita[9], A Remillieux[5], F Ricci[12], I Ricciardi[6], P Ruggi [2], G Russo[6], S Solimeno[6], A Spallicci[7], R Stanga[3], R Taddei[2], D Tombolato[1], M Tonelli[11], A Toncelli[11], E Tournefier[1], F Travasso[10], G Vajente [11], D Verkindt[1], F Vetrano[3], A Viceré[3], J Y Vinet[7], H Vocca[10], M Yvert[1] and Z Zhang[2]

[1] Laboratoire d'Annecy-le-Vieux de Physique des Particules, 74941 Annecy-le-Vieux, France;
[2] European Gravitational Observatory (EGO), 56021 Cascina (PI) Italy;
[3] INFN Sez. di Firenze/Urbino and/or Università di Firenze, 50019 Sesto Fiorentino, and/or Osservatorio Astrofisico di Arcetri, 51125 Firenze, and/or Università di Urbino, 61019 Urbino, Italy;
[4] INFN, Laboratori Nazionali di Frascati, 00044 Frascati (RM) Italy;
[5] LMA, 69622 Villeurbanne, Lyon, France;
[6] INFN Sez. di Napoli and/or Università di Napoli "Federico II", 80126 Napoli, and/or Università di Salerno, 84084 Fisciano (SA), Italy;

Corresponding authors: É. Chassande-Mottin, CNRS, Observatoire de la Côte d'Azur, ARTEMIS, France (*Eric.Chassande-Mottin@obs-nice.fr*) and I. Fiori, INFN and Università di Pisa, Italy (*Irene.Fiori@pi.infn.it*).

[7] Observatoire de la Côte d'Azur, 06034 Nice, France;
[8] Laboratoire de l'Accélérateur Linéaire and/or CNRS-IN2P3 and Université de Paris Sud, 91898 Orsay, France;
[9] ESPCI, 75005 Paris, France;
[10] INFN Sez. di Perugia and/or Università di Perugia, 06123 Perugia, Italy;
[11] INFN Sez. di Pisa and/or Università di Pisa, 56127 Pisa, Italy;
[12] INFN Sez. di Roma and/or Università di Roma "La Sapienza", 00185 Roma, Italy.

Abstract. We propose a new method for the detection of spectral lines in random noise. It mimics the processing scheme of matching filtering i.e., a whitening procedure combined with the measurement of the correlation between the data and a template. Thanks to the original noise spectrum estimate used in the whitening procedure, the algorithm can easily be tuned to the various types of noise. It can thus be applied to the data taken from a wide class of sensors. This versatility and its small computational cost make this method particularly well suited for real-time monitoring in gravitational wave experiments. We show the results of its application to Virgo C4 commissioning data.

Persistent and narrow spectral peaks, known as *lines* are a typical feature of the data from gravitational wave interferometric detectors (ITF). Lines can originate from the ITF functioning (e.g., mirror or suspension resonant modes) and operation (e.g., calibration lines) or from environmental perturbations (e.g., vacuum pumps). Detecting the presence of such lines in ITF readout and control channels is very important for the studying and monitoring of the detector performance. In this paper we present a simple and versatile processing tool to address this issue.

Optimal strategies can be designed for the detection of sinusoidal signals in random noise [1]. However, they are computationally expensive and thus not very well-suited for monitoring purposes. In addition, the amplitude of the lines we are dealing with, are usually large as compared to the noise level. A statistic which is sub-optimal, but computationally acceptable, suffices in this context. Contrarily to [2, 3], we are not interested here in the removal of the line once detected. Though estimates of the line characteristics (such as the central frequency, amplitude and width) can be obtained from the proposed tool, we don't discuss this point here and concentrate on the detection issue.

In Sect. 1, we give a general description of the line detection problem and point out its main difficulties. In Sect. 2, we detail the proposed line search method. In Sect. 3, we give several rules of thumb for adequately setting the free parameters of the line search. In Sect. 4, we present the results of its application to Virgo C4 commissioning data.

1. A tricky detection problem

ITFs are sophisticated apparatus controlled by a complex network of feedback loops. One way of locating the origin of a line in such a system is to check for the coincident occurrence of the line in channels including the ITF readouts, the ITF controls and the environmental sensors.

The line search has thus to be applied to many different channels. Generally, the spectral lines are superimposed to a broadband random noise, hereafter referred to as "background" noise. The power spectral density (PSD) of the background noise can have very different shapes. In most cases, we only have a rough idea of the PSD shape

which may also change with time. Therefore, the problem to address is the statistical detection of a signal (lines) in a random noise of unknown PSD which thus needs to be estimated from the data. This turns out to be difficult because lines and background noise mix at all times i.e., there is no "background noise only" data. The proposed method tackles with this difficulty.

