



## **Iterative baseline correction algorithm for dead time truncated one- dimensional solid-state MAS NMR spectra**

Maxime Yon, Franck Fayon, Dominique Massiot, Vincent Sarou-Kanian

### **► To cite this version:**

Maxime Yon, Franck Fayon, Dominique Massiot, Vincent Sarou-Kanian. Iterative baseline correction algorithm for dead time truncated one- dimensional solid-state MAS NMR spectra. Solid State Nuclear Magnetic Resonance, 2020, pp.101699. <10.1016/j.ssnmr.2020.101699>. <hal-02987857>

**HAL Id: hal-02987857**

**<https://hal.science/hal-02987857v1>**

Submitted on 4 Nov 2020

**HAL** is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.



HAL Authorization

# Iterative baseline correction algorithm for dead time truncated one-dimensional solid-state MAS NMR spectra

Maxime Yon<sup>1</sup>, Franck Fayon<sup>1</sup>, Dominique Massiot<sup>1</sup>, Vincent Sarou-Kanian<sup>1</sup>

<sup>1</sup>CNRS, CEMHTI UPR3079, Université d'Orléans, F-45071, Orléans, France.

## Abstract

We present an algorithm suitable for automatically correcting rolling baseline coming from time-domain truncation induced by the dead time in pulse-acquire one-dimensional MAS NMR spectra. It relies on an iterative estimation of the baseline restricted in the time-domain by the dead time duration combined with a histogram filter allowing adaptive selection of the baseline points. This method does not make any assumption regarding the NMR resonances line shapes or widths and does not modify the acquired free induction decay points. This makes it suitable for accurate deconvolution and quantification of single-pulse MAS NMR spectra. The baseline correction accuracy is evaluated on synthetic solid-state spectra of <sup>19</sup>F, <sup>71</sup>Ga, and <sup>23</sup>Na by comparing the fitted baseline to the theoretical one. The versatility of the algorithm is also exemplified on three additional solid-state spectra of <sup>23</sup>Na and <sup>71</sup>Ga. The algorithm is made available to the community through a user-friendly standalone Matlab® application.

## Introduction

Magic angle spinning NMR spectroscopy is a well-established analytical method in solid-state chemistry and material science allowing to study the chemical environment of different nuclei in a wide range of samples. The simple pulse-acquire experiment remains the cornerstone of a large part of NMR analyses as it minimizes the line shape distortion due to transverse relaxation and allows to record quantitative spectra. In the case of observation of quadrupolar nuclei, this sequence is often used with large acquisition spectral width to record both the central transition (CT) and the spinning sidebands of the satellite transitions (ST) which are spread over a large frequency range by the first-order quadrupolar interaction. Indeed acquiring both central and satellite transitions is useful to determine accurately the quadrupole coupling constant ( $C_q$ ) and the asymmetry parameter ( $\eta_q$ ), especially when they are too small to be measured from the second-order quadrupolar line shape of the central transition.<sup>1,2</sup>

However, the pulse-acquire sequence suffers from one major drawback as the dead time between the end of the radiofrequency pulse and the beginning of the actual acquisition prohibits the recording of the first points of the free induction decay (FID).<sup>3</sup> The restoration of these first missing points in the time domain is required to obtain, by Fourier transformation (FT), a pure absorption spectrum with a flat baseline used for accurate integration or fitting of the NMR lines. This restoration is usually performed by a first-order phase correction in the frequency domain (analogous to a shift in the time

domain) followed by a baseline correction of the resulting rolling sinc convolution<sup>4,5</sup>. The amplitude of the baseline distortion is proportional to the number of missing points (NMP) and thus the ratio between dead time (DE) and dwell time (DW). As the minimum DE is fixed by the hardware, the baseline distortion gets stronger and more difficult to correct with the increase of the acquisition spectral width. Moreover, additional baseline distortions can also arise from various sources such as acoustic ringing<sup>6</sup>, the discrete nature of the Fourier transformation<sup>7</sup>, or low-quality filters<sup>7</sup>. Unlike the dead time truncation, these latter distortions induce signal contamination of the FID points.

Various algorithms either based on frequency domain correction<sup>8–13</sup> or time-domain backprediction<sup>14,15</sup> and software<sup>16–18</sup> have been proposed to perform automatic baseline correction. However, these algorithms often fail in the case of significant time-domain truncation and wide ranges of shapes and intensities of the NMR lines common in solid-state MAS spectra. The most common solution for the baseline correction of solid-state 1D spectra acquired with a large spectral width remains to use a spline correction based on user-defined baseline points,<sup>19</sup> however this process is strongly user-dependent, lack of reproducibility and can introduce biases.

In this article, we introduce an automatic baseline correction algorithm for 1D NMR spectra affected by dead time truncation. The algorithm relies on an iterative smoothing spline fit constrained by the dead time duration and a histogram filter used to discard the high-intensity points from the baseline fit. This allows obtaining a baseline correction that does not modify the acquired points of the FID which is the main disadvantage of the baseline correction through model-free smooth fitting<sup>3</sup>. The model does not require any user inputs when the baseline deformation is due to time-domain truncation only. Nevertheless, if the FID is contaminated by other sources of baseline distortion (ringing effects...), the user is free to remove or adjust the time-domain constraint and optimize the smoothing parameter of the fitting function or the NMP to obtain a meaningful baseline correction.

This method does not aim to compute the mathematically exact baseline modulation but to allow an easy, rapid, and reproducible baseline correction. The quality and the interpretation of the baseline correction remain at the judgment of the user.

## Baseline correction algorithm

### Theoretical considerations

The algorithm is based on both the invertibility and the linearity properties of the Fourier transformation that, in case of 1D NMR acquisition, can be illustrated by the following equation:

$$f(t) = f_1(t) + f_2(t) \xrightarrow{FT} S(\nu) = S_1(\nu) + S_2(\nu) \xrightarrow{IFT} f(t) = f_1(t) + f_2(t) \quad (1)$$

With  $f(t)$  the ideal FID,  $f_1(t)$  the FID experimentally acquired after the dead time duration and  $f_2(t)$  the part of the FID that could not be experimentally recorded (Figure 1, left). The FT of these signals

leads to the spectra  $S(\nu)$ ,  $S_1(\nu)$  and  $S_2(\nu)$ , respectively (Figure 1, right).  $S_1(\nu)$  resembles an experimental spectrum processed with first-order phase correction. Here, the zero filling to  $t = 0$  of the  $f_1$  signal avoids the need for first-order phase correction. The shape of  $S_2(\nu)$  corresponds to the convolution of the spectrum  $S(\nu)$  and a sinc function whose periodicity is equal to the inverse of the dead time. Finally, the addition of  $S_1(\nu)$  and  $S_2(\nu)$  allows retrieving  $S(\nu)$  and  $-S_2(\nu)$  *de facto* corresponds to the so-called baseline distortion.

These considerations imply that the correction of a rolling baseline purely induced by a time-domain truncation of the beginning of the FID must originate from a non-zero signal strictly limited in the time-domain by the dead time duration. This baseline correction must not induce any modification of the acquired FID points.

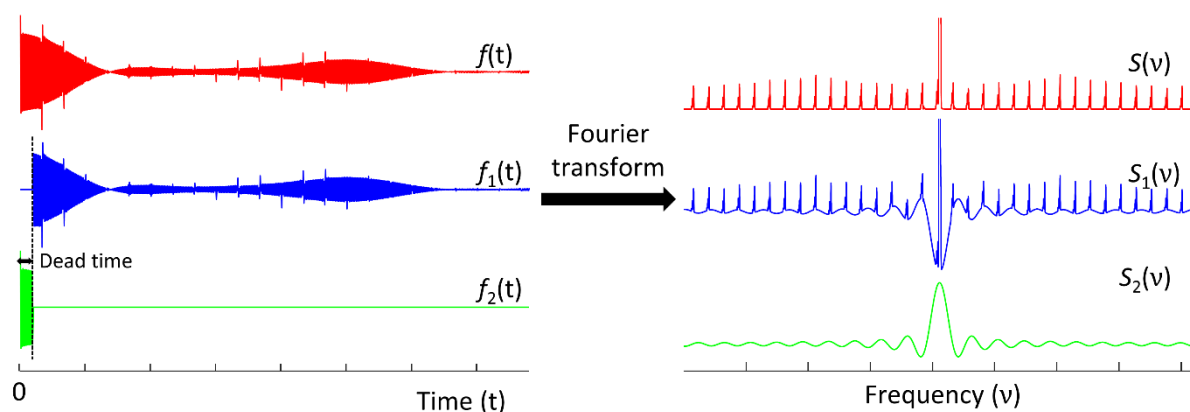


Figure 1: (Left) numeric models of a typical NMR free induction decay signal: in red  $f(t)$  the full ideal signal, in blue  $f_1(t)$  the signal acquired experimentally after the dead time and in green  $f_2(t)$  the part of the FID that could not be experimentally recorded. (Right) the three corresponding spectra obtained by Fourier transformation of the FID signals.

