Increasing solid-state NMR sensitivity : instrumentation, fast acquisitions and signal processing

Guillaume Laurent

Sorbonne Université

École doctorale 397

Laboratoire de Chimie de la Matière Condensée de Paris (LCMCP, UMR 7574)

Équipe Spectroscopies, Modélisations, Interfaces pour l'Environnement et la Santé (SMiLES)

Increasing solid-state NMR sensitivity

Instrumentation, fast acquisitions and signal processing

Par Guillaume P. Laurent

Thèse de doctorat de Physique et Chimie des Matériaux

Dirigée par Christian Bonhomme

Présentée et soutenue publiquement le 13 février 2020

Devant un jury composé de :

Dr. Dimitrios Sakellariou	Professeur, KU Leuven, Belgium	Président
Dr. Marc-André Delsuc	Directeur de Recherche, IGBMC, Strasbourg	Rapporteur
Dr. Luminita Duma	Chargée de Recherche, GEC, Compiègne	Rapporteuse
Dr. Catherine Achard	Maîtresse de Conférences, ISIR, Paris	Examinatrice
Dr. Alain Louis-Joseph	Ingénieur de Recherche, PMC, CNRS, IPP, Palaiseau	Examinateur
Dr. Justine Vieville	Application Manager, RS2D, Mundolsheim	Invitée
Dr. Christian Bonhomme	Professeur, LCMCP, Paris	Directeur

A tous ceux qui n'ont pas confiance en eux

To all those who are not self-confident

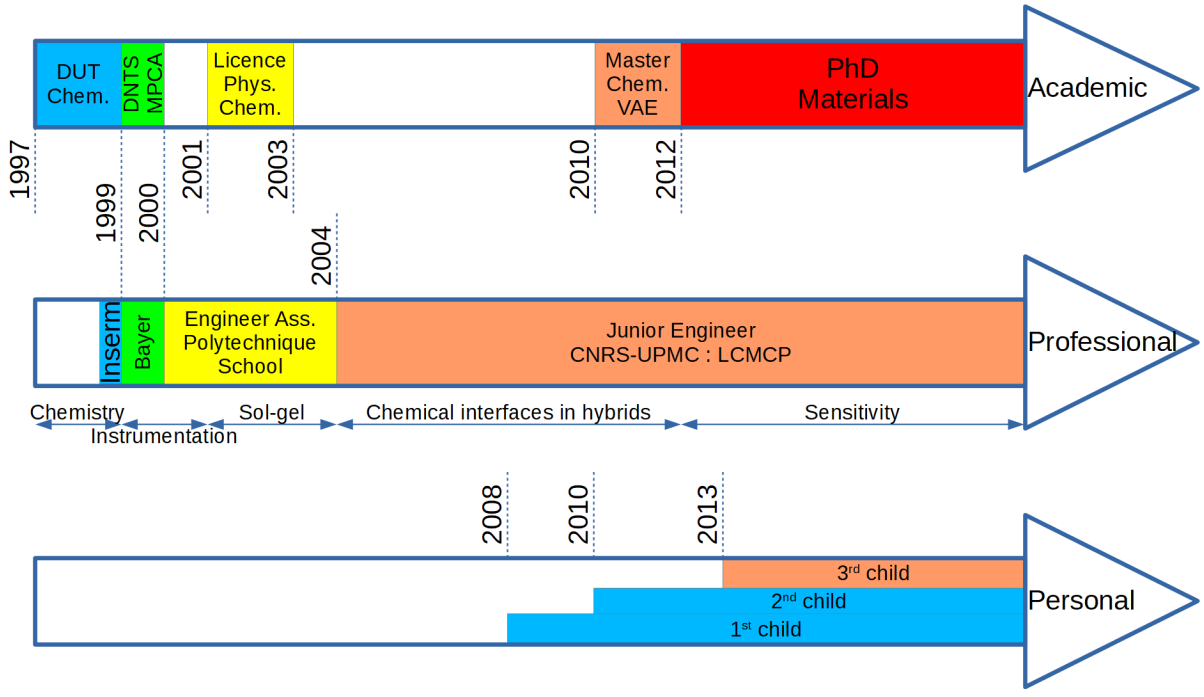
Preface

'Everything is under control!'

Guillaume P. Laurent (1979-), PhD candidate

My itinerary is uncommon (see figure below): I started with an associate degree in Chemistry in 1999 before specialising in Physical-Chemical Analysis Methods (MPCA) in 2000, which was done by work-based learning. In 2003, I validated a Physical-Chemistry Licence, simultaneously to my work as an engineer assistant (AI) in Condensed Matter Physics laboratory (PMC). In 2012, I obtained a Master's degree by validating my experience. Hence, during my academic curriculum, I had only a few mathematics and no electromagnetism courses at all. This PhD started in the continuation of my master.

Why starting a PhD? As I have already a permanent CNRS junior engineer position (so-called IE), this is a good question. There are three reasons for this: firstly for the search of knowledge, secondly as a personal challenge, and thirdly to have a career evolution as a senior engineer (IR). This PhD was done in partial time in addition to my junior engineer job, which explains why it took 7.5 years. Reading a complex publication or trying to find a bug in a program with three children screaming loudly around, was not an easy task. In such a case, I could really say 'Everything is under control!'



Remerciements

Avant tout, je voudrai remercier les **membres du jury** d'avoir accepté d'évaluer ce travail, et particulièrement les rapporteurs pour leur lecture attentive.

Cette thèse s'étant écoulée sur une longue période, j'ai épuisé trois **directeurs de laboratoire** : Clément Sanchez, Florence Babonneau et Christian Bonhomme. Je les remercie infiniment pour leur confiance et pour avoir accepté d'aménager mon temps de travail, ainsi que François Ribot en tant que directeur adjoint. L'**École Doctorale** 397 m'a soutenu année après année. Christian Bonhomme était également mon **directeur de thèse**. Un énorme merci à lui pour ses conseils toujours pertinents, sa patience, et surtout pour son humour. Ce n'est pas donné à tout le monde de faire partie de l'École du Rire.

Cédric Lorthioir et Sylvie Masse ont vraiment été d'un soutien sans faille au **quotidien**. Merci à eux pour leurs avis. J'exprime ma gratitude aux collègues de **repas** qui m'ont encouragé chaque jour pendant les six derniers mois. Il en est de même pour toute l'équipe **SMiLES**, et particulièrement Thierry Azaïs. Je remercie également les **chercheurs** des autres groupes, et notamment Cédric Boissière, Gervaise Mosser, Nadine Nassif et Capucine Sassoye. Tous le personnel de la **plateforme RMN** m'a énormément soutenu : Aurelie Bernard, Cristina Coelho-Diogo, Isabelle Correia, Flavien Guenneau, Franck Lacour, Olivier Lequin, Andreï Nossov, François Ribot, Baptiste Rigaud et Claire Troufflard.

Un laboratoire ne peut rien faire sans une **équipe administrative** performante : Fatima Bodih, Hélène Gervais, Diana Lesueur et Corinne Pozzo di Borgo. Je n'oublie pas l'**équipe technique**. J'en profite pour remercier le personnel et les directions successives de l'**Institut des Matériaux de Paris Centre**.

Au cours de ces années, j'ai vu passer de nombreux **doctorants** qui m'ont encouragé. En particulier, Virgile Barret-Vivin, Leslie Dos Santos, François Portier et Stanislas Von Euw m'ont fourni des échantillons à analyser. Je remercie mes **stagiaires** : Samy Liso (2015), Pierre-Aymeric Gilles (2017), et Winh-Chhunn Teh (2019).

Le mot de la fin revient à ma **famille**, qui a supporté mes humeurs et mon manque de temps. Mon père et mon beau père ont beaucoup œuvré pour les travaux de la maison.

Résumé long en français

A. Introduction

La Résonance Magnétique Nucléaire (RMN) en phase solide est un domaine en constante évolution, notamment depuis une vingtaine d'années. Le champ magnétique des appareils commerciaux atteint maintenant $B_0 = 28.2$ T (1.2 GHz pour le ^{1}H), et $B_0 = 35.2$ T (1.5 GHZ pour le ^{1}H) pour l'aimant hybride résistif/superconducteur du Laboratoire National des Très Hauts Champs Magnétiques (NHMFL, Tallahassee, FL, USA). En parallèle, la rotation à l'angle magique (*MAS*) est de plus en plus rapide, à savoir 111 kHz pour les sondes commerciales et 170 kHz pour celles en développement de diamètre 0.5 mm (dans le groupe d'A. Samoson). S'ajoutent à ces améliorations techniques des développements méthodologiques importants, avec des séquences d'impulsions permettant des découplages ou recouplages homo- ou hétéronucléaires toujours plus performants, permettant d'aller sonder les proximités des noyaux au sein de l'échantillon. Ainsi, la combinaison d'ultra-haut champ magnétique, de MAS ultra-rapide, et d'acquisition inverse sur le ^{1}H a permis d'étudier des objets toujours plus complexes, telles que des protéines de 153 résidus, des prions ou même des capsides virales (*1*).

Malgré ces innovations, le principal défaut de la RMN reste sa faible sensibilité, qui est toujours une question ouverte au sein de la communauté. Cet inconvénient est particulièrement présent dans trois cas parmi d'autres : les noyaux de rapport gyromagnétique faible ou intermédiaire, qui ne bénéficient que marginalement de l'augmentation du champ magnétique ; les noyaux présentant de fortes anisotropies, notamment les spins quadripolaires, dont le spectre s'étale sur plusieurs centaines de kHz et est masqué par le bruit ; et d'une manière générale les spins rares, que ce soit du fait de l'abondance naturelle intrinsèque, de la dilution au sein de l'échantillon ou de la faible quantité de matière associée.

Dans le Chapitre I, après avoir rappelé les bases de la RMN en phase solide, nous nous sommes intéressés aux notions de bruit et d'artefacts. Le bruit en RMN est principalement lié à l'agitation thermique des électrons dans la bobine et dans le préamplificateur (bruit Johnson). La répartition du bruit ne dépend pas de la fréquence, ce que l'on appelle un bruit blanc. De plus, son intensité suit une loi gaussienne. Lors de l'acquisition du spectre,

différents types d'artefacts peuvent apparaître, que ce soit un pic à fréquence nulle, une image miroir des pics, des fantômes à différentes fréquences ou une "traînée" verticale sur les spectres nD (bruit-t_1). Le procédé de traitement peut également créer des artefacts, comme c'est le cas lors de l'utilisation de la covariance.