2. A simple line detection algorithm

The basic ingredient of the proposed method is the *spectrogram*. Let us assume that the data $x(t+t_s n), n \geq 0$ are sampled at Nyquist rate $f_s = 1/t_s$ and collected during the time period from t to $t+\mathcal{T}$. We divide this time period into $\mathcal{N}$ non-overlapping intervals of equal duration $\mathcal{T}/\mathcal{N} = t_s N$ (which thus contain the same number of samples N). The spectrogram $S(t,f)$ is defined for $f \in [0, f_s/2]$ by [4]

$$S^2(t,f) \equiv \frac{1}{\mathcal{N}} \sum_{k=0}^{\mathcal{N}-1} \left| t_s \sum_{j=kN}^{(k+1)N-1} x(t+t_s j)\, h(t_k - t_s j)\, e^{-2\pi i\, t_s j f} \right|^2 ,$$

where $t_k \equiv t_s(kN + (N-1)/2)$ is the center of the kth interval and $h(t)$ is a window function (e.g., Hanning type) centered around $t=0$ and scaled to unit L^2 norm.

Matched filtering [5] is an efficient method for detecting deterministic signals in random noise. It can be viewed as a two-step process: a whitening of the data followed by a scalar product with a template. The line search algorithm mimics this structure. We assume that $S(t,f)$ has been computed for some given time t and we detail now the detection procedure.

step 1. background PSD estimate

We use $S(t,f)$ to get a robust estimate of the background PSD: (1) the frequency axis $f \in [0, f_s/2]$ is tiled into intervals. Their size is chosen sufficiently small that each of them contains only a few lines ($\lesssim 5$) and that the background PSD can be considered almost linear within the interval. (2) In each interval, we remove the N_q points with the largest amplitude (these "outliers" are essentially corresponding to the few frequency peaks belonging to the interval). (3) We make a least mean square linear fit of the remaining spectrogram data points. The collection of the fits performed in all frequency intervals yields the estimate of the background noise PSD, $\hat{S}(t,f)$. Clearly, step (2) prevents that the lines (of large amplitude) bias this estimate.

step 2. whitening and scalar product with template

Roughly speaking, the lines can be described by the general model $a(t)\cos\varphi(t)$ where the envelope $a(t)$ is slowly varying and the phase $\varphi(t)$ is well approximated by a linear function for time periods of duration not smaller than $\mathcal{T}$.

The template matching to a line of duration $\mathcal{T}$ with a constant envelope $a(t) = C^{st}$ and a constant frequency $f_0 = (2\pi)^{-1} d\varphi/dt$ is a "simple" cosine function of the same duration. The FFT (of time base $\mathcal{T}$) does exactly the scalar product with this template. The above definition of the spectrogram is not identical to this (because of the division of $\mathcal{T}$ into intervals and the modulus averaging) but it is similar. This suggests to use $S(t,f)$ in place of the exact template match. We consider the following statistic $W(t,f) \equiv S(t,f)/\hat{S}(t,f)$ which includes both whitening and template match.

step 3. detection and post-processing

The line detection is then made by comparing the statistic to a threshold: $W(t,f) > \eta$.

Once the three steps are completed, we restart the procedure from step 1 to process the spectrogram computed over the next data chunk.

3. Choosing the parameters

Some tuning of the free parameters is required to ensure that the procedure works properly. There is a total of six free parameters. The computation of $S(t,f)$ uses two of them: the total observation time $\mathcal{T}$ (typically $\mathcal{T} \approx 300s$) and the FFT time base $N \equiv \mathcal{T} f_s/\mathcal{N}$ (set to a power of 2 for efficiency) which also sets the frequency resolution (the orders of magnitude are: $f_s \approx 10$ kHz and we choose $N \sim 10^5$ data points corresponding to a frequency resolution of 100 mHz).

The PSD fit requires a value for N_q. A good value is given by the mean number of lines per frequency interval times the peak average width (at half height, expressed in bin).

The threshold η determines the minimum detectable signal-to-noise ratio (or, more specifically, the ratio of the line amplitude to the neighboring background noise level) and at the same time, the rate of false alarms. Typically, we choose $\eta \simeq 3$ to 4.

The two remaining parameters are related to the tiling of the frequency axis mentioned in Sect. 2. We discuss the question of choosing their value in the next two sections.