### The algorithm flowchart

The flowchart of the baseline correction protocol is presented in Figure 2 exemplifying its different steps, which are detailed in the following subsections.

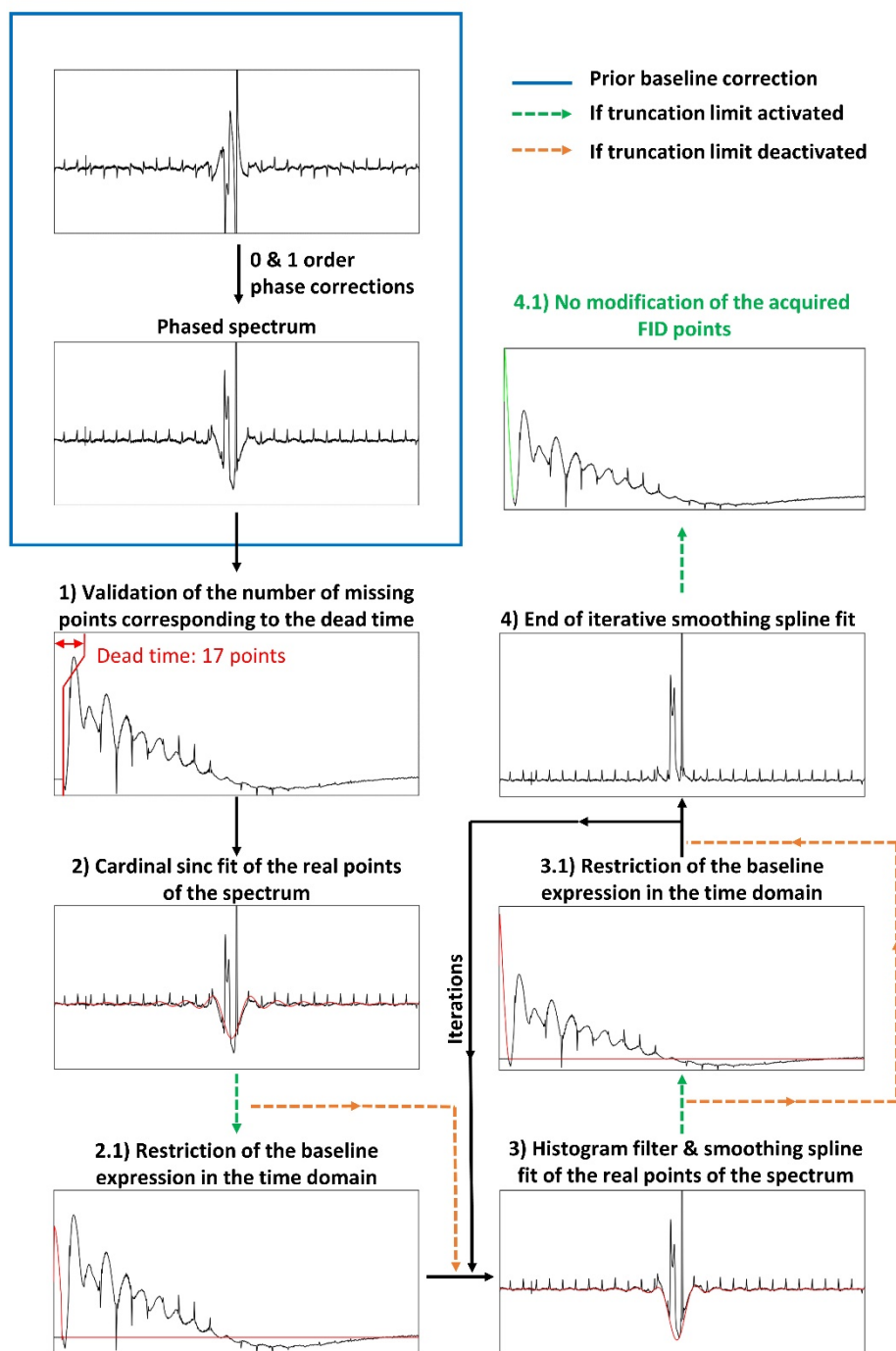


Figure 2: Flowchart of the baseline correction protocol for 1D NMR spectra. The zero and first-order phase corrections in the blue rectangle have to be performed before the baseline correction. The green arrows indicate the baseline protocol when the truncation limit option is activated. The orange arrows are used when the truncation limit option is deactivated. In the restriction steps (2.1 and 3.1), the inverse Fourier transform points of the baseline fit (in red) at time superior to the dead time are set to zero. In panel 4.1, the time domain signal due to the baseline correction is represented in green, the rest of the points are not modified.

### Preliminary requirements

The baseline correction procedure is dedicated to phased spectra, it is then mandatory to perform zero- and first-order phase corrections before the baseline correction as illustrated in the blue rectangle in

Figure 2. The algorithm requires the number of missing points (NMP) at the beginning of the FID. In practice, the interface load only the real part of the spectrum and its inverse Fourier transform resemble a symmetric echo where the NMP need to be even and its default value is calculated by  $NMP = (\text{round}((DE/DW)/2)) \cdot 2$  with DW following the convention of  $DW = 1/(2 \cdot swh)$  with  $swh$  the acquisition spectral width. In the case of the Bruker data format, the values of DE and DW are retrieved automatically and displayed in  $\mu s$  while for spectrum imported as simple two-column ASCII file, the NMP value needs to be set manually. The application of additional digital filters can also slightly modify the value of NMP. In such cases, the overall dead time duration can be determined from the very low signal intensity region of the inverse Fourier transform of the spectrum as illustrated in figure 2 step 1 and the value of NMP can be suitably adjusted. The NMP value is only used when the truncation limit option is activated which is the default setting. In this case, all the baseline estimations obtained by the successive fits will undergo an inverse Fourier transformation and their time-domain points of coordinate superior to NMP will be set to zero enforcing the rule stated in theoretical considerations. This step avoids any modification of the acquired FID points and is suitable if the baseline roll is induced by the dead time truncation only. Conversely, when the FID points are contaminated by other sources of baseline distortion, the user is free to deactivate the truncation limit or virtually increase the value of NMP, which will remove or decrease this time-domain constraint. The deactivation or virtual modification of the NMP will make the correction parametric and the user will need to optimize either the smoothing parameter of the smoothing spline fit function or the NMP. The influence of these two options will be discussed in the results.

### Baseline detection

The successive fits performed in the frequency domain required to differentiate the data points corresponding to the baseline from those belonging to the NMR resonances. The identification of the baseline points is performed in two steps: a simple local minimum detection and a thresholding operation. These operations do not aim at precisely identifying the baseline points, which would be difficult without making any assumption regarding the shapes, intensities, or number of NMR lines, but only to remove the majority of the positive high-intensity points of the resonances. The local minimum detection is performed on a spectrum smoothed by a moving average lowpass filter (span = number of spectrum points /1000) to reduce the number of baseline points and increase the fitting speed. The thresholding operation is based on the histogram of the real points of the spectrum, which allows it to be adaptive. Indeed, the flatter the baseline, the weaker its effect on the width of the low intensity points distribution. This reduction induces a decrease of the threshold value and thus the elimination of less and less intense resonances points. Figure S1 illustrates this effect. This operation assumes that the spectrum is sparse: the highest number of points per intensity range corresponds to the noise points. The threshold value  $TV$  is calculated with:

$$TV = Max_{hist} + FWHM_{hist} \quad (2)$$

with  $FWHM_{hist}$  the full width at half maximum of the histogram and  $Max_{hist}$  the maximum of the histogram. This threshold value was found to be a good compromise between robustness and selectivity on multiple tests with NMR spectra exhibiting quadrupolar line shapes, large CSA, crystalline or disordered environments, and various SNR. This filter is increasingly selective as the baseline gets better (Fig. S1).

