Supplementary material

Denoising applied to spectroscopies – part I: concept and limits

Guillaume LAURENT (1*), William WOELFFEL (2), Virgile BARRET-VIVIN (1), Emmanuelle GOUILLART (2), Christian BONHOMME (1)

(1) Sorbonne Université, Collège de France, CNRS, Laboratoire de Chimie de la Matière Condensée de Paris, LCMCP, F-75005, Paris, France.

(2) Saint-Gobain, CNRS, Surface du Verre et Interfaces, SVI, F-93300, Aubervilliers, France.

(1) Case Courrier 174, 4 place Jussieu, 75252 Paris Cedex 05, France.

(2) 39 quai Lucien Lefranc, BP 135, 93303 Aubervilliers Cedex, France.

(*) Corresponding author

guillaume.laurent@sorbonne-universite.fr, orcid.org/0000-0002-8127-5326

william.woelffel@saint-gobain.com

vigile_bv@msn.com

emmanuelle.gouillart@saint-gobain.com

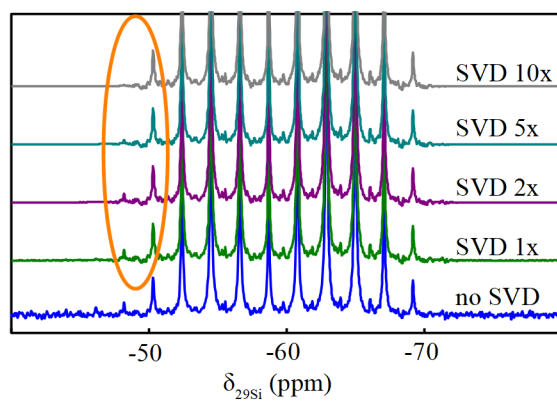christian.bonhomme@sorbonne-universite.fr, orcid.org/0000-0003-0802-6961