La sensibilité peut être définie comme la capacité à distinguer de faibles différences de concentration. Elle est intimement liée à la limite de détection. L. Currie a formalisé mathématiquement trois seuils (*2*) : la limite de décision (L_c), la limite de détection (L_d) et la limite de quantification (L_q), correspondant à 1.64, 3.29 et 10 σ_{bruit}, respectivement, où σ_{bruit} est l'écart-type du bruit. Le rapport signal-sur-bruit (*SNR*) peut quant à lui être défini de plusieurs manières. Il est appelé *PSNR* quand il est mesuré sur la hauteur du pic de signal (H_{signal}), tandis que le bruit est mesuré soit par sa moyenne quadratique (*RMS*), soit sur la hauteur du bruit pic-pic ($h_{bruit\_pic\_pic}$).

En mathématiques et en électronique (*3*) $$SNR = \frac{\sigma_{signal}^2}{\sigma_{bruit}^2} \quad (1)$$

En spectroscopies[1] (*4*) $$PSNR_{rms} = \frac{H_{signal}}{\sigma_{bruit}} \quad (2)$$

En chimie analytique (*5*) $$PSNR_{max} = \frac{H_{signal}}{h_{bruit\ \ pic\ \ pic}/2} \quad (3)$$

Dans le cas d'un bruit gaussien, $PSNR_{rms} \sim 3.3\ PSNR_{max}$. Tandis que $PSNR_{rms}$ est utilisé couramment en RMN, cette formule est inadaptée en présence d'artefacts non gaussiens, et la formule $PSNR_{max}$ doit être préférée. Le traitement du signal a été présenté dans la suite de ce chapitre, en donnant les différentes étapes, depuis la sonde jusqu'au spectre. L'apodisation, la Transformée de Fourier (FT) et le phasage ont aussi été expliqués.

Depuis 2010, la solution usuelle pour augmenter la sensibilité est l'utilisation de la Polarisation Nucléaire Dynamique (DNP). Cette technique requiert l'achat de nouveaux équipements coûteux. D'autres alternatives existent, complémentaires de la DNP, telles que (i) l'instrumentation, (ii) les acquisitions rapides et (iii) le traitement du signal. Ces trois approches seront utilisées dans ce manuscrit. (i) Les Bobines en Rotation à l'Angle Magique (MACS) seront présentées dans le Chapitre II pour l'acquisition de spectres associés à des échantillons en quantité limitée. (ii) Les échos Carr-Purcell-Meiboom-Gill (CPMG, Chapitre III) et l'Échantillonnage Non-Uniforme (NUS, Chapitre IV) permettront de réduire

[1] En RMN, l'écart-type du bruit est doublé pour des raisons historiques.

le temps d'acquisition dans la dimension directe et dans les dimensions indirectes des spectres multi-dimensionnels (nD), respectivement. (iii) Enfin, les spectres seront débruités par Décomposition en Valeurs Singulières (SVD, Chapitre V) et le temps de calcul sera optimisé sur processeurs (CPU) et cartes graphiques (GPU, Chapitre VI).

Au cours de cette étude, nous avons appliqué cette méthodologie sur des échantillons d'intérêt dans plusieurs contextes chimiques. Un premier domaine concerne les HydroxyApatites (HAp), un phosphate de calcium hydroxylé, qui est le principal constituant inorganique des os et des dents des mammifères. Ce matériau a une complexité chimique importante du fait de la substitution possible à la fois par des anions (CO_3^{2-},…) ou des cations (Na+,…), ainsi qu'une complexité structurale (cristalline ou partiellement amorphe), qui nécessite donc une sensibilité accrue pour être caractérisée au mieux. La deuxième catégorie concerne les matériaux hybrides organiques/inorganiques, entre autres à base de silice, préparés à basse température par voie sol-gel. La communauté a un fort besoin de quantification sur des films minces ou des membranes, par exemple dans le domaine de l'énergie. Le dernier contexte chimique concerne les verres bioactifs contenant des noyaux quadripolaires tels que le ^{91}Zr ou le ^{87}Sr, ce dernier permettant de lutter contre l'ostéoporose.

B. Microbobines

Le but du Chapitre II était d'analyser des échantillons de l'ordre de 100-200 µg, ce qui correspond par exemple à la quantité de matière obtenue en récupérant 2 cm^2 d'un film mince de 300 nm de silice déposée sur un substrat. De plus, dans le cas d'échantillons enrichis, leur coût peut être tel qu'il interdit l'utilisation d'une plus grande masse d'échantillon. Un dernier exemple concerne les calculs rénaux, dont le point de départ, appelé plaque de Randall, ne mesure que quelques dizaines de microns de long et nécessite d'être parfaitement caractérisé afin d'établir le bon diagnostic et le traitement médical adéquat.

Après avoir présenté les solutions commerciales pour analyser des échantillons de taille de plus en plus réduite, nous nous sommes intéressés aux microbobines. Le terme microbobine est réservé aux bobines dont le diamètre est inférieur à 1 mm. En ajustant la taille de la bobine à celle de l'échantillon, on maximise le taux de remplissage de la bobine et on optimise le signal obtenu. De plus, d'après le principe de réciprocité, plus la distance entre l'échantillon et la bobine est faible, plus l'excitation et la réception seront efficaces. La microbobine concentre le champ électro-magnétique B_1 en son centre et permet un gain en

sensibilité pour une même puissance appliquée P donné par $PSNR \propto B_1/\sqrt{(P)}$.

Les Bobines en Rotation à l'Angle Magique (MACS) permettent de combiner l'utilisation de microbobines au plus près de l'échantillon et les sondes MAS standard (*6*). Un circuit auto-résonant est centré grâce à un insert en céramique thermiquement conductrice (Shapal-M) au sein d'un rotor de gros diamètre, lui même placé en rotation à l'angle magique. Par couplage inductif entre la sonde et la microbobine, l'énergie est transférée de l'un à l'autre, comme entre les circuits primaires et secondaires d'un transformateur électrique. Deux géométries ont été utilisées : un solénoïde enroulé autour d'un capillaire dont le circuit était fermé par un condensateur (Figure 1a) (*6*) et un Résonateur à Ligne de Transmission (TLR), constitué d'une double piste en quasi-spirale, séparées par un substrat jouant le rôle de condensateur (Figure 1b) (*7*). Ces deux géométries ont l'avantage de ne pas nécessiter de modification de la sonde. De plus, les TLR sont mécaniquement moins fragiles que les solénoïdes.

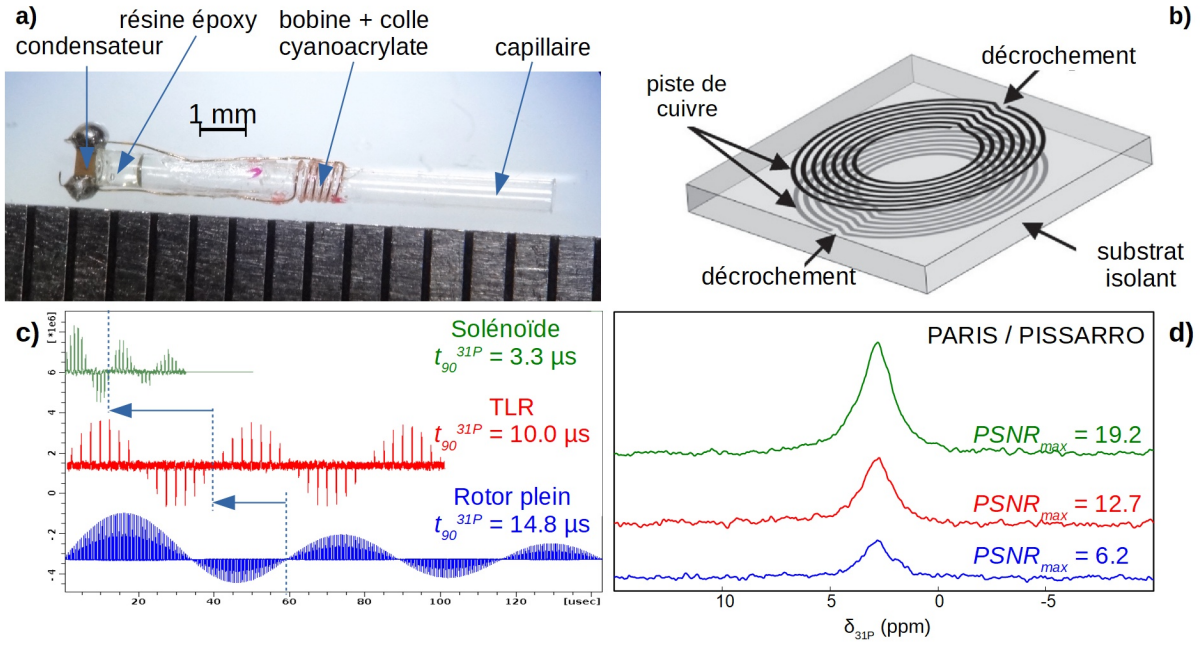


Figure 1: Instrumentation: a) microbobines solenoïde et b) TLR ; c) comparaison des profiles de nutation ; d) spectres ^{31}P obtenus avec la séquence d'impulsions PARIS/PISSARRO sur un échantillon d'HAp partiellement amorphe. Paramètres : B_0 = 11.7 T, rotor de 7 mm, *MAS* = 5 kHz, délai de recyclage *RD* = 10 s, *PARIS* = 10 s.

Après une étude de l'ensemble des paramètres impliqués dans la conception des microbobines, nous avons fourni une feuille de calcul LibreOffice permettant de choisir le condensateur à souder sur les solénoïdes, en fonction de la fréquence de résonance désirée, du diamètre du capillaire, du matériau du fil, de son diamètre, du pas entre les spires, du nombre

de spires, de la position de la bobine et de la taille du condensateur. Des messages d'erreur s'affichent quand le pas est incorrect, quand le diamètre du fil est trop faible par rapport au courant de peau, quand les courants de Foucault génèrent trop de chaleur et quand il n'existe pas de valeur du condensateur adéquate. La résistance et le facteur qualité du système sont également calculés automatiquement.