### Sinc baseline fitting

A sinc function fit (Figure 2 – step 2) can be performed to correct the major deviations of the baseline allowing to increase the selectivity of the threshold filter as well as making the iterative fit converge faster. This approach results from the relation between the experimental spectrum and the convolution of the ideal spectrum with a sinc function for which the periodicity is the inverse of the dead time. It is generally suitable for spectra dominated by few central intense resonances such as the half-integer spin quadrupolar nuclei spectra containing few intense CT resonances and numerous low-intensity ST resonances. This step can be skipped if not relevant for the baseline correction.

### Iterative smoothing spline fit

The main baseline fit (Figure 2 – step 3) is conducted with the Matlab smoothing spline fit function which estimates  $\hat{S}(v)$  by minimizing the following equation:

$$p \sum_i w_i (S_i - \hat{S}(v_i))^2 + (1 - p) \int \left( \frac{\partial^2 \hat{S}}{\partial^2 v} \right)^2 dx \quad (3)$$

with  $p$  the smoothing parameter,  $S_i$  and  $w_i$  the intensity and the weight of the  $i$ -th baseline point, respectively. The smoothing parameter  $p$  ranges between 0 and 1, such that the spline function tends to a straight line when  $p \rightarrow 0$ , while it tends to fit perfectly the baseline points when  $p \rightarrow 1$ . The use of a small  $p$  value induces an important constraint leading to the slow evolution of the recursive process. However, the iterative (recursive) loop ensures that when the truncation limit option is activated, the time domain restriction is the main constraint on the baseline fit. The smoothing parameter  $p$  is then set to a default value of  $p = 1/N^2$  with  $N$ , the number of spectrum points, allowing to avoid overfitting of the baseline points which would be entirely discarded by the time domain constraint. The value  $-\log_{10}(p)$  is displayed in the user interface to allow manual modification of this parameter required when the truncation limit is deactivated. During this fit, the intensity of each baseline point ( $S_i$ ) is weighted by  $w_i$ . This weighting is used to overweight the baseline points near the intense resonances that induce strong baseline distortions. The weighting values  $w_i$  are computed by applying an important Gaussian broadening on the spectrum to obtain the spectrum  $S_g$  (Fig. S2) and use its intensity to compute  $w_i$  normalized between 1 and 10 according to:

$$w_i = 9 \frac{|S_{g_i}|}{\max(S_g)} + 1 \quad (4)$$

## End of fit and correction of the imaginary part

The iterative fit automatically ends when the normalized root mean squared error between the iterative fit  $\hat{S}(\nu)$  and a flat line reaches  $10^{-6}$  according to the equation:

$$nRMSE = \frac{\sum_{i=1}^N |S_i|}{N \cdot \max(\text{spectrum})} < 10^{-6} \quad (5)$$

With  $S_i$  the fit value of the  $i$ -th point. The user can modify the value of this stopping criterion if needed. The baseline-corrected spectrum can be finally saved in the format of importation. The imaginary part of the spectrum is also corrected by using a Hilbert transform to bring it in absorption mode, it is then corrected with the same baseline and brought back in dispersion mode by a second Hilbert transform.

## Methodology

The performances of the baseline correction algorithm toward various SNR values and dead time durations were evaluated on three model spectra synthesized with Dmfit<sup>19</sup> and representative of typical cases encountered in solid-state NMR experiments. Their NMR parameters are derived from experimental data.

- Cryst- $\text{NaH}_2\text{PO}_4$ : a  $^{23}\text{Na}$  ( $I = 3/2$ ) spectrum of a crystalline sample ( $\text{NaH}_2\text{PO}_4$ ) with two crystallographic sites at  $\delta_{\text{iso}} = -2.1$  and  $-6$  ppm with quadrupolar coupling constants  $C_q$  of 1.57 and 2.36 MHz respectively. The spectrum was simulated for a magnetic field  $B_0 = 9.4$  T, a MAS frequency  $\nu_{\text{MAS}} = 25$  kHz, and a spectral width of 833.333 kHz ( $\text{DW} = 0.6 \mu\text{s}$ ).
- Glass- $\text{CaAl}_2\text{O}_4$ : a  $^{27}\text{Al}$  ( $I = 5/2$ ) spectrum of a structurally disordered sample ( $\text{CaAl}_2\text{O}_4$ ) exhibiting  $\text{AlO}_4$ ,  $\text{AlO}_5$ ,  $\text{AlO}_6$  structural units with characteristic Czjzek distributions and proportions of 84.8, 12.9, and 2.3 % respectively. The spectrum was simulated for  $B_0 = 17.6$  T,  $\nu_{\text{MAS}} = 33.35$  kHz, and a spectral width of 2 MHz ( $\text{DW} = 0.25 \mu\text{s}$ ).
- Cryst- $\text{ZrF}_4$ : a  $^{19}\text{F}$  ( $I = 1/2$ ) spectrum of a crystalline sample ( $\beta\text{-ZrF}_4$ ) with seven crystallographic sites with relative proportion of 16.05, 9.64, 18.31, 15.57, 14.95, 6.6 and 18.88 % for site 1 to 7 respectively and exhibiting different chemical shift anisotropies. The spectrum was simulated for a  $B_0 = 17.6$  T,  $\nu_{\text{MAS}} = 35$  kHz, and a spectral width of 500 kHz ( $\text{DW} = 1 \mu\text{s}$ ).

The use of synthetic spectra allows comparing the baseline obtained by the algorithm to the theoretical one obtained by the subtraction of the truncated and the theoretical (non-truncated) spectra. This comparison allows computing the goodness of fit (GoF), giving a quantitative metric to evaluate the baseline correction accuracy. The goodness of fit is calculated by:

$$GoF = 1 - \frac{\sqrt{\sum_{i=1}^N |S_i - S_i^{\text{ref}}|^2}}{\sqrt{\sum_{i=1}^N |S_i^{\text{ref}} - \overline{S^{\text{ref}}}|^2} \cdot \max(S)} \quad (6)$$

With  $S_i^{\text{ref}}$  the theoretical baseline and  $S_i$  the fitted baseline.



The effect of the baseline correction on the quantification of the different chemical environments was also evaluated by reconstructing the spectra using Dmfit and by calculating the relative quantification errors.

These metrics have been calculated for various SNR: 10000, 200, 100, 50, and 25 (obtained by adding white Gaussian noise to the spectra) and various NMP: 10, 20, and 40. Twenty replicates with different Gaussian white noise were generated for each condition (types, numbers of missing points, and SNR) leading to a total number of spectra of 900 ensuring statistical significance. The spectra with a NMP of 20 and the 5 SNR levels are displayed in Figure 3(a-c).