Figure S1: Influence of SVD iterations for a matrix of 3901×128 points (n = 128) with k = 22 singular values. 4028 points ^{29}Si CPMG MAS solid-state NMR spectrum of the 50 / 50 MTEOS / TEOS sample. Orange ellipsis shows slight decrease of noise.
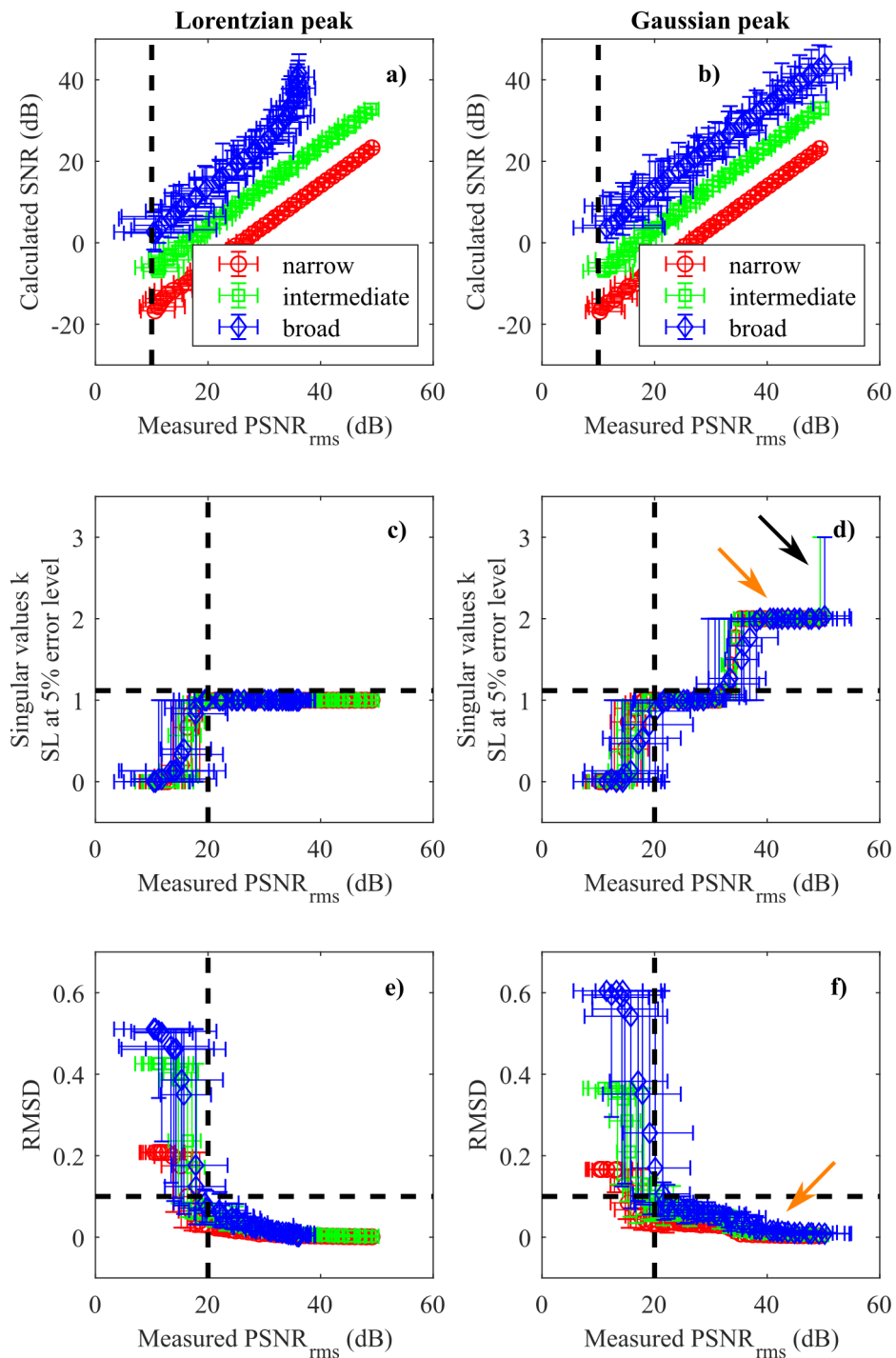
Figure S2: 7380 simulated spectra with known added homoscedastic white Gaussian noise, **with truncation of FID at 5 T$_2$**. a), c) and e) Lorentzian peak; b), d) and f) Gaussian peak; a) and b) comparison of SNR definitions; c) and d) automatic threshold value with SL at 5% error level; e) and f) root mean square deviation of denoised spectra. Error bars correspond to the repetition of 30 simulated spectra with the same level of added noise. 