Les géométries solénoïdes et TLR ont été comparées sur la même sonde, et les solénoïdes ont permis un champ RF plus important à la même puissance (Figure 1c), et donc un *PSNR* plus important (Figure 1d). Des séquences d'impulsions basse puissance (PARIS/PISSARRO et TEDOR) ont été utilisées car la Polarisation Croisée (CP) entraînait la détérioration des TLR, du fait de la puissance RF nécessaire. Un gain en temps d'un facteur ~ 5 a été obtenu avec MACS.

C. Échos Carr-Purcell-Meiboom-Gill

Dans le Chapitre III, nous nous sommes intéressés aux échantillons en quantité plus importante, mais dont les spins étaient soit dilués par l'abondance naturelle et la teneur dans l'échantillon (films hybrides organiques/inorganiques utilisés dans le domaine de l'énergie), soit présentaient de larges anisotropies (noyaux quadripolaires dans des verres bioactifs).

Une première approche d'acquisition rapide a consisté à utiliser des échos Carr-Purcell-Meiboom-Gill (CPMG) qui permettent, grâce à un train d'impulsions 180°, de refocaliser régulièrement l'aimantation. Entre chaque écho, l'aimantation décroît selon la relaxation transversale apparente T_2^*, tandis que l'enveloppe globale du train d'échos suit la "vraie" relaxation transversale T_2 (Figures 2a1 et 2a2). Dans le cadre de l'étude des noyaux quadripolaires, la phase de la séquence d'impulsion est modifiée et cette dernière se nomme QCPMG. Le choix des délais est important pour éviter les artefacts et les chemins de cohérence indésirés (*8*). De même, une synchronisation avec la rotation à l'angle magique doit être mise en place.

Plusieurs protocoles de traitement du signal peuvent être utilisés (*9*). Le plus simple est d'appliquer la FT sur le peigne d'échos, ce qui donne directement un peigne de pics fins, dont l'enveloppe reproduit la forme du spectre sans écho (Figure 2a3). C'est la méthode par spikelets. Alternativement, les échos peuvent être d'abord repliés en deux puis sommés tous

ensemble avant de faire la FT (Figure 2b), ce qui donne directement l'enveloppe globale du spectre. C'est la méthode par superposition. Les gains attendus sont indiqués ci-dessous.

$$G_{spikelets} = \frac{1}{\sqrt{1+2N}}\left[1+2\sum_{n=1}^{N}\cos\left(2\omega n\tau_{echo}\right)e^{-2n\frac{\tau_{echo}}{T_2'}}\right] \quad 4$$

$$G_{superposition} = \frac{1}{\sqrt{1+2N}}\left[1+2\sum_{n=1}^{N}e^{-2n\frac{\tau_{echo}}{T_2'}}\right] \quad 5$$

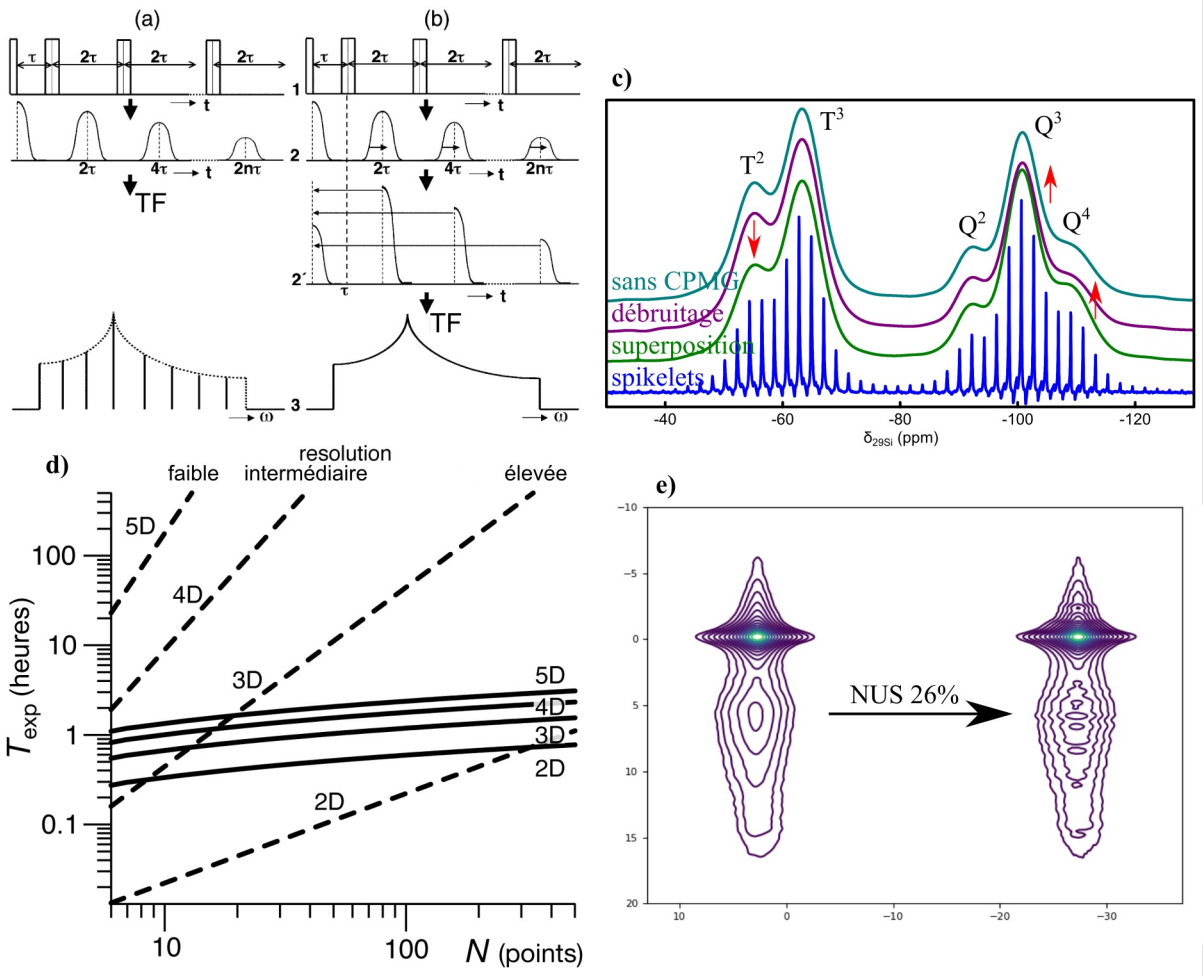


Figure 2: Acquisitions rapides : a) traitement CPMG avec la méthode par spikelets et b) par superposition ; c) comparaison des méthodes CPMG sur un spectre ^{29}Si CP d'un hybride organique/inorganique ; d) temps d'acquisition et résolution obtenues pour des expériences 2D à 5D avec US (tirets) et avec NUS (traits pleins) ; e) spectre CP 2D ^{1}H-^{31}P d'un échantillon d'HAp partiellement amorphe obtenu avec l'échantillonnage hybride et le meilleur algorithme de reconstruction. Paramètres : c) et e) B_0 = 7.0 T, rotor de 4 mm, *MAS* = 14 kHz, *RD* = 1 s.

Il est également possible d'appliquer une apodisation globale sur le train d'écho pour limiter l'effet de troncature, ainsi qu'une apodisation sur chaque écho afin de diminuer le bruit. L'inconvénient majeur de ces deux méthodes est qu'elles sont sensibles aux différences

de T_2 "vrai" entre les sites. Ainsi, des distorsions d'intensité apparaissent sur les spectres bleus et verts de la Figure 2c, mis en évidence par les flèches rouges. Les espèces T^2 sont minorées, tandis que les espèces Q^3 et Q^4 sont majorées, sur une silice hybride.

Nous avons développé un programme Python de plus de 2300 lignes de code permettant tout d'abord de synthétiser un train d'échos parfaitement connu, y compris en y introduisant un déphasage ou une mauvaise synchronisation. Dans un deuxième temps, les délais de l'écho sont calibrés, une apodisation individuelle et globale est appliquée, et les résultats des méthodes par spikelets et par superposition sont obtenues. En supplément, une nouvelle méthode est fournie, par somme partielle des échos et débruitage SVD. Cette dernière méthode permet une quantification améliorée (spectre violet sur la Figure 2c). Un gain en temps allant de 3 à 100 a été obtenu avec le (Q)CPMG.

D. Échantillonnage non-uniforme

Dans le Chapitre IV, une autre approche est utilisée afin d'augmenter la vitesse d'acquisition : l'Échantillonnage Non-Uniforme (NUS). L'une des forces de la RMN est de pouvoir obtenir des cartes multidimensionnelles (nD) qui corrèlent plusieurs interactions entre elles. L'inconvénient est que ces cartes sont acquises point à point dans les dimensions indirectes, ce qui augmente considérablement le temps d'acquisition. En échantillonnage uniforme (Uniform Sampling, US), la solution est de limiter le nombre de tranches (tirets sur la Figure 2d), ce qui entraîne une troncature du signal. Une perte de résolution est observée pour une durée d'acquisition dans la dimension indirecte $AQ < 3\ T_2^*$ et même une perte de sensibilité pour $AQ < 1.26\ T_2^*$ (*10*). Contrairement au CPMG qui optimise la dimension directe, NUS s'intéresse aux dimensions indirectes des nD. En n'acquérant qu'une partie des points, NUS réconcilie résolution, sensibilité et temps d'acquisition (traits pleins sur la Figure 2d).

Le choix du schéma d'échantillonnage a un impact considérable sur la qualité du spectre final. En effet, ce dernier est convolué par la Fonction de Dispersion des Points (PSF), qui est la FT du schéma d'échantillonnage. Les artefacts présents sur la PSF masquent les pics d'intérêts sur le spectre final. Les schémas de type Poisson permettent une répartition aléatoire dans l'espace à échantillonner, tout en évitant d'avoir deux points côte à côte. Cela diminue les artefacts et facilite la détection des pics peu intenses. Toutefois, même les schémas de type Poisson ne sont pas tous équivalents, et le programme NUSscore permet de

les classer en fonction de leur qualité (*11*).