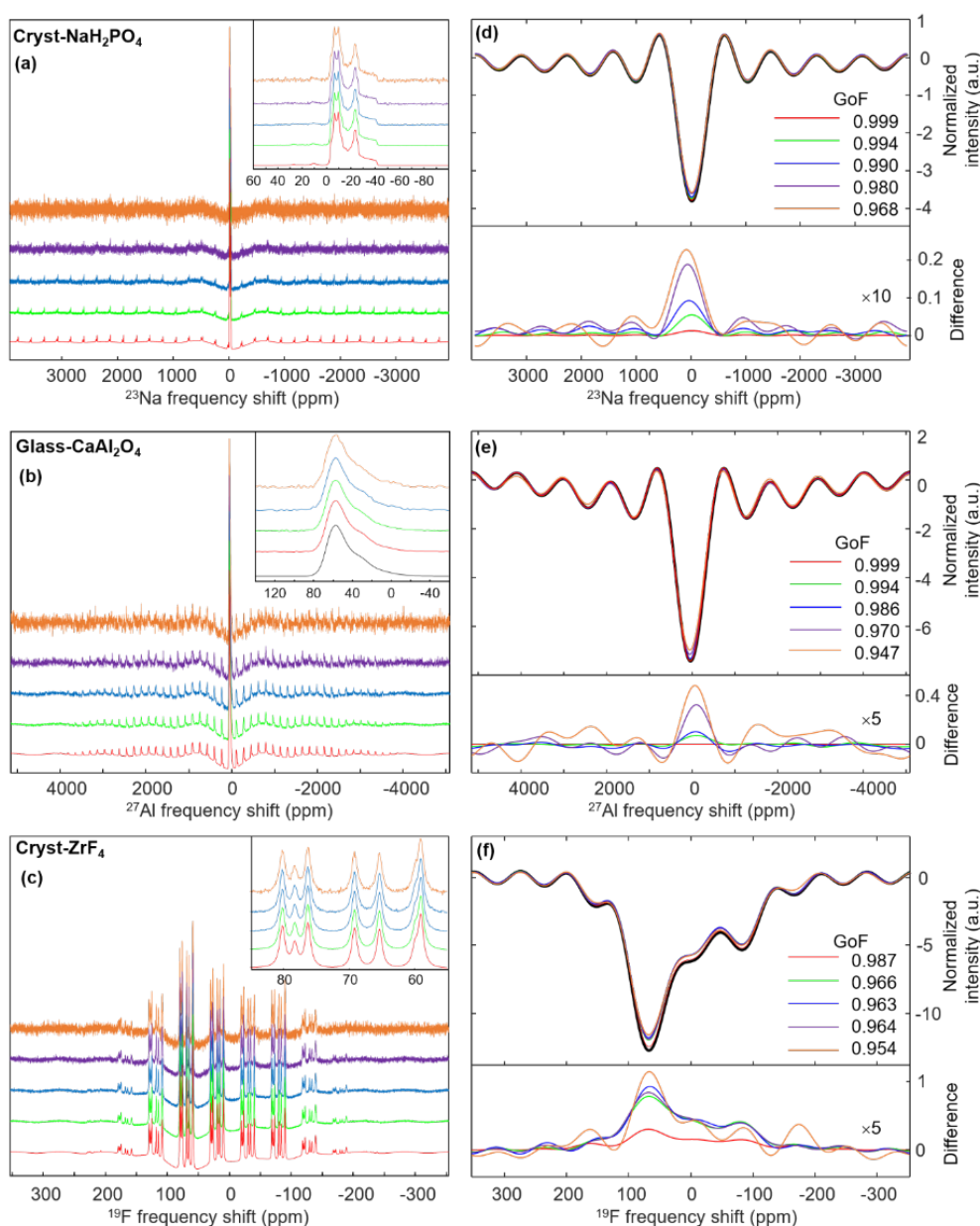


Figure 3: Synthetic NMR spectra truncated with NMP = 20 used to evaluate the baseline correction algorithm and the corresponding mean baselines obtained by automatic fitting. Red, green, blue, purple, and orange spectra have a signal-to-noise ratio (SNR) of 10000, 200, 100, 50, and 25, respectively. (a) Cryst- $\text{NaH}_2\text{PO}_4$ :  $^{23}\text{Na}$  spectra; (b) glass- $\text{CaAl}_2\text{O}_4$ :  $^{27}\text{Al}$  spectra; (c) Cryst- $\text{ZrF}_4$ :  $^{19}\text{F}$  spectra. (d,e,f) Corresponding mean baselines obtained by fitting the various SNR spectra with 20 replicates for each SNR value and default values of the smoothing parameter  $p$  of 9, 8, 8 respectively. The top of each panel shows the mean

fitted baseline and the theoretical baseline (in black) normalized by the spectrum maximum intensity. The corresponding mean GoFs are also displayed on the figure. The bottom of each panel displays the baseline differences between the mean fitted baseline and the theoretical one with an enhancement factor of 10, 5, and 5 for the panel d, e, and f respectively.

## Results

### Evaluation on synthetic spectra

Figure 3 panel d, e, and f display the mean fitted baselines of the 20 replicates for the three simulated spectra with various SNR and 20 missing points. This NMP corresponds to dead time durations of 12  $\mu$ s for Cryst- $\text{NaH}_2\text{PO}_4$ , 5  $\mu$ s for glass- $\text{CaAl}_2\text{O}_4$ , and 20  $\mu$ s for Cryst- $\text{ZrF}_4$ . The baseline errors displayed at the bottom of the three panels are scaled as a percentage of the maximum spectrum intensity, and are lower than 0.5% for Cryst- $\text{NaH}_2\text{PO}_4$  and glass- $\text{CaAl}_2\text{O}_4$ , and reach 1.1% for Cryst- $\text{ZrF}_4$  with the lowest SNR of 25. The highest error values are always localized close to the most intense resonances: the central transition for the quadrupolar nuclei and the most intense spinning sidebands for Cryst- $\text{ZrF}_4$ . The corresponding goodnesses of fit (GoF) are also included in figure 3 and are all above 0.94. It is worth noting that even at low SNR the baseline error remains approximately centered at zero indicating that the baseline remains well defined in the center of the noise distribution. The GoF for all SNR and NMP conditions are summarized in Figure 4 and are always higher than 0.9 and higher than 0.95 for all SNR above 50. The error bars on each value are defined by the standard deviation on the twenty replicates. The span of the error bars indicates the variability of the baseline and decreases when the SNR increases. Despite these variations induced by the noise distribution, the baseline error remains mainly constant for a defined spectrum, NMP and SNR. For the same NMP, the GoF increases with the SNR due to the decrease in noise and thus in uncertainty in the baseline definition. At constant SNR, the GoF decreases when the NMP increases due to the increase of the baseline distortions as shown in Supplementary figure 3, but also due to the increased freedom of the baseline induced by the less restrictive time-domain constraint. The GoFs can be further decreased in case of inaccurate zero-order phase adjustment before baseline correction, however, this additional error is negligible compared to the error induced by the phase bias itself as illustrated in Supplementary figure 4.

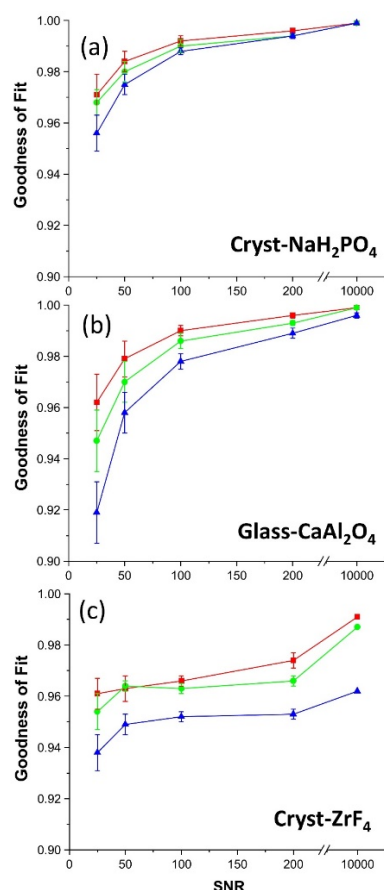


Figure 4: Goodness of fit comparing the fitted baselines with the theoretical ones as a function of the signal-to-noise ratio (SNR) for Cryst- $\text{NaH}_2\text{PO}_4$  (a), glass- $\text{CaAl}_2\text{O}_4$  (b), and Cryst- $\text{ZrF}_4$  (c) spectra. The red, green, and blue lines correspond respectively to NMP= 10, 20, and 40. The error bars correspond to the standard deviation calculated over the twenty replicates.

Although the goodness of fit allows assessing the accuracy of the baseline correction, the end-point user will be more interested in the impact of the baseline correction on the quantification. For this reason, the mean baseline errors for each NMP and SNR were added to the theoretical spectra and each resulting spectra was modeled with Dmfit. The amplitudes of the lines were optimized to get the proportions of each environment and calculate the percentage of integration error due to the baseline correction errors. The percentage of integration error are summarized in supplementary tables 1 to 3. For each type of spectra, the integration error increases when the SNR decreases or when the number of missing points increases, which correlates with the evolution of the GoF values. The integration errors for the two  $^{23}\text{Na}$  sites of Cryst- $\text{NaH}_2\text{PO}_4$  remain inferior to 0.5% confirming the high GoF values and allowing relative quantification with a maximum error of 0.25%. For Cryst- $\text{ZrF}_4$ , the results are similar with a maximum integration error of the 1.32% for a SNR of 25 and 40 missing points leading to a maximum quantification error of 0.25%. Regarding the glass- $\text{CaAl}_2\text{O}_4$  spectra, the error percentage of the integration based on the CT resonances is more important but remains under 4% for the two major environments ( $\text{AlO}_4$  and  $\text{AlO}_5$ ), and reach 58% for the minor one ( $\text{AlO}_6$ ) with a SNR of 25 and NMP=40. This important relative integration error results in two unfavorable conditions. (i) The proportion of this environment is 2.3% making the relative error more sensitive. Indeed, the quantification error on the

proportion of  $\text{AlO}_6$  is only 1.3% which remains satisfying. (ii) The line shape overlaps with the right side long tailing of the two other intense resonances that may make the baseline correction difficult. The baseline correction induces a more important quantification error (1.3%) for the  $^{27}\text{Al}$  spectrum compared to the  $^{23}\text{Na}$  and  $^{71}\text{Ga}$  spectra (0.25%) probably due to the broad, stepless and poorly resolved  $^{27}\text{Al}$  lines characteristic of Czjzek distributions.