3.1. Tiling types

The estimate of the background noise in step 1 of our algorithm relies on a tiling of the frequency axis. In principle, any tiling method could be used but for practical reasons (simple coding) we restrict the choice to the ones described in this section.

The FFT algorithm samples uniformly the frequency axis according to $f_k = f_s k/N$ for $k = 0, \ldots, N/2$. We propose to divide this discrete axis in non-overlapping intervals. These intervals are organized by successive blocks of intervals of equal size. Each block is identified by the index $j \in \{-1, 0, \ldots, j_{\max}\}$ and it is made of N_b contiguous intervals of size M_j. Each interval $\mathcal{I}_{i,j}$ is labelled by $i \in \{0, \ldots, N_b - 1\}$ and the block index j. We set $\mathcal{I}_{i,j} = \{k_{i,j}, \ldots, k_{i,j} + M_j - 1\}$ with $k_{i,-1} = iN_n$ and $M_{-1} = N_n$ for $j = -1$, and $k_{i,j} = 2^j(N_b + i)N_n$ and $M_j = 2^j N_n$ for $j \geq 0$. Therefore, the tiling is completely determined by the two parameters N_n and N_b, respectively the smallest interval size (in bins) and the number of intervals within a block. Since N is a power of 2, N_n and N_b must also be so, and we have the condition $N_n N_b \leq N/2$. We also have $j_{\max} = \log_2(N/(2N_n N_b)) - 1$.

Two types of tiling are thus defined, depending on the product $N_n N_b$, as exemplified in Figure 1. If $N_n N_b < N/2$, we obtain a *dyadic* tiling for which the size of the intervals is increasing with the frequency. If instead $N_b N_n = N/2$, the frequency axis is tiled *uniformly* (all intervals are of size N_n).

3.2. Match the tiling to the data

The performance of the algorithm depends critically on the accuracy of the estimate $\hat{S}(t,f)$ of the background noise PSD, which itself depends on the selected tiling. The tiling choice results from a trade-off balancing the two following competing factors:

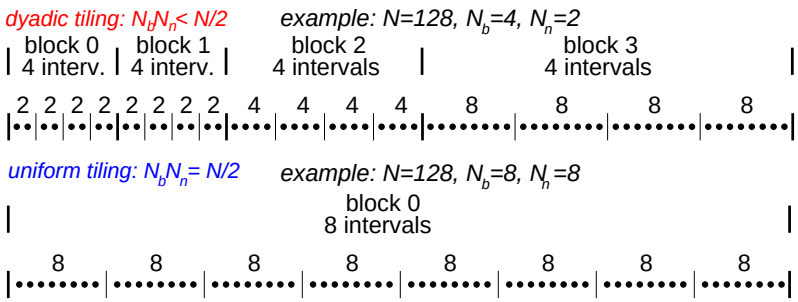

Figure 1. Sketch view of two examples of frequency axis tiling.

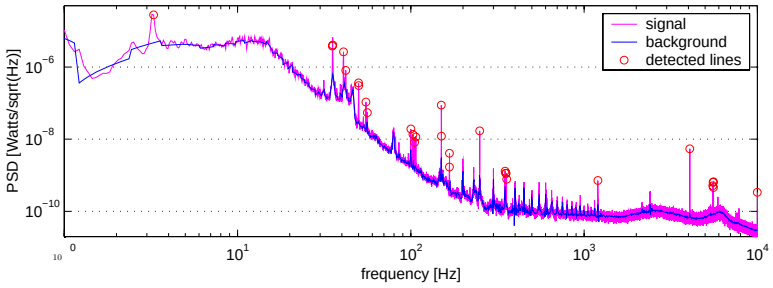

Figure 2. Result of the line search on the dark fringe photo-diode signal of Virgo C4 commissioning run ($f_s = 20$ kHz). The parameter set is: $\mathcal{N} = 20$, $N = 262144$ i.e., total observation time $\mathcal{T} \simeq 260s$, $N_b = 8$, $N_n = 64$ i.e., dyadic tiling, $N_q = 32$ and the detection threshold $\eta = 4$.

long frequency intervals provide a better fit (because of the larger number of data points), whereas short intervals allow to better follow curved (non-flat) PSD.

The main readout channel of the ITF i.e., the dark fringe (DF) signal is of special importance, and deserves a specific attention. The noise PSD observed in the DF signal has a characteristic shape: it is curved at low frequency (where the number of lines is also larger) and it becomes almost flat at high frequencies. Consequently, we need small intervals at low frequencies and large intervals at high frequencies. The dyadic tiling is thus preferred for the DF channel as illustrated in Figure 2.