La reconstruction du spectre obtenu par NUS est un problème à part entière. De nombreux algorithmes sont disponibles : l'Entropie Maximum (MaxEnt) et l'Acquisition Comprimée (CS) sont les plus fréquemment utilisées. Les deux algorithmes fonctionnent de manière opposée : MaxEnt commence par un spectre constitué de zéros et reconstruit la dimension temporelle correspondante, tandis que CS effectue la FT de la dimension temporelle et reconstruit le spectre. De plus, chacun a de nombreuses variantes. A l'heure actuelle, il n'est pas clair si l'un surpasse l'autre. Par ailleurs, le nombre d'itérations pour reconstruire le spectre est difficile à optimiser et des artefacts sont introduit au-delà du nombre optimum.

Après avoir appliqué NUS sur un échantillon mobile de gélatine, qui est la forme dénaturée de la protéine principale dans les os de mammifères, nous nous sommes intéressés à des HAp constituant la partie minérale de l'os. Tandis que le schéma par défaut du logiciel TopSpin (moteur aléatoire 54321) s'est révélé inadapté, l'échantillonnage de type Poisson s'est révélé plus performant. Enfin, un schéma hybride US/NUS a été mis en place sur un échantillon d'HAp amorphe/cristallin présentant à la fois des pics larges et fins en quatre fois moins de temps (Figure 2e).

E. Décomposition en valeurs singulières

Une fois le spectre obtenu, si le *PSNR* n'est pas satisfaisant, il est possible d'effectuer du traitement du signal et par exemple de débruiter par Décomposition en Valeurs Singulières (SVD, Chapitre V). Au préalable, il est nécessaire de convertir le spectre 1D en matrice de Hankel ou de Toeplitz. La SVD est ensuite appliquée sur cette matrice X qui est décomposée en trois autres matrices (rectangles hachurés en orange sur la Figure 3a) : une matrice carrée U de la taille du nombre de lignes ; une matrice Σ de la taille de X, mais avec des valeurs uniquement sur la diagonale ; et une matrice carrée V^T de la taille du nombre de colonnes. Les valeurs de Σ sont classées par ordre de taille. Tandis que les premières correspondent aux pics d'intérêt, les suivantes correspondent à du bruit. En ne sélectionnant que les k premières valeurs singulières, c'est à dire en diminuant le rang de la matrice Σ, il est possible de supprimer le bruit tout en conservant le signal (rectangles bleus pleins sur la Figure 3a). L'effet du débruitage SVD est présenté sur la Figure 3b.

$$X = U \Sigma V^T$$
$$X \approx X_k = U_k \Sigma_k V_k^T$$

6

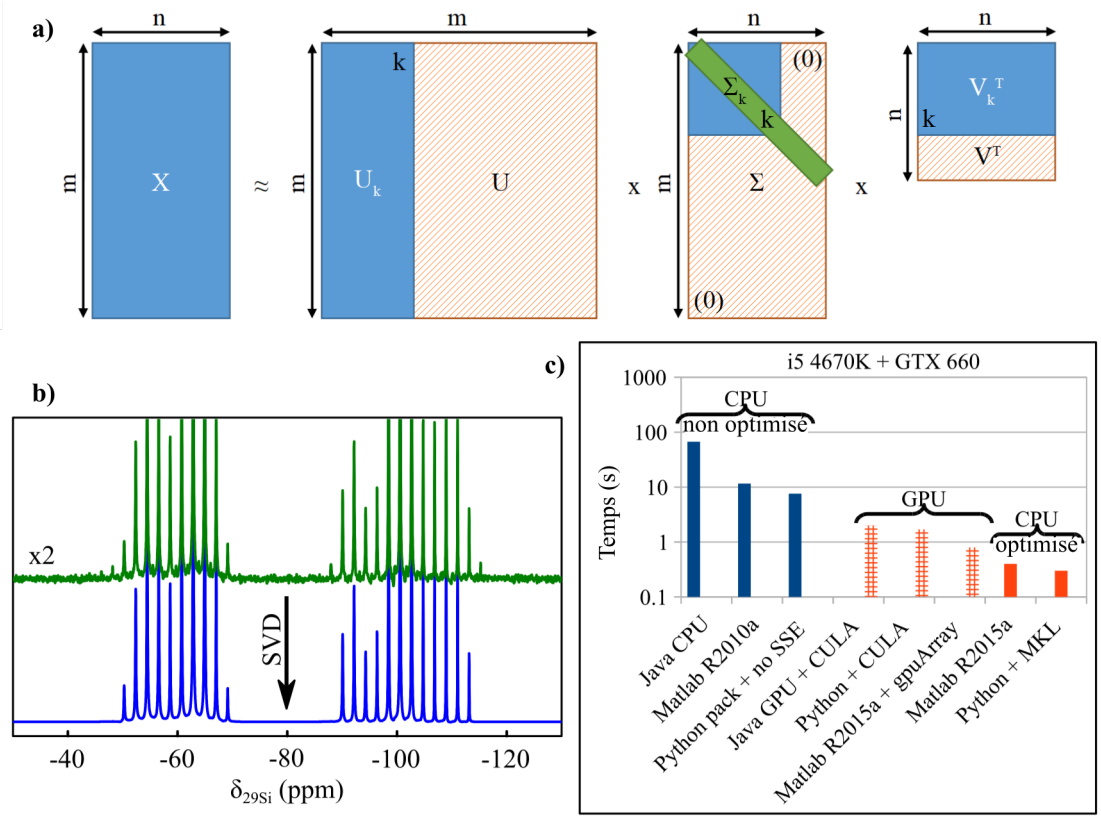


Figure 3: Traitement du signal : a) principe de la SVD (rectangles hachurés oranges), valeurs singulières (diagonale verte) et approximation de rang réduit (rectangles pleins bleus) ; b) débruitage par SVD sur un spectre ^{29}Si CP d'un hybride organique/inorganique ; c) comparaison des temps de calculs pour une matrice de 1025 × 1024 = 1.0×10^6 points. Paramètres : b) B_0 = 7.0 T, rotor de 4 mm, *MAS* = 14 kHz, *RD* = 1 s.

La forme initiale de la matrice conditionne la qualité du débruitage et les meilleurs résultats sont obtenus pour une matrice carrée dans la dimension temporelle. Le seuillage automatique a été implémenté grâce aux indicateurs de Malinowski (*12*). Le premier, *IND*, est basé sur l'écart-type résiduel, tandis que le deuxième, *SL*, est un test de Fisher donnant la probabilité que la valeur singulière appartienne au bruit.

Grâce à la simulation de 6 × 7380 spectres avec un niveau de bruit contrôlé, il a été prouvé que le seuil de détection est de $PSNR_{max} \sim 2$, que ce soit pour des pics lorentziens ou gaussiens. Toutefois, les pics gaussiens sont systématiquement surestimés d'un facteur ~ 20 %, et leur forme est modifiée conduisant à une forme intermédiaire entre gaussienne et lorentzienne. Nous avons appliqué ce débruitage sur des spectre ^{29}Si d'un échantillon sol-gel organique/inorganique, et nous avons confirmé qu'une quantification équivalente au spectre classique pouvait être obtenue en 2,3 fois moins de temps, ce qui a été publié (*13*).

F. Diminuer le temps de calcul

Dans le Chapitre VI, nous avons cherché à diminuer le temps de calcul du SVD qui peut durer plusieurs centaines de secondes. Les premiers tests ont été effectués à l'aide d'un programme Java disponible en ligne, soit en version processeur (CPU), soit en version carte graphique (GPU) (*14*). Seules les GPU fabriqués par Nvidia, disposant de la technologie CUDA, étaient compatibles. Il s'est avéré qu'un GPU d'entrée de gamme de 2008 pouvait exécuter l'opération de SVD en trois fois moins de temps qu'un CPU de milieu de gamme de 2013, et ce, sur la même matrice. Ce résultat était particulièrement surprenant au vu de la puissance de calcul théorique plus importante du CPU en question.

Après avoir comparé des CPU et des GPU s'étalant sur 10 ans et 6 ans, respectivement, nous avons mis en avant que le traitement SVD avec Java sur le CPU était monotâche, tandis qu'un GPU est intrinsèquement massivement parallèle. Afin d'aller plus loin, nous avons utilisé plusieurs versions de Matlab. L'utilisation de l'algorithme "diviser pour mieux régner" s'est avéré cruciale pour utiliser au mieux l'ensemble des cœurs du CPU. Nous avons également mis en avant un croisement entre le temps de calcul sur le CPU et celui sur le GPU, à cause du temps de latence nécessaire au transfert des données vers le GPU.

Une étape supplémentaire a été franchie sous Python en testant différentes bibliothèques et en activant les instructions matérielles SSE3 et AVX2 sur le CPU. Bien que ces instructions soient présentes sur les CPU depuis de nombreuses années, la bibliothèque utilisée n'y fait pas forcément appel. Un gain cumulé d'un facteur 100 en temps de calcul a été obtenu : 9 pour l'algorithme, 3 pour les instructions matérielles, 2 pour la bibliothèque Intel MKL, et 2 avec la simple précision (Figure 3c). Ce travail a donné lieu à publication (*15*).

Par la suite, nous nous sommes intéressés à l'erreur introduite par la simple précision, qui peut s'avérer de fait problématique pour les données couvrant une large gamme dynamique, mais cela doit toutefois être confirmé par des mesures sur des spectres expérimentaux. Nous avons également cherché à utiliser le calcul hybride CPU/GPU.

G. Conclusion

Cette thèse a abordé plusieurs approches complémentaires pour augmenter la

sensibilité de la RMN en phase solide. Les gains obtenus sont résumés dans le Tableau 1. Cela ouvre des portes indéniables en termes d'analyse d'échantillons à la chimie toujours plus compliquée, avec la contrepartie de devoir passer plus de temps à concevoir des programmes et à traiter les données.

Tableau 1: Gain en temps mesuré pour chaque méthode d'amélioration.