## Experimental spectra

The baseline correction algorithm was applied to two practical cases to illustrate different baseline distortion problems. In line with the synthetic spectra previously used, the first example is the experimental  $^{23}\text{Na}$  spectrum of  $\text{NaH}_2\text{PO}_4$  recorded at 9.4 T with a MAS frequency of 25 kHz ( $\text{SNR} \approx 1000$ ). The DW and the DE are 0.6  $\mu\text{s}$  and 10  $\mu\text{s}$ , respectively, such that the NMP is initialized as 18. Here the baseline distortion is purely due to dead time truncation, and the original spectrum (Figure 5, blue line) exhibits an intense baseline distortion due to the intense CT resonances of the two different environments. The baseline correction (red line) obtained by a sinc function fit, followed by the smoothing spline iterative fitting, is excellent. The spectral modeling of the two chemical environments lineshape performed with DMfit<sup>19</sup> and including all transitions of  $^{23}\text{Na}$  are displayed in purple and orange. The resulting spectrum (in green) reveals an excellent agreement with the corrected one for both the central transition (top right panel) and the satellite transition (top left panel) demonstrating that NMR line shapes are perfectly preserved by the baseline correction. In some cases when the spectrum contains numerous very broad low-intensity features, the baseline correction algorithm may introduce small distortions of the line shapes. This is exemplified in Figure S5 with the  $^{71}\text{Ga}$  spectrum of an Yttrium Gallium Garnet sample ( $\text{Y}_3\text{Ga}_5\text{O}_{12}$ ). The small mismatch between the corrected spectrum (in black) and the total deconvolution (in green) in the center of the spectrum is due to an underestimation of the baseline induced by the consideration of the dip between the two central resonance as a baseline point. However, it is unlikely that a manual correction would have given a better result. These first examples show the great efficiency of the baseline correction algorithm when the truncation limit option is activated which allows an automatic easy, rapid, and reproducible baseline correction.

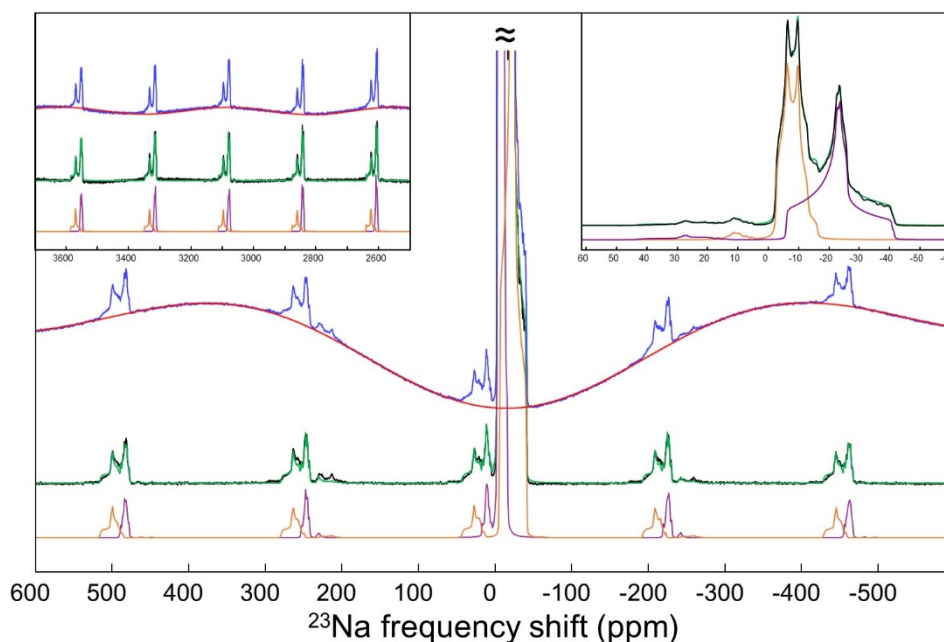


Figure 5:  $^{23}\text{Na}$  spectrum of  $\text{NaH}_2\text{PO}_4$  recorded at 9.4 T at a MAS spinning rate of 25 kHz. The spectrum was acquired with a DW of 0.6  $\mu\text{s}$  and a DE of 10  $\mu\text{s}$ . In blue: original spectrum; in red: the fitted baseline; in black: the corrected spectrum. In green: the full DMfit<sup>19</sup> decomposition obtained with the int2QUAD model including all  $^{23}\text{Na}$  transitions and two chemical environments ( $\nu_{\text{iso}}^{(1)} = -2.1$  ppm,  $C_Q^{(1)} = 1.57$  MHz,  $\eta_Q^{(1)} = 0.47$ ;  $\nu_{\text{iso}}^{(2)} = -6.0$  ppm,  $C_Q^{(2)} = 2.36$  MHz,  $\eta_Q^{(2)} = 0.95$ ) displayed in purple and orange respectively. The total correction was performed in a few tens of seconds on a laptop computer (Intel i7-4800MQ-CPU @ 2.7 MHz with 16 Go of RAM) with the default value of the smoothing parameter  $p$  of 10.

The second example is a  $^{71}\text{Ga}$  spectrum of  $\beta\text{-Ga}_2\text{O}_3$  ( $I=3/2$ ) recorded at 20 T with a MAS frequency of 60 kHz (SNR $\approx$ 500). The DW and the DE are 0.26  $\mu\text{s}$  and 5  $\mu\text{s}$ , respectively, which corresponds to an initial NMP of 20. In this case, the time domain truncation of the signal alone cannot explain the asymmetrical baseline distortions present between 1800 and 3000 ppm in the original spectrum (Figure 6, the blue line in the top left panel). This important distortion results from signal contamination of the first points of the FID probably due to pulse ringing since no Gallium probe signal was found. When trying to correct such spectrum with the truncation limit activated, the user will see that the baseline correction fails to produce a flat baseline meaning that the time domain truncation only cannot explain the baseline distortion. In this case, the user can either deactivate the truncation limit and optimize the smoothing parameter  $p$  or increase the NMP value. These settings will either remove or decrease the time domain constraint on the baseline expression allowing the modification of the FID points. Increasing the number of NMP allows the user to know how many points of the FID need to be modified to obtain a flat baseline and thus to determine the equivalent  $T_2^*$  filter duration required to remove the additional component of the baseline distortion. Both options need to be used with care: the modification of the early FID points will remove any broad component: artifacts of broadly distributed signals.

In the case of the  $^{71}\text{Ga}$  spectrum of  $\beta\text{-Ga}_2\text{O}_3$  a NMP increased to 40 corresponding to an equivalent dead time of 10  $\mu\text{s}$  allowed obtaining a satisfying baseline as shown in figure 6 in red. The spectral modeling including all transitions of the  $^{71}\text{Ga}$  and two chemical environments (in purple and orange lines), as well

as the resulting spectrum, show a very good agreement with the corrected one for both the CT (top right panel) and the STs (top left panel) resonances.