41 level of noise were used ranging from -20 dB to +20 dB. SVD was applied on time data before **zero-filling**; SNR and PSNR were obtained on frequency data. Dashed black lines represent characteristic values; black arrow shows artefacts detection; orange arrows highlight the second singular value for Gaussian peaks.
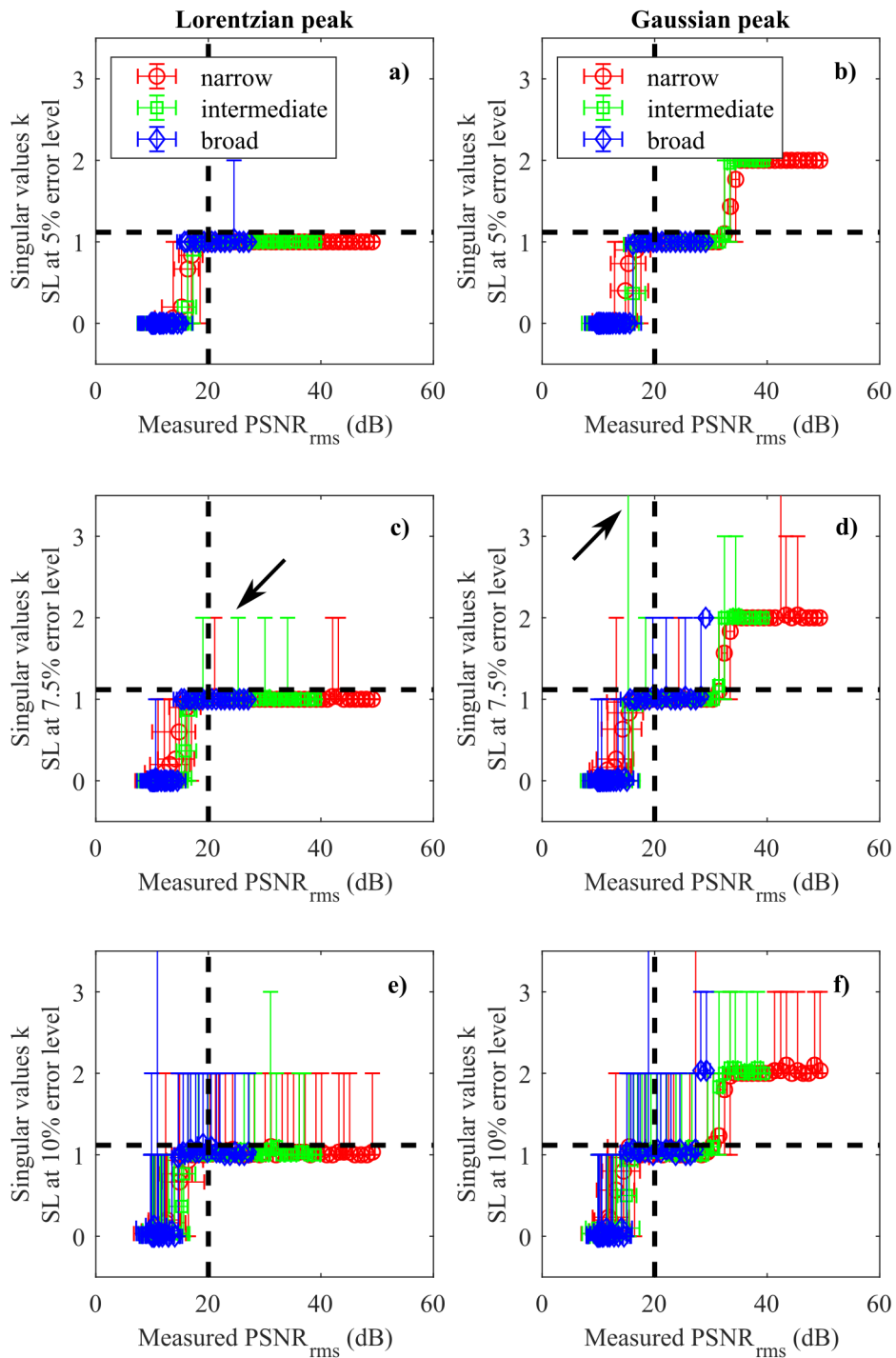
Figure S3: 3×7380 simulated spectra with known added homoscedastic white Gaussian noise. a), c) and e) Lorentzian peak; b), d) and f) Gaussian peak; automatic threshold value with a) and b) SL=5 %; c) and d) SL=7.5 %; e) and f) SL=10 %. Error bars correspond to the repetition of 30 simulated spectra with the same level of added noise. 41 level of noise were used ranging from -20 dB to +20 dB. SVD was applied on time data; PSNR was obtained on frequency data. Dashed black lines represent characteristic values; black arrows show artefacts detection. Data sets are available in files Data_SL5.mat, Data_SL7.5.mat, and Data_SL10.mat of http://doi.org/10.5281/zenodo.1406172.