Domaine	**Instrumentation**	**Acquisitions rapides**		**Traitement du signal**	
Méthode	MACS	CPMG	NUS	SVD	CPU/GPU
Gain en temps	~ 5	~ 3-100	~ 4	~ 2,3	~ 100

H. Bibliographie

1. L. B. Andreas *et al.*, *Proc. Natl. Acad. Sci.* **113**, 9187–9192 (2016).
2. L. A. Currie, *Anal. Chem.* **40**, 586–593 (1968).
3. N. Priyadarshani, S. Marsland, I. Castro, A. Punchihewa, *PLoS ONE*. **11**, e0146790, 1–26 (2016).
4. S. G. Hyberts, S. A. Robson, G. Wagner, *J. Biomol. NMR*. **55**, 167–178 (2013).
5. A. Shrivastava, V. B. Gupta, *Chron. Young Sci.* **2**, 21–25 (2011).
6. D. Sakellariou, G. Le Goff, J.-F. Jacquinot, *Nature*. **447**, 694–697 (2007).
7. J. A. Lehmann-Horn, J.-F. Jacquinot, J. C. Ginefri, C. Bonhomme, D. Sakellariou, *J. Magn. Reson.* **271**, 46–51 (2016).
8. I. Hung, Z. Gan, *J. Magn. Reson.* **204**, 256–265 (2010).
9. J. Z. Hu, R. A. Wind, *J. Magn. Reson.* **163**, 149–162 (2003).
10. D. Rovnyak, M. Sarcone, Z. Jiang, *Magn. Reson. Chem.* **49**, 483–491 (2011).
11. P. C. Aoto, R. B. Fenwick, G. J. A. Kroon, P. E. Wright, *J. Magn. Reson.* **246**, 31–35 (2014).
12. E. R. Malinowski, *Factor analysis in chemistry* (Wiley, Hoboken, NJ, USA, 3rd ed., 2002).
13. G. Laurent, W. Woelffel, V. Barret-Vivin, E. Gouillart, C. Bonhomme, *Appl. Spectrosc. Rev.* **54**, 602–630 (2019).
14. P. P. Man, C. Bonhomme, F. Babonneau, *Solid State Nucl. Magn. Reson.* **61–62**, 28–34 (2014).
15. G. Laurent, P.-A. Gilles, W. Woelffel, V. Barret-Vivin, E. Gouillart, C. Bonhomme, *Appl. Spectrosc. Rev.* **55**, 173–196 (2020).

Table of contents

List of figures and tables

List of abbreviations

1D, 2D, nD	one-, two-, multi-Dimensional
CS	Compressed Sensing
CSA	Chemical Shift Anisotropy
CP	Cross-Polarisation
CPMG	Carr-Purcell-Meiboom-Gill
CPU	Central Processing Unit
D	Dipolar interaction
DNP	Dynamic Nuclear Polarisation
FT	Fourier Transform
GPU	Graphics Processing Unit
HAp	HydroxyApatite
HETCOR	HETeronuclear CORrelation
HMQC	Heteronuclear Multiple Quantum Correlation
INEPT	Insensitive Nuclei Enhanced by Polarization Transfer
J	Scalar interaction
MACS	Magic Angle Coil Spinning
MAS	Magic Angle Spinning
MaxEnt	Maximum Entropy
MTEOS	MethylTriEthoxySilane
NMR	Nuclear Magnetic Resonance
NUS	Non-Uniform Sampling
PARIS	Phase-Alternated Recoupling Irradiation Scheme
PISSARRO	Phase-Inverted Supercycled Sequence for Attenuation of Rotary ResOnance
PSNR	Peak Signal-to-Noise Ratio
Q	Quadrupolar interaction
R^3	Rotary Resonance Recoupling
REDOR	Rotational Echo Double Resonance
SNR	Signal-to-Noise Ratio
SSB	Spinning SideBand
SVD	Singular Value Decomposition
TEDOR	Transferred-Echo DOuble-Resonance
TEOS	TetraEthylOrthoSilicate
TLR	Transmission Line Resonator
US	Uniform Sampling
Z	Zeeman interaction

Chapter I. General introduction

'Here's how I understand music. If you can play the same bunch of noise twice, it's music. To go beyond that is supercilious and pontificating.'

Johnny Rotten (1956-), singer of Sex Pistols

During the past two decades, developments of solid state Nuclear Magnetic Resonance (NMR) have been particularly spectacular both in terms of new dedicated equipment and methodology. Firstly, ultra-high static magnetic are now available both commercially (28.2 T, 1.2 GHz for ^{1}H, Bruker company) (*1*) and in dedicated institutions (35.2 T, 1.5 GHz for ^{1}H, National High Magnetic Field Laboratory, Tallahassee, USA) (*2*). Such high magnetic fields allow high resolution NMR. The subsequent development of (pre-)amplifiers and probes offers broadband capabilities for the user, including multidimensional and multinuclear experiments. Secondly, Magic Angle Spinning (MAS) is nowadays the most important technique to retrieve high resolution in solid state NMR, usually in combination with decoupling from abundant nuclei (such as ^{1}H and ^{19}F). Ultra-fast rotation frequencies have been reached recently with excellent stability allowing for very efficient averaging of NMR anisotropies in multidimensional experiments: 110 kHz (JEOL, 0.75 mm), 111 kHz (Bruker, 0.7 mm), 170 kHz (Samoson's probe, 0.5 mm, still in development) (*3*). Smaller is the rotor and higher is the maximum spinning speed. As an example, the combination of ultra-fast MAS and ultra-high magnetic field led recently to major progress in the study of biosolids (including proteins, prions, viral capsid…) with inverse (^{1}H) detection as major improvement (*4*, *5*). Moreover, the ultra-fast MAS regime allowed for the development of very efficient low power decoupling methods associated to smart and robust recoupling pulse schemes in the NMR sequences.

Nevertheless, NMR *sensitivity* remains a vivid debate in the community and an unsolved-open question. This point remains nowadays the major NMR drawback at least in three cases (among others). The first case concerns intermediate to low gamma NMR active nuclei for which brute force increase of sensitivity by increasing B_0 is usually of (very) moderate help. The second one corresponds to nuclei subjected to very strong anisotropies (mostly in the case of quadrupolar interaction, if $I > ½$). Even ultra-fast MAS is useless here as numerous overlapped spinning sidebands will make the NMR spectra non tractable: in this case, static experiments have to be preferred even though highly reduced sensitivity is expected. Finally, the third case is much more general and is related to the "limited" number of spins. This can occur for instance with low mass natural-synthetic samples or costly labeled ones.

As a matter of fact, there are several ways to tackle the general problem of NMR sensitivity. Since 2010, the most popular approach has been Dynamic Nuclear Polarization

(DNP) at moderate-high magnetic field which involves new equipment in the laboratories (gyrotrons or klystrons). This is only one option among others, including (i) instrumentation, (ii) fast acquisitions and (iii) signal processing. In this manuscript, points (i), (ii) and (iii) have been considered to improve solid-state NMR sensitivity, focusing on microcoils adapted to mass limited samples and MAS (i), CPMG echoes for 1D experiments and Non Uniform Sampling (NUS) for 2D experiments (ii), and denoising using graphical cards capabilities (iii), respectively. The chemical context will be described in Subchapter E.

Hereafter, the following nomenclature will be followed to differentiate parts of this manuscript. From top level to lowest level, Chapter I, Subchapter A, Section A.1, Subsection A.1.a or A.1.a.i will be used. This chapter is divided in five subchapters, focussed on NMR context (Subchapter A), signal, noise and artefacts (Subchapter B), signal processing in NMR (Subchapter C), outline of this manuscript (Subchapter D), and materials and methods (Subchapter E).

A. Nuclear magnetic resonance

NMR basics will be described in Section A.1, as well its intrinsic limited sensitivity. Specificities of solid-state NMR will be investigated in Section A.2. Finally, the strategies to improve sensitivity with brute force, hyperpolarisation and spin polarisation techniques will be provided in Sections A.3.

A.1. Principle

NMR is a spectroscopy probing chemical environment of atoms up to a few angstroms away. Only magnetically active isotopes, with a nuclear spin quantum $I > 0$, can be analysed by NMR. Hence, $I = 0$ nuclei such as ^{12}C, ^{16}O, and ^{28}Si are simply unobservable by NMR. For spin $I > 0$, $2I + 1$ energy levels are degenerated. $I = ½$ nuclei such as ^{1}H, ^{13}C, ^{29}Si, and ^{31}P are the simplest ones to analyse. $I > ½$ nuclei such as ^{17}O ($I = 5/2$), ^{27}Al ($I = 5/2$), ^{43}Ca ($I = 7/2$), ^{87}Sr ($I = 9/2$), and ^{91}Zr ($I = 5/2$) are denoted quadrupolar ones and their signature is more complex. In this dissertation, spectra of ^{1}H, ^{29}Si, ^{31}P, ^{87}Sr, and ^{91}Zr nuclei will be acquired. In presence of a magnetic field, Zeeman effect occurs, separating energy levels (Figure I.1a) (*6*). A superconducting NMR magnet operating at 16.4 T is presented on (Figure I.1b).

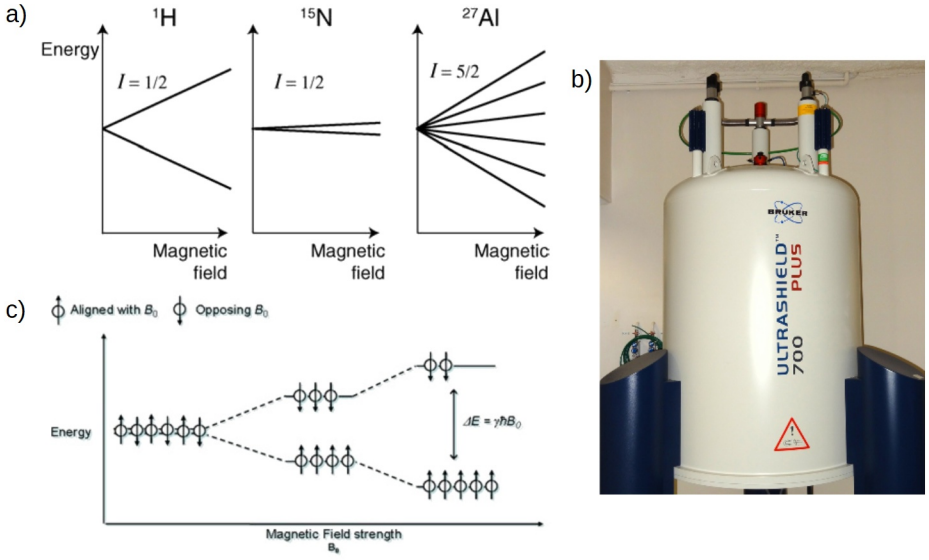


Figure I.1: Zeeman effect: a) energy levels in function of magnetic field and nuclei (*6*), b) 16.4 T NMR magnet (*7*), c) Boltzmann distribution in function of Zeeman effect.