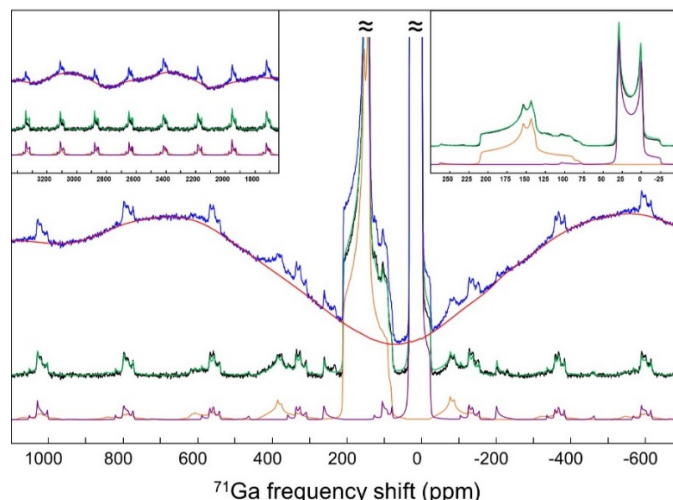


Figure 6:  $^{71}\text{Ga}$  spectrum of  $\beta\text{-Ga}_2\text{O}_3$  recorded at 20 T at a MAS spinning rate of 60 kHz. The spectrum was acquired with a DW of 0.26  $\mu\text{s}$  and a DE of 5  $\mu\text{s}$ . In blue: original spectrum; in red: the fitted baseline; in black: the corrected spectrum. In green the DMfit<sup>19</sup> decomposition obtained with the int2QUAD model including all  $^{71}\text{Ga}$  transitions and two chemical environments ( $\nu_{iso}^{(1)} = 40.3$  ppm,  $C_Q^{(1)} = 8.36$  MHz,  $\eta_Q^{(1)} = 0.085$ ;  $\nu_{iso}^{(2)} = 210.9$  ppm,  $C_Q^{(2)} = 11.23$  MHz,  $\eta_Q^{(2)} = 0.84$ ) and displayed in purple and orange respectively. The total correction was performed in a total time of 25 seconds on a laptop computer (Intel i7-4800MQ-CPU @ 2.7 MHz with 16 Go of RAM) with the default value of the smoothing parameter  $p$  of 10.

## Conclusion

We presented a baseline correction algorithm suitable for time-domain truncated pulse-acquire solid-state MAS NMR spectra acquired with large spectral width. It is shown to be of great efficiency for a wide range of nuclei ( $^{23}\text{Na}$ ,  $^{27}\text{Al}$ ,  $^{71}\text{Ga}$ ,  $^{19}\text{F}$ ...) and acquisition parameters. It allows accurate quantification of the corrected spectra with quantification errors below 0.25% for  $^{23}\text{Na}$  spectra of  $\text{NaH}_2\text{PO}_4$  or  $^{19}\text{F}$  spectra of  $\text{ZrF}_4$  and below 1.7% for  $^{27}\text{Al}$  spectra of  $\text{CaAl}_2\text{O}_4$  with Czjzek distributions even with SNR level as low as 25. It is made available to the community through a user-friendly standalone Matlab® application allowing to open Bruker and ASCII data and can be download at this address: [http://nmr.cemhti.cnrs-orleans.fr/baseline\\_corrector/default.aspx](http://nmr.cemhti.cnrs-orleans.fr/baseline_corrector/default.aspx) A video tutorial showing how to perform the baseline correction on some of the examples described in this publication is also available on the web page.

## Acknowledgments

The authors wish to express their gratitude towards Dr. Babacar Diallo and Dr. Alexey Novikov for the numerous test and comments of the baseline correction application.

## References

1. Kunath, G., Losso, P., Steuernagel, S., Schneider, H. & Jäger, C.  $^{27}\text{Al}$  satellite transition

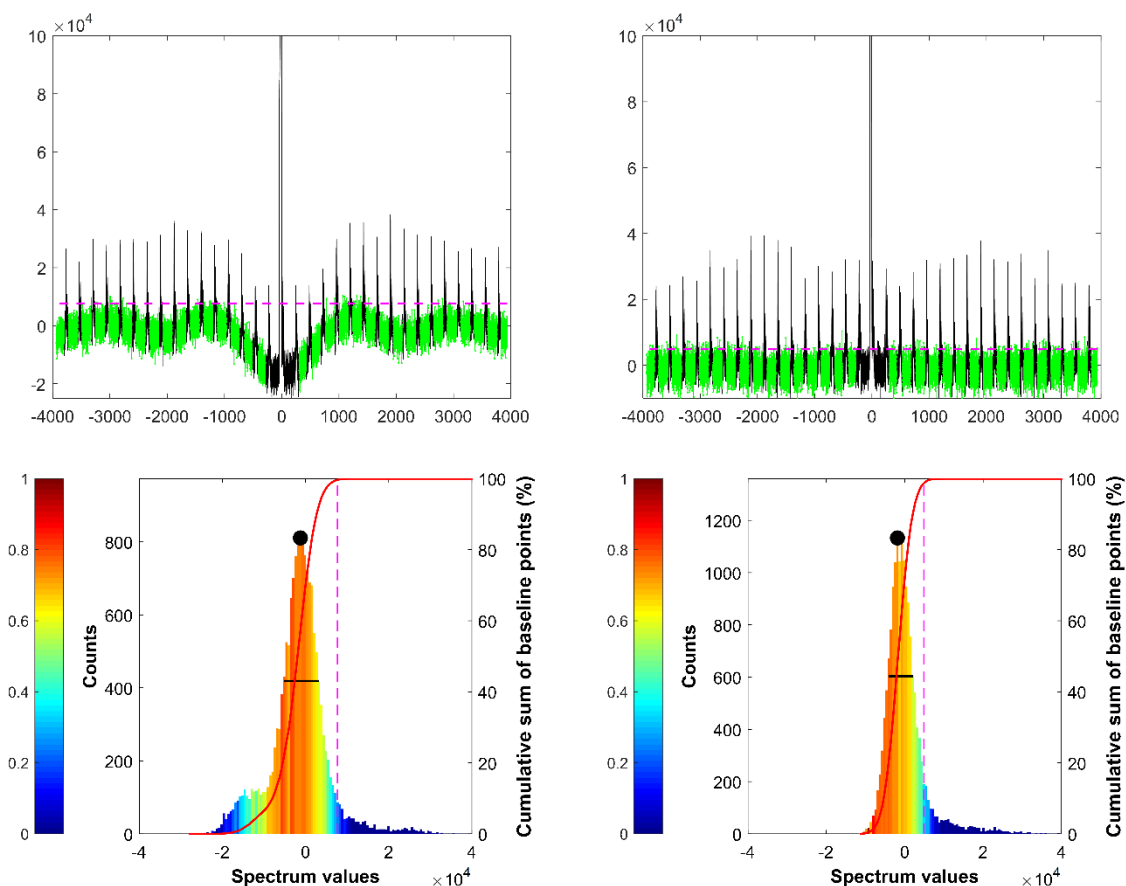
- spectroscopy (SATRAS) of polycrystalline aluminium borate  $9\text{Al}_2\text{O}_3 \cdot 2\text{B}_2\text{O}_3$ . *Solid State Nucl. Magn. Reson.* **1**, 261–266 (1992).
2. Skibsted, J., Nielsen, N. C., Bildsoe, H. & Jakobsen, H. J. Satellite transitions in MAS NMR-spectra of quadrupolar nuclei. *J. Magn. Reson.* **95**, 88–117 (1991).
  3. Heuer, A. & Haeberlen, U. A new method for suppressing baseline distortions in FT NMR. *J. Magn. Reson.* **85**, 79–94 (1989).
  4. Smith, M. E., Taulelle, F. & Massiot, D. Getting More Out of MAS NMR Studies of Quadrupole Nuclei : Part 2. *Bruker Rep.* 16–19 (1990).
  5. Alemany, L. B., Massiot, D., Sherriff, B. L., Smith, M. E. & Taulelle, F. Observation and accurate quantification of  $^{27}\text{Al}$  MAS NMR spectra of some  $\text{Al}_2\text{SiO}_5$  polymorphs containing sites with large quadrupole interactions. *Chem. Phys. Lett.* **177**, 301–306 (1991).
  6. Gerothanassis, I. P. Methods of avoiding the effects of acoustic ringing in pulsed fourier transform nuclear magnetic resonance spectroscopy. *Prog. Nucl. Magn. Reson. Spectrosc.* **19**, 267–329 (1987).
  7. Tang, C. An Analysis of Baseline Distortion and Offset in NMR Spectra. *Journal of Magnetic Resonance, Series A* **109**, 232–240 (1994).
  8. Güntert, P. & Wüthrich, K. FLATT-A new procedure for high-quality baseline correction of multidimensional NMR spectra. *J. Magn. Reson.* **96**, 403–407 (1992).
  9. Pearson, G. A. A general baseline-recognition and baseline-flattening algorithm. *J. Magn. Reson.* **27**, 265–272 (1977).
  10. Brown, D. E. Fully Automated Baseline Correction of 1D and 2D NMR Spectra Using Bernstein Polynomials. *Journal of Magnetic Resonance, Series A* **114**, 268–270 (1995).
  11. Dietrich, W., Rüdél, C. H. & Neumann, M. Fast and precise automatic baseline correction of one- and two-dimensional nmr spectra. *J. Magn. Reson.* **91**, 1–11 (1991).
  12. Friedrichs, M. S. A model-free algorithm for the removal of baseline artifacts. *J. Biomol. NMR* **5**, 147–153 (1995).
  13. Bao, Q. *et al.* A new automatic baseline correction method based on iterative method. *J. Magn. Reson.* **218**, 35–43 (2012).
  14. Marion, D. & Bax, A. Baseline correction of 2D FT NMR spectra using a simple linear prediction extrapolation of the time-domain data. *J. Magn. Reson.* **83**, 205–211 (1989).
  15. Koehl, P. Linear prediction spectral analysis of NMR data. *Prog. Nucl. Magn. Reson. Spectrosc.* **34**, 257–299 (1999).
  16. Carlos Cobas, J., Bernstein, M. A., Martín-Pastor, M. & Tahoces, P. G. A new general-purpose fully automatic baseline-correction procedure for 1D and 2D NMR data. *J. Magn. Reson.* **183**, 145–151 (2006).
  17. Chang, D., Banack, C. D. & Shah, S. L. Robust baseline correction algorithm for signal dense NMR spectra. *J. Magn. Reson.* **187**, 288–292 (2007).
  18. Golotvin, S. & Williams, A. Improved Baseline Recognition and Modeling of FT NMR Spectra. *J. Magn. Reson.* **146**, 122–125 (2000).
  19. Massiot, D. *et al.* Modelling one- and two-dimensional solid-state NMR spectra. *Magn. Reson. Chem.* **40**, 70–76 (2002).