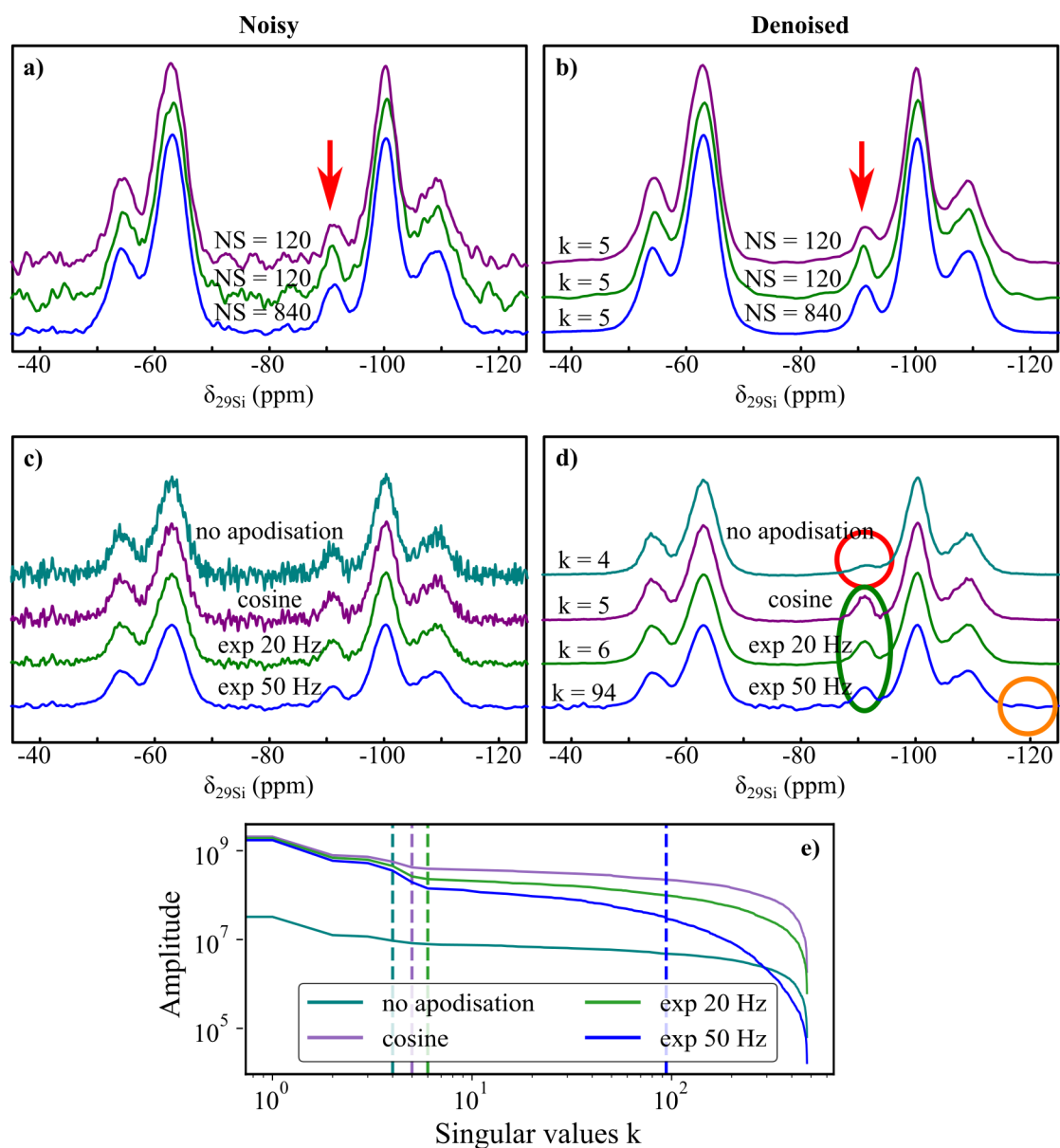
Figure S4: a) and b) Influence of number of scans (NS) on SVD with manual thresholding and 50 Hz exponential apodisation; c) and d) influence of pre-processing before SVD with automatic thresholding and apodisation as annotated on spectra; a) and c) noisy spectra; b) and d) denoised spectra; e) singular values k of spectra in c), in logarithmic scale. ^{29}Si MAS solid-state NMR spectrum of the 50 / 50 MTEOS / TEOS sample. NS: number of scans; k: number of singular values. Red arrows evidence variations of peak amplitude; red and green ellipsis highlight not detected and detected peaks, respectively; orange circle shows ineffective denoising; dashed vertical lines correspond to automatic thresholding of data with the same color, at an SL error level of 7.5 %.