The difference of energy between two successive energy levels is given by:

$$\Delta E = h\nu_0 = \frac{h}{2\pi}\gamma B_0 \qquad \text{I.1}$$

where ν_0 is called the Larmor frequency, γ is the gyromagnetic factor specific of each isotope, and B_0 is the magnetic field. Nuclear spins behave like small magnets and align against magnetic field, similarly to a compass needle giving north of Earth. According to their energy level, they align either on one direction or on the other (Figure I.1c). The probability to measure spin magnetisation on each energy level depends on the temperature and of the energy difference. Their sum gives a macroscopic magnetisation. According to Boltzmann distribution, only one nucleus over 10^5 is visible at room temperature (*6*). Furthermore, isotope natural abundance decreases this value. For instance, ^{13}C, ^{17}O, and ^{29}Si have a natural abundance of 1.1, 0.037, and 4.7 %, respectively. As a consequence, NMR has an intrinsically *low sensitivity*, which is its main drawback.

Macroscopic magnetisation can then be manipulated by applying radio-frequency pulses at Larmor frequency, simultaneously exciting all nuclei of the same kind. During acquisition, the out of equilibrium magnetisation precesses around the magnetic field, which generates a current into the observing coil (Figure I.2a) (*6*), the so-called *Free Induction Decay* (FID). The resulting signal is the sum of all the individual frequencies, arising from the respective shielding constants depending on chemical environment. After Fourier Transform

(FT), the spectrum exhibits multiple peaks whose position is called the chemical shift δ, expressed in ppm, depending on the analysed compound and on the reference compound frequencies, ν and ν_0, respectively:

$$\delta(ppm) = 10^6 \frac{\nu - \nu_0}{\nu_0} \qquad \text{I.2}$$

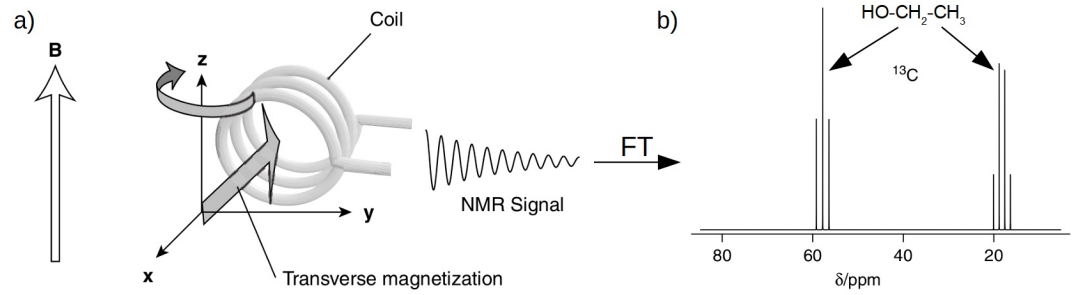


Figure I.2: a) Free induction decay resulting from the macroscopic magnetisation evolving into the observing coil; b) typical ^{13}C solution-state NMR spectrum of ethanol obtained after FT, adapted from (*6*).

For instance, it is possible to discriminate the carbons of ethanol (Figure I.2b). The CH_2 carbon resonates at a higher chemical shift than the CH_3 carbon, due to the proximity with the electronegative oxygen. The fine structure is due to scalar coupling with ^1H nuclei. Moreover, multiple pulses can be combined into a pulse sequence, in order to transfer magnetisation from one spin to another, filter some undesired coherences, excite some directly unobservable coherences, and much more. These NMR capabilities are evidenced by the huge list of experiments acronyms. This is the main advantage of NMR: it is a *powerful* technique (*8*).

A.2. Solid-state NMR

In solution-state NMR, Brownian motion average spin interactions. In solid-state NMR, the number of degrees of freedom is reduced, and the physical interactions detailed hereafter are much less averaged. However, the image of a solid totally immobile is absolutely wrong, but the fluctuation density is considerably reduced. Consequently, broadening of peaks occurs.

Chemical Shift Anisotropy (CSA) is one of the involved interactions, which can be understood by the anisotropic electronic density around a carbonyl bond (Figure I.3a) (*9*). When this ellipsoid is aligned against three directions of space in presence of a magnetic field,

three different chemical shifts are observed. Typical values for ^{13}C are 30 ppm and 100 ppm for aliphatic and aromatic carbons, respectively (*10*).

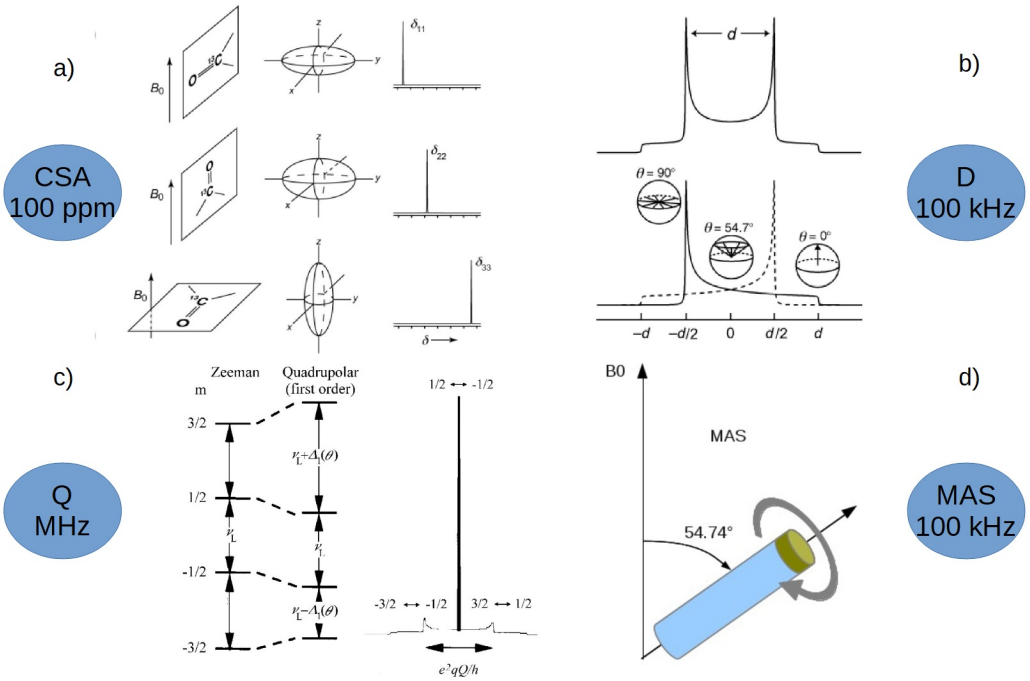


Figure I.3: Anisotropies in solid-state NMR: a) Chemical Shift Anisotropy (CSA) (*9*), b) heteronuclear Dipolar interaction (D) (*11*), c) first order Quadrupolar interaction (Q) for a spin $I = 3/2$ (*12*). d) Averaging by Magic Angle Spinning (MAS). Typical values are indicated in blue ellipses.

Dipolar coupling (D) involves two spins and their respective orientation θ against the magnetic field (Figure I.3b, Equation I.3) (*11*).

$$D \propto \frac{\gamma_1 \gamma_2}{r^3}\left(3\cos^2\theta - 1\right) \qquad \text{I.3}$$

The higher are the γ of the involved nuclei, or the shorter is the distance r between the two spins, the higher is the dipolar interaction. In the case of a spin pair, two symmetric axial shapes are superimposed: a characteristic Pake doublet is obtained (*13*). A mark value to keep in mind is $D_{IA}^{HH} = 122$ kHz, for two immobile ^1H nuclei at 1 Å from each other (*14*).

In the case of quadrupolar nuclei ($I > ½$), ($2I$+1) energy levels have to be considered in the presence of the static magnetic field B_0. In solid state NMR, the central transition ($m = ½ \leftrightarrow m = -1/2$) is not subjected to the first order Quadrupolar interaction (Q), leading to a sharp and intense resonance in the corresponding powder spectrum (Figure I.3c) (*12*). This is no more the case for the so-called satellite transitions ($m \neq \pm 1/2$) which are usually spread

over hundreds of kHz. Each satellite is associated to a lineshape comparable to the one observed for CSA. Q is due to the Electric Field Gradient (EFG) present around the nucleus. In presence of angle and bond distortions, nucleus symmetry is polarized away from a sphere. In the case of "strong" quadrupolar interaction (when compared to the principal Zeeman interaction), second-order perturbation theory has to be applied for accurate calculation of the various energy levels. In this case, even the central transition becomes subjected to spectral broadening and the corresponding lineshape cannot be described by a CSA-like one. First order quadrupolar effect spreads over several megahertz. In extreme cases, such as ^{33}S or ^{127}I, it can be of the same order than Zeeman interaction, and Nuclear Quadrupole Resonance (NQR) reveals to be a suitable alternative (*15–18*).

Magic Angle Spinning (MAS), artificially reintroduces mobility inside the sample (*19–21*). The magic angle θ is defined by Equation I.4:

$$3\cos^2\theta - 1 = 0 \Leftrightarrow \theta = 54.74\,° \qquad \text{I.4}$$

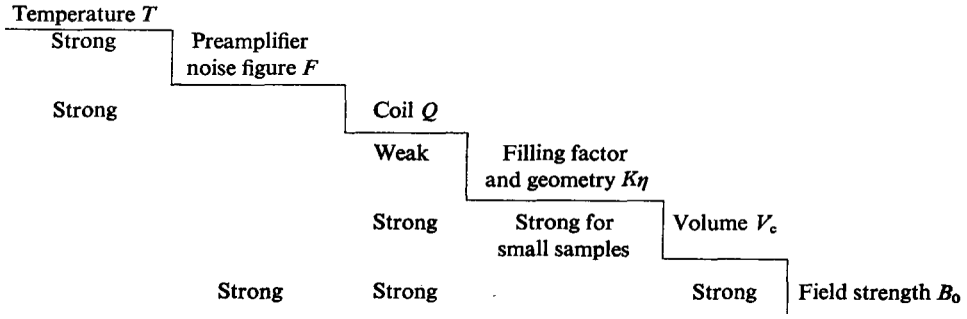
This angle is involved in all first order interactions and the magic value corresponds to the big diagonal of a cube, connecting the two opposite corners. By successively following all three directions of space, efficient averaging of the broadening interactions occurs. With ultra-fast MAS, peaks are considerably narrowed, but to the detriment of sample amount (*22*).