## Supplementary information:

### Iterative baseline correction algorithm for dead time truncated one-dimensional solid-state MAS NMR spectra

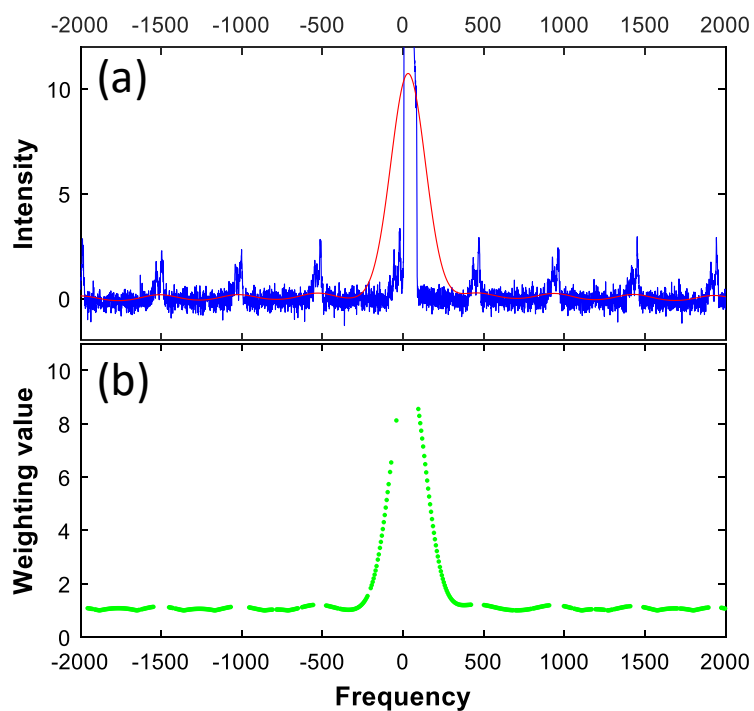
Maxime Yon<sup>1</sup>, Franck Fayon<sup>1</sup>, Dominique Massiot<sup>1</sup>, Vincent Sarou-Kanian<sup>1</sup>

<sup>1</sup>CNRS, CEMHTI UPR3079, Université d'Orléans, F-45071, Orléans, France.

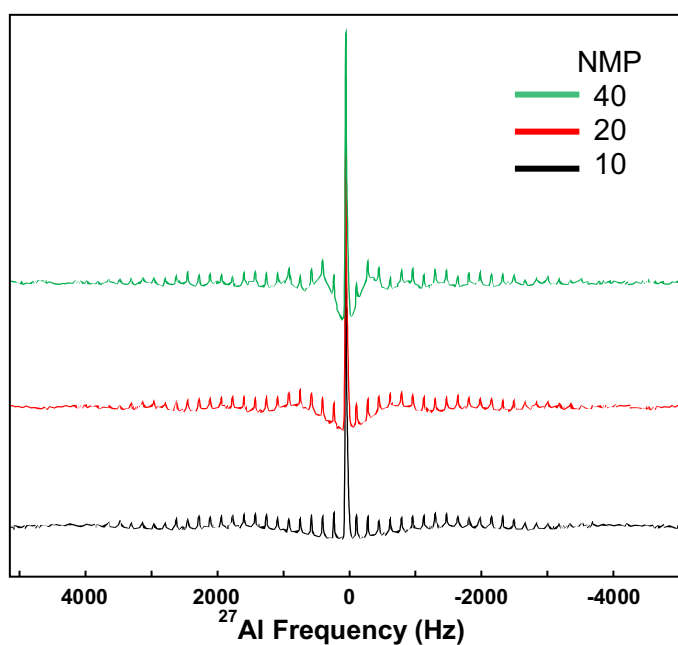


**Figure S1:** Noise distribution before and after baseline correction showing how the adaptive threshold filter allows discarding the main resonances high-intensity points. The color code in the bottom distribution shows the proportion of resonance signal divided by the proportion of noise points. The peak areas were determined via automatic peak detection on a theoretical spectrum of high SNR.