A different case is exemplified by a signal from a seismometer in Figure 3. The PSD of the background noise for such signal is much flatter than DF (compare the respective dynamical ranges) and its curvature is not localized specifically at low frequency. The uniform tiling is thus the right choice.

In summary, the tiling type should be chosen according to the general shape of the background PSD. However, for a given tiling type (dyadic or uniform), the size of the typical interval is also important. The fit made with intervals of small sizes are able to follow PSDs which are curved locally as illustrated in Figure 4. If the size is too large, $\hat{S}(t, f)$ may not fit well the data and this causes an increase of the rate of false alarms.

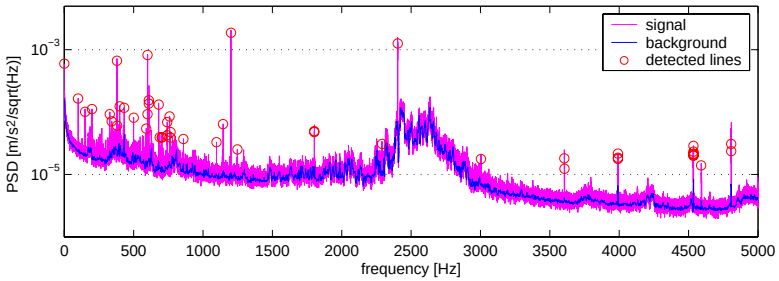

Figure 3. Result of the line search on one seismometer signal (f_s = 10 kHz). The parameter set is: $\mathcal{N} = 20$, $N = 131072$ i.e., total observation time $\mathcal{T} \simeq 260s$) $N_b = 1024$, $N_n = 64$ i.e., uniform tiling $N_q = 32$ and the detection threshold $\eta = 4$.

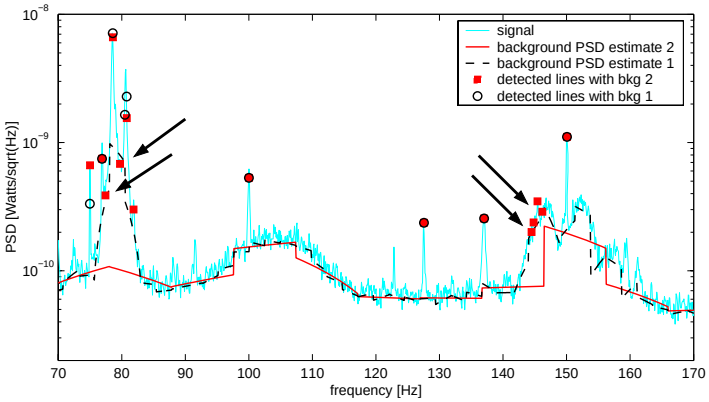

Figure 4. Line search on the photo-diode signal used in C4 to control the beam splitter mirror position. Two choices of the tiling size are made: 1) $N = 262144$, $N_b = 32$, $N_n = 32$, $N_q = 16$, $\eta = 3$ (dashed line and empty circles); 2) $N = 262144$, $N_b = 32$, $N_n = 128$, $N_q = 64$, $\eta = 3$ (continuous line and filled squares). The coarser choice (2) has difficulty to follow the curved shape of the background PSD. This causes false alarms which are indicated by arrows.

4. Line monitoring in Virgo data

We used the proposed algorithm to monitor the lines of the Virgo DF in the C4 commissioning run data (a total of 120 hours). The line search was applied over $\mathcal{T} \simeq 260$ s long data chunks with a FFT time base $Nt_s \simeq 13$ s (which corresponds to $N = 262144$ and a frequency resolution $\simeq 0.07$ Hz). We chose the tiling parameters $N_b = 8$ and $N_n = 64$ (dyadic tiling), a rejection parameter $N_q = 32$ and a threshold $\eta = 4$.

A number of 40 lines were detected during the entire run. (This number is roughly constant when ITF is in stable operation). An overview is presented in the time-frequency diagram in Figure 5 where 1 dot is 1 detected line. Isolated dots are probably false alarms, vertical stripes correspond to ITF losses of lock, blank regions correspond to unlocked periods. We see that some lines appear non-stationary.

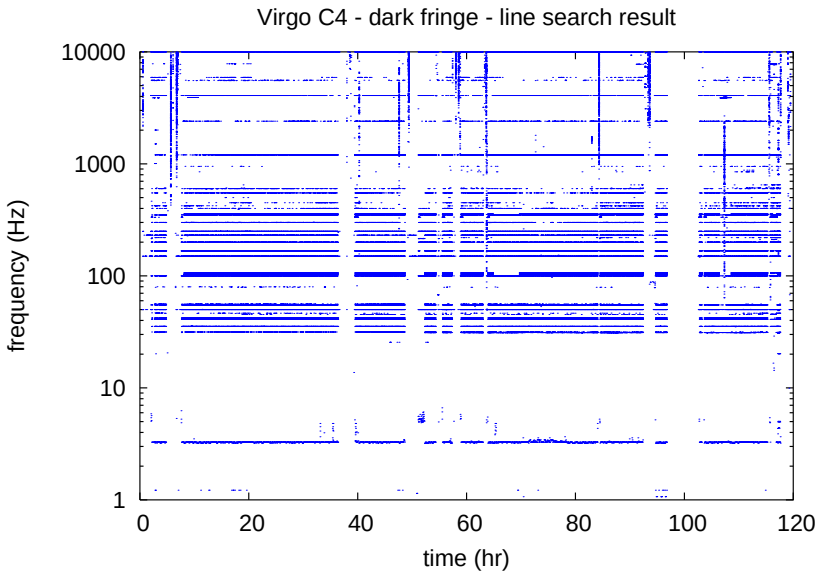

Figure 5. Lines detected in the Virgo DF during the entire C4 run (1 blue dot = 1 detection).

Freq. [Hz]	*explanation*	*Freq.* [Hz]	*explanation*
0.4	mirror suspensions θ_x mode	353.0	calibration line
3.2	mirror suspensions θ_x mode	355.0	calibration line
31.5	IB susp. vertical mode	357.0	calibration line
35.2	IB susp. θ_x mode	600.8	IB tower vacuum pump
40.9	IB susp. θ_z mode	1201.6	IB tower v. p. (1^{st} harmonic)
55.0	IB wires violin mode	2402.4	IB tower v. p. (2^{nd} harmonic)
56.1	IB wires violin mode	5543.9	NE mirror 1^{st} symmetric mode
103.0	calibration line	5544.8	WE mirror 1^{st} symmetric mode
105.0	calibration line	5583.9	WI mirror 1^{st} symmetric mode
107.0	calibration line	5585.8	NI mirror 1^{st} symmetric mode
230.8	water chiller pump		

Table 1. Subset of 21 lines detected in Virgo C4 DF which have been identified (acronyms: IB ≡ Input Bench, NI ≡ North Input, WI ≡ West Input, NE ≡ North End, WE ≡ West End.)

We extracted a subset of 38 persistent lines i.e., lasting at least 3% of the whole run duration. Among these: 11 are 50 Hz harmonics, 21 have been associated to mirror internal modes, mechanical resonances of the mirror suspensions, vibrations of vacuum and water pumps, and 6 have not been identified yet. Identified lines are listed in Table 1.

Conclusions

We propose a robust and versatile line detection algorithm and demonstrate its validity on Virgo C4 commissioning data. The computational resources required are reasonably cheap such that the processing pipeline is feasible in real-time. We are presently studying the possibility to use this algorithm as the detection engine of a "line monitor" for Virgo, and eventually interface it to a database which would help to track the history of the detected lines.

References

[1] B. Allen, M. A. Papa, and B. F. Schutz. Optimal strategies for sinusoidal signal detection. *Phys. Rev.*, D66:102003, 2002.

[2] A. C. Searle, S. M. Scott, and D. E. McClelland. Spectral line removal in the LIGO Data Analysis System (LDAS). *Class. Quantum Grav.*, 20:S721–S730, 2003. Proc. of the 7th Grav. Wave Data Anal. Workshop (Kyoto, Japan).

[3] A. M. Sintes and B. F. Schutz. Coherent line removal: Filtering out harmonically related line interference from experimental data, with application to gravitational wave detectors. *Phys. Rev.*, D58:122003, 1998.

[4] The Virgo Collaboration (corresponding authors: G. M. Guidi and É. Chassande-Mottin). Data analysis methods for non-Gaussian, non-stationary and non-linear features and their application to Virgo. *Class. Quantum Grav.*, 20:S915–S924, 2003. Proc. of the 7th Grav. Wave Data Anal. Workshop (Kyoto, Japan).

[5] S. Kay. *Fundamentals of Statistical Signal Processing: Detection theory.* Prentice Hall, New Jersey (US), 1998.