A.3. Sensitivity increase

A.3.a. Brute force

As we said earlier, the main drawback of NMR is its sensitivity, whereas its main advantage is the variety of available experiments. Hoult and Richards summarised in Table I.1 the different parameters on which the experimenter can play to increase sensitivity (*23*), namely temperature, preamplifier noise figure, coil quality factor *Q*, filling factor and geometry, sample volume, and magnetic field strength. Each parameter on the diagonal correlates to the ones on the same row or column. Many of them are interdependent, and can be strongly or weakly coupled.

Table I.1: Interactions of parameters involved in NMR signal-to-noise ratio (*23*).



Nowadays, superconducting magnets are the mean of choice to enhance sensitivity with a high magnetic field, up to soon 28.2 T (*1*). Hybrid resistive/superconducting magnets have to be preferred to further increase magnetic field strength. Such a magnet at 35.2 T is accessible at the National High Magnetic Field Laboratory (NHMFL, Tallahassee, FL, USA) (*2*). However the Signal-to-Noise Ratio (*SNR*) is only moderately increasing with magnetic field strength B_0 to the power 1.5, whereas the increase is faster with gyromagnetic factor γ to the power 2.5 (Equation I.5) (*6*).

$$SNR \propto |\gamma|^{\frac{5}{2}} (B_0)^{\frac{3}{2}} \qquad \text{I.5}$$

Cryoprobes are another technique to increase sensitivity, by cooling down the probe electronic circuit, hence decreasing noise roughly by a factor 3 (*24*).

A.3.b. Hyperpolarisation

Hyperpolarisation consists to circumvent the low macroscopic magnetization imposed by Boltzmann distribution of Zeeman levels, by using a highly polarised source, either polarised light or electrons, albeit transiently. Nikolaou *et al.* reviewed these hyperpolarisation techniques in the context of biomedicine (*25*). Para-hydrogen hyperpolarisation was successfully transferred to ^{1}H, ^{13}C, and ^{15}N via a temporary association to a transition metal centre (*26*). Polarised light magnetisation was transferred to Xenon, giving hyperpolarised Xenon, itself transferred to other nuclei (*27*). Since a few years, Dynamic Nuclear Polarisation (DNP) has gained a lot of interest, either in solution (*28–30*) or in solid-state NMR (*31–33*). A very complete review on DNP applied to solid-state NMR can be found in (*34*) Except for DNP, where commercial equipments are available, other ones are rather confidential, needing home-made designs.

A.3.c. Spin polarisation transfer

Another way to improve sensitivity is spin manipulation. As ^{1}H nucleus gyromagnetic ratio γ is almost the highest one, its magnetisation is proportionally high (Equation I.1). ^{1}H magnetisation can be transferred to low-gamma nuclei X with various techniques. Insensitive Nuclei Enhanced by Polarisation Transfer (INEPT) (*35*) involves through-bonds scalar coupling. Heteronuclear Overhauser Effect SpectroscopY (HOESY) (*36*) can be used to probe through-space proximities in solution-state. The most common spin polarisation transfer method in solid-state NMR is the through space Cross Polarisation (CP) (*37*). This magnetisation transfer allows a maximum theoretical sensitivity gain of γ_{1H} / γ_X. Further gain can be obtained with inverse acquisition on ^{1}H, with the enhancement factor ξ defined in Equation I.6, where W is the full width at half maximum of each nucleus (*38*, *39*).

$$\xi \approx \left(\frac{\gamma_{1H}}{\gamma_X}\right)^{\frac{3}{2}} \left(\frac{W_X}{W_{1H}}\right)^{\frac{1}{2}} \qquad \text{I.6}$$

Polarisation transfer has been often used in solution-state (*40–42*) or in solid-state NMR (*43–45*) and is the basic element of many pulse sequences. In conjunction, heteronuclear decoupling was useful to narrow lines (*46–48*). In solution-state NMR, further improvement was obtained with pulsed field gradients for coherence filtering (*49*). In solid-state NMR, in addition to MAS explained in previous section, further narrowing was obtained with homo-nuclear decoupling such as Lee-Goldburg (*50*) or Decoupling Using Mind-Boggling Optimization (DUMBO) (*51*).

In Subchapter A, we highlighted that NMR is suffering from a low sensitivity, whereas it is a powerful technique. In solid-state NMR, MAS is a useful tool to decrease broadening of peaks. It acts in complement of magnetic field strength increase, CP and decoupling.

B. Signal, noise and artefacts

Not only a weak signal is obtained by NMR on the sample of interest, but noise and artefacts can mask it and have therefore to be controlled. Characteristics of noise will be investigated in Section B.1. Some sources of NMR artefacts will be presented in Section B.2. Finally, signal-to-noise ratio will be defined in Section B.3.

B.1. Noise characteristics

In general in physics and electronics, noise has a random nature with a Gaussian distribution (*52*). Stationary noise exhibits no drift against time. Noise is called ergodic if all statistically possible values are obtained at long times. Finally, noise is homoscedastic if its standard deviation σ is constant against time (*53*). Homoscedastic noise is different from white noise, which will be explained in next paragraph. Histograms are useful to analyse noise distribution. A Gaussian noise is depicted on Figure I.4a (*54*). 68 % of the values are in the interval [-σ; σ], 95 % between [-2σ; 2σ], and 99.7 % of peaks are in the range [-3σ; 3σ]. According to Gaussian probabilities, some rarely events can exceed this range.

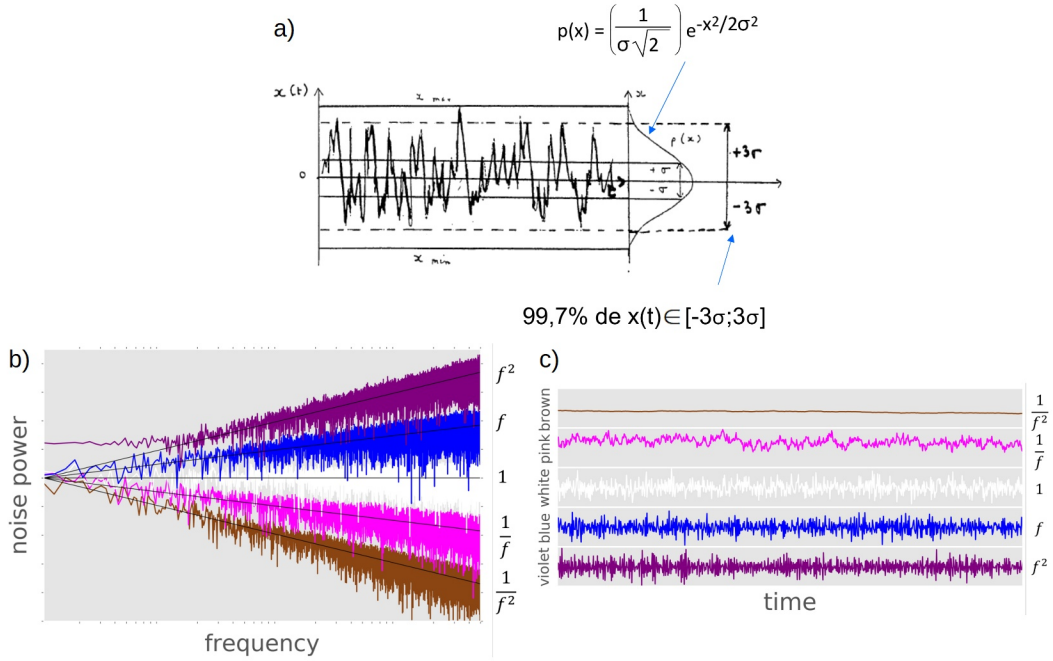


Figure I.4: a) Gaussian noise with probability rule $p(x)$ (*54*); b) Noise colours power in function of frequency; c) repartition of noise against time for each colour (*55*).

Noise can also be described by colours, according to its power spectral density. Figure I.4b presents such an analysis with frequency in logarithmic scale on abscissa and observed power at the corresponding frequency in linear ordinate (*55*). The corresponding time evolutions are presented on Figure I.4c. Brown noise is a low frequency perturbation whose power decreases with the square of the frequency f, giving only small baseline oscillations. Red noise is similar to brown noise, but with non-Gaussian distribution. Pink noise power decreases more slowly at $1/f$ and its time evolution looks smoothed. On the contrary, blue and violet noises are high frequency ones, whose power is increasing with f or

f^2, respectively. The result is the presence of high frequency bursts during evolution time. White noise is very common and its aspect is the most regular of all noise colours. It has no frequency dependence. In normal conditions, without non-uniform sampling, zero-filling, apodisation or artefact, NMR noise is stationary, ergodic, homoscedastic, Gaussian and white (*56*).

In NMR, the main source of noise is thermal fluctuations of electrons inside the coil and the electronic circuit, so-called Johnson noise. Clark thereby stated about an early spectrometer that: 'At maximum gain, the noise of the receiver from 2-30 Mc referred to the input is roughly that of a 1000-Ω resistor at 300 K' (*57*). Johnson noise can be defined as:

$$V_{Johnson}^2 = 4\,R\,k\,T \int_0^\infty \frac{[g(f)]^2}{1+(2\pi f C R)^2}\,df \qquad \text{I.7}$$

where V is the observed voltage, R and C are the circuit resistance and capacitance, respectively, k is the Boltzmann constant, T is the temperature, f is the frequency and g accounts for any amplification or attenuation of the electronics (*58*). Moreover, shot noise is present when a current is flowing into the electronic circuit, independently of temperature and resistance (*58*), with e being the electron charge and I the averaged current:

$$V_{shot}^2 = 2\,e\,I\,R^2 \int_0^\infty [g(f)]^2\,df \qquad \text{I.8}$$

As evidenced here, Johnson noise is proportional to kT. Thus, by decreasing coil and electronics temperature, sensitivity is improved. That is the principle of cryoprobes (*24*). Part of the resistance arises from the sample itself whose temperature can also be lowered, limiting the induced noise. That is an additional benefit of solid-state DNP which is usually done at 100 K (*59*). This gain comes in addition to the improved macroscopic magnetisation at low temperature according to Boltzmann distribution.

B.2. Artefacts

Artefacts are unexpected signals. On the contrary to noise, they are limited to specific parts of the spectrum but can mask signal peaks as well. They can arise from instance from electronic imperfections (Figure I.5a) (*60*). When an offset is present on digitaliser (top left panel), a transmitter spike is present on spectrum (top right panel), and when real and imaginary channels are non-equivalently balanced (bottom left panel), a mirror image appears (bottom right panel). This has been addressed by the Cyclically Ordered Phase Sequence

(CYCLOPS) pulse program (*61*). These artefacts are almost negligible on modern NMR spectrometers. Moreover, many experiences use 180° pulses, such as J-resolved spectroscopy, which can give phantom and ghost peaks (Figure I.5b), arising from miscalibrated pulses, but eliminated by the EXORCYCLE (*62*).

t_1-noise is a scintillation artefact proportional to the signal, characterised by vertical lines on intense peaks of two-dimensional spectra, and occurring when small variations of pulses lengths or phases are experienced (*63*). Such a trace is shown on left of Figure I.5c for a Dipolar Heteronuclear Universal-Quantum Correlation (D-HUQC) (*64*). This can be minimised by decreasing the time between two successive scans (right spectrum). More generally, t_1-noise can be reduced by improving RF stability, for instance by sample and RF coil temperature regulation. Artefacts can also be induced by data processing, as for covariance spectroscopy, where diagonal, peak-ridge and peak-free noise was highlighted (Figure I.5d) (*65*).

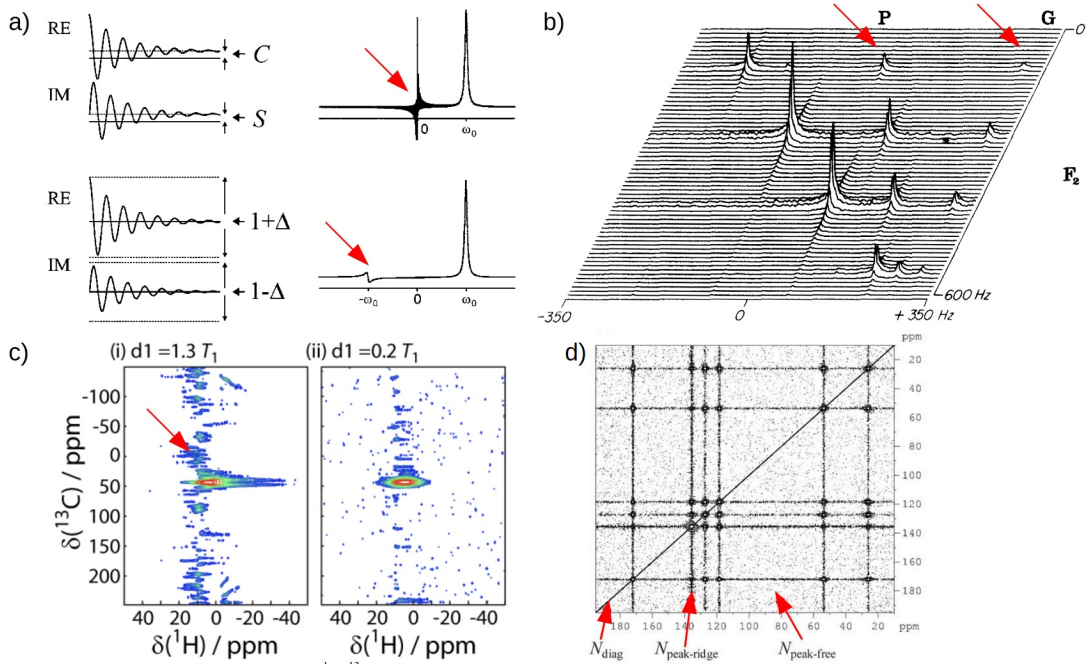


Figure I.5: Artefacts present on NMR spectra: a) transmitter spike (top right) and mirror image (bottom right) and their origin (left panel), adapted from (*60*); b) phantom (P) and ghost (G) on 2D J-resolved spectra, adapted from (*62*); c) t_1-noise on 2D D-HUQC for two acquisition rates, adapted from (*64*); d) diagonal, ridge and peak-free noise obtained by covariance, adapted from (*65*). Red arrows highlight artefacts.

Furthermore, other artefacts sources are present. In particular, NMR is working in the radio frequency domain, and the 100 MHz range is especially sensitive to Frequency

Modulated (FM) emissions. This corresponds to ^{13}C or ^{27}Al nuclei at a magnetic field of 9.4 T. Such perturbations are especially visible on old-fashioned unshielded magnets. Even on modern spectrometers, this effect is easily observable at low temperature, in only 30 s of acquisition *without* any nuclear or electron excitation, due to an increased probe quality factor (Figure I.6) (*30*). These FM radios are good examples of band-specific artefacts.

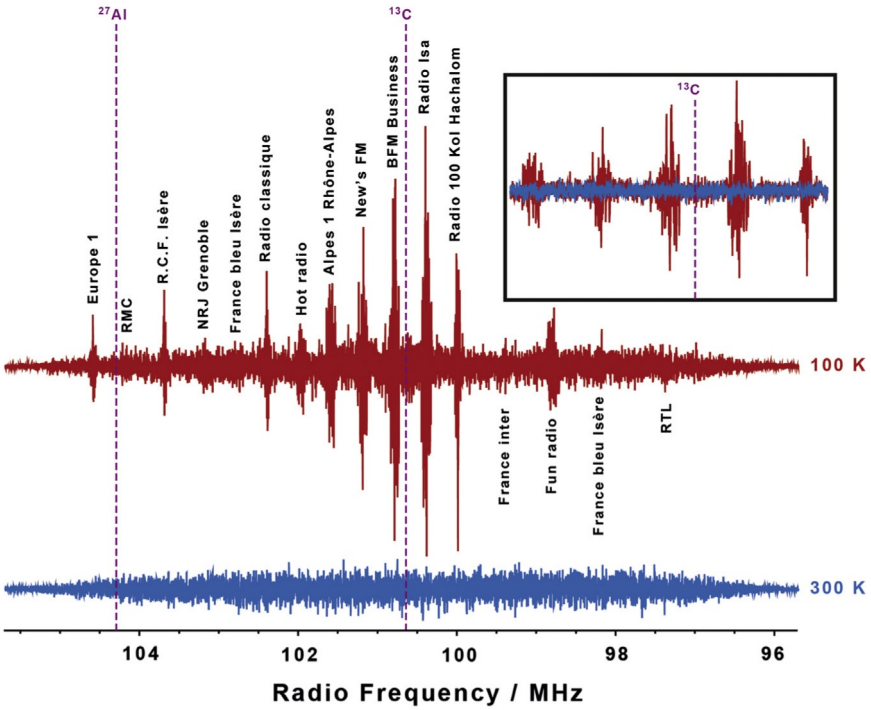


Figure I.6: Acquisition of sample-free NMR signal, without pulse applied by the spectrometer, in the 100 MHz range corresponding to ^{27}Al and ^{13}C nuclei at B_0 = 9.4 T (dashed vertical lines) at 300 K (in blue) and at 100 K (in red). The insert is a zoom on ^{13}C region. Vertical names correspond to attributed FM radios emitting at Grenoble, France (*32*).

B.3. Signal-to-noise ratio

Sensitivity can be defined as the capacity to distinguish slight differences of sample concentration. It is intimately correlated to the detection limit (*66*). L. Currie mathematically formalised three thresholds: critical limit to take a decision of presence when peak position is already known (L_c), detection limit when peak position is unknown (L_d) and quantification limit (L_q), corresponding to 1.64, 3.29 and 10 σ_{noise}, respectively, where σ_{noise} is noise standard deviation. Signal-to-Noise Ratio (*SNR*) can be defined in various ways, either with signal standard deviation (σ_{signal}) or with signal peak height (H_{signal}). In the latter case, the acronym *PSNR* is preferred. Noise and artefacts can be measured by their Root Mean Square (*RMS*) or by their peak-to-peak value ($h_{noise\_peak\_peak}$).

In mathematics and electronics (*67*) $$SNR = \frac{\sigma^2_{signal}}{\sigma^2_{noise}}$$ I.9

In spectroscopies[1] (*56*) $$PSNR_{rms} = \frac{H_{signal}}{\sigma_{noise}}$$ I.10

In analytical chemistry (*68*) $$PSNR_{max} = \frac{H_{signal}}{h_{noise\_peak\_peak}/2}$$ I.11

In case of Gaussian noise, $PSNR_{rms} \sim 3.3\ PSNR_{max}$. Despite $PSNR_{rms}$ is commonly used in NMR, this definition is not suitable in presence of artefacts, and $PSNR_{max}$ formula has to be preferred. Moreover, an important warning has to be made about detection limit. If *SNR* is under detection level and signal is artificially increased by a non-linear processing, such as a wrongly calibrated maximum entropy algorithm, no detection limit enhancement will be obtained, despite *SNR* improvement (Figure I.7) (*69*). Further details on *SNR* will be provided in Subsection B.3.c of Chapter V.

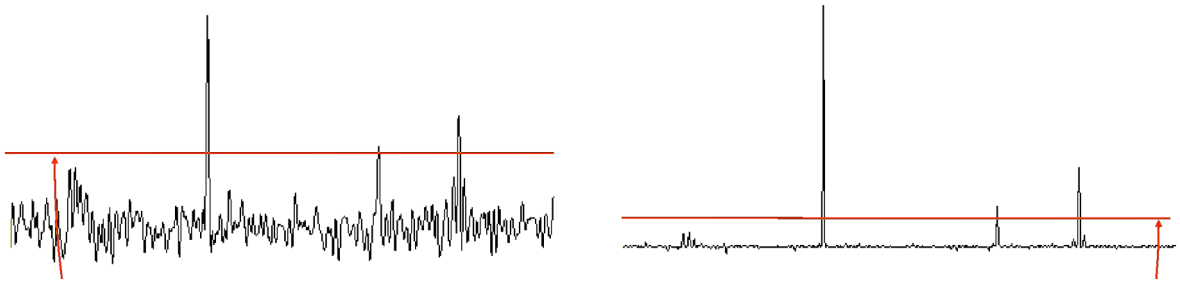


Figure I.7: Influence of linear (left) and non-linear processing (right): signal and noise were both scaled down, but small peaks on the left of the spectrum are still under detection threshold (in red), adapted from (*69*).

In Subchapter