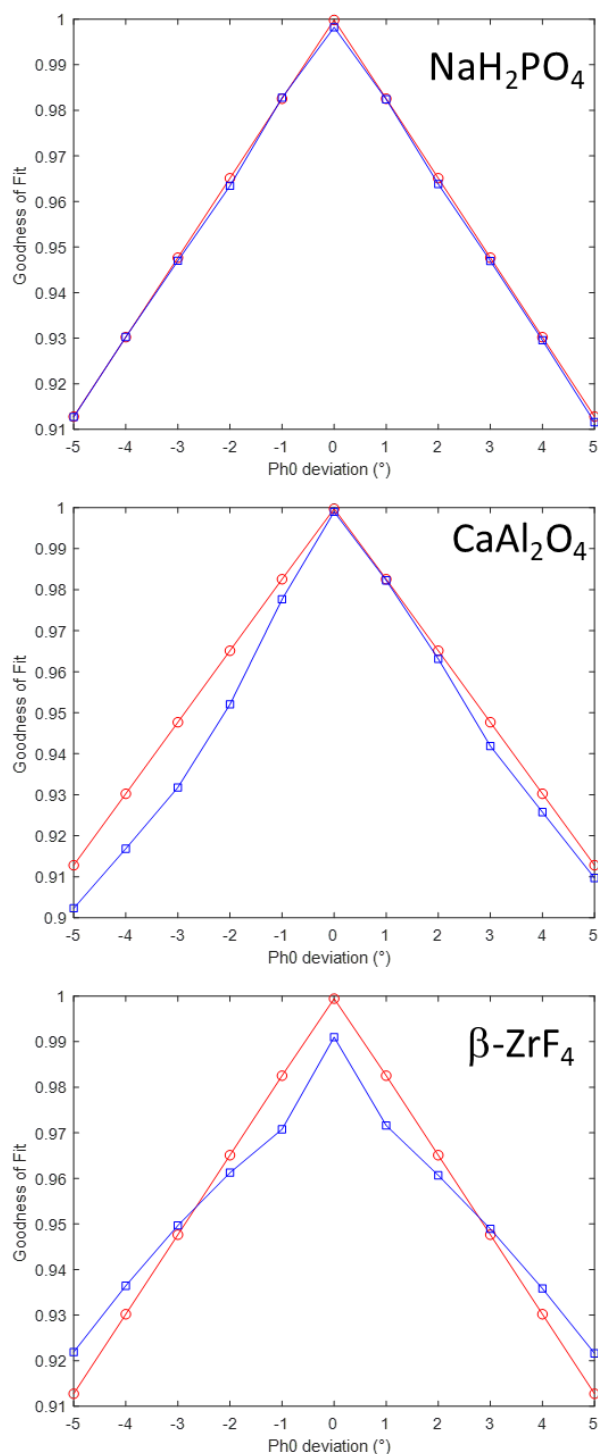




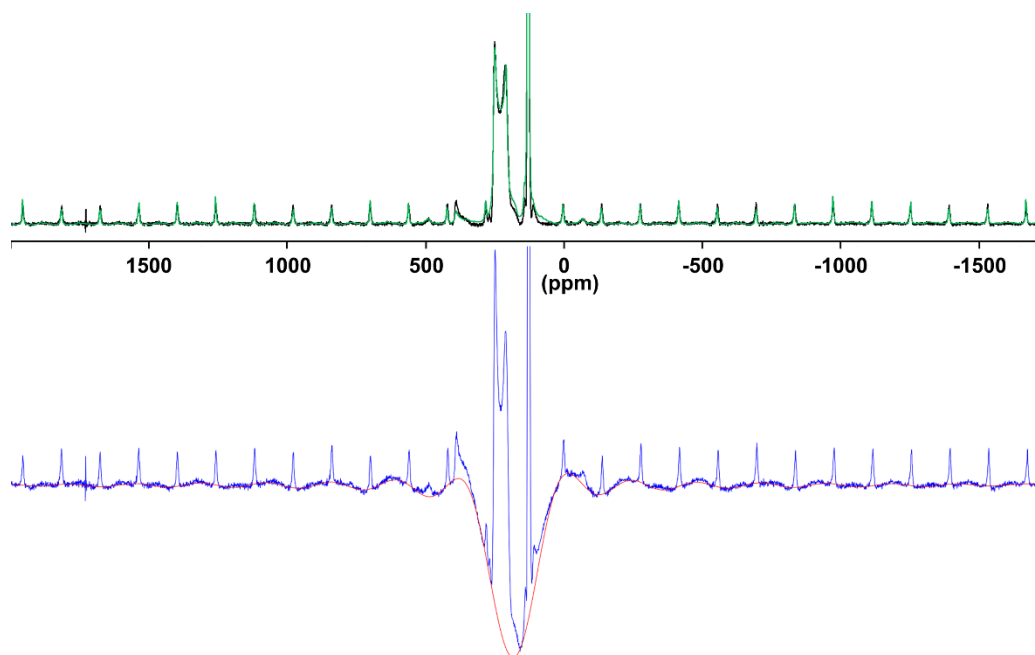
**Figure S2:** (a) Blue line: synthetic  $^{23}\text{Na}$  solid-state NMR spectrum with a signal-to-noise ratio of 20. Red line: same spectrum after a rough Gaussian apodization. (b) Weighting values calculated from the roughly broadened spectrum in (a).



**Figure S3:** glass- $\text{CaAl}_2\text{O}_4$   $^{27}\text{Al}$  spectra with a SNR of 10 000 and NMP = 10, 20, and 40 for the black, red, and green spectra respectively.



**Figure S4:** Goodness of fit comparing the NMR lineshape errors due to inaccurate zero-order phase adjustment in red, and inaccurate zero-order phase adjustment plus baseline correction in blue, for the three synthetic spectra and a NMP of 10. The error introduced by the baseline correction of the non-perfectly phased spectra (the difference between the blue and red curves) is much smaller than the error due to the phase error itself which corresponds to the red curve. Thus in case of improper zero-order phase adjustment, the error due to the baseline correction is negligible compared to the error induced by the phase bias itself.



**Figure S5:**  $^{71}\text{Ga}$  spectrum of an Yttrium Gallium Garnet sample ( $\text{Y}_3\text{Ga}_5\text{O}_{12}$ ) recorded at 20 T at a MAS spinning rate of 60 kHz. The spectrum was acquired with a dwell time of 0.3  $\mu\text{s}$  and a dead time of 10  $\mu\text{s}$ . In blue the original spectrum, in red the baseline correction obtained by cardinal sinc fit followed by the smoothing spline algorithm with the truncation limit option activated. In black the corrected spectrum. In green the DMfit modeling obtained with the int2QUAD model including all  $^{71}\text{Ga}$  transitions and two chemically different environments.

**Table S1:** Error percentage of the integration of each  $^{23}\text{Na}$  sites for the various SNR (25, 50, 100, 200 and 10000) and various NMP (10, 20, 40) corresponding to dead time durations of 6, 12 and 24  $\mu\text{s}$ , respectively. The errors whose absolute values are inferior to 1 are highlighted in green.

$\text{NaH}_2\text{PO}_4$						
NMP SNR	site 1: -2.1 ppm			site 2: -6 ppm		
	10	20	40	10	20	40
25	0,26	0,36	0,42	-0,26	-0,36	-0,42
50	0,26	0,32	0,36	-0,26	-0,32	-0,36
100	0,20	0,24	0,26	-0,20	-0,24	-0,26
200	0,18	0,22	0,22	-0,18	-0,22	-0,22
10000	0,16	0,18	0,18	-0,16	-0,18	-0,18

**Table S2:** Error percentage of the integration of the CT transition of each chemical environment of  $^{27}\text{Al}$  for the various SNR (25, 50, 100, 200 and 10000) and various NMP (10, 20, 40) corresponding to dead time durations of 2.5, 5 and 10  $\mu\text{s}$ , respectively. The errors whose absolute values are: inferior to 1 are highlighted in green, comprised between 1 and 5 are highlighted in orange and superior to 5 are highlighted in red.

$\text{CaAl}_2\text{O}_4$									
NMP SNR	CT $\text{AlO}_4$			CT $\text{AlO}_5$			CT $\text{AlO}_6$		
	10	20	40	10	20	40	10	20	40
25	0,38	0,72	2,00	-0,15	-0,74	-2,85	-13,27	-22,54	-58,03
50	0,22	0,50	1,05	-0,21	-0,67	-1,60	-6,93	-14,67	-29,78
100	0,16	0,16	0,58	-0,16	-0,27	-0,91	-4,84	-4,37	-16,45
200	0,09	0,13	0,31	-0,15	-0,21	-0,42	-2,40	-3,57	-9,01
10000	0,00	0,01	0,06	0,03	-0,05	0,00	0,03	-0,05	-2,25

**Table S3:** Error percentage of the integration of each  $^{19}\text{F}$  site for the various SNR (25, 50, 100, 200 and 10000) and various NMP (10, 20, 40) corresponding to dead time durations of 10, 20 and 40  $\mu\text{s}$ , respectively. The errors whose absolute values are: inferior to 1 are highlighted in green, comprised between 1 and 5 are highlighted in orange.

$\text{ZrF}_4$																					
NMP SNR	site 1			site 2			site 3			site 4			site 5			site 6			site 7		
	10	20	40	10	20	40	10	20	40	10	20	40	10	20	40	10	20	40	10	20	40
25	0,06	0,12	0,87	-0,93	-0,93	-1,24	0,22	0,22	0,27	0,00	-0,06	-0,90	-0,13	-0,13	-0,80	-0,30	-0,45	-0,91	0,42	0,48	1,32
50	0,06	0,06	0,87	-0,93	-0,93	-1,24	0,22	0,22	0,27	-0,06	-0,06	-0,90	-0,13	-0,13	-0,80	-0,30	-0,30	-0,91	0,48	0,48	1,32
100	0,06	0,06	0,81	-0,93	-1,04	-1,24	0,22	0,22	0,27	0,00	-0,06	-0,83	-0,13	-0,13	-0,74	-0,30	-0,45	-0,91	0,42	0,48	1,27
200	0,06	0,06	0,87	-0,73	-0,93	-1,14	0,16	0,22	0,33	0,00	-0,06	-0,83	-0,07	-0,13	-0,80	0,00	-0,30	-0,91	0,26	0,42	1,22
10000	0,00	0,00	0,56	-0,21	-0,31	-1,04	0,00	0,05	0,16	0,00	0,00	-0,64	-0,07	-0,07	-0,60	0,45	0,45	-0,45	-0,05	0,00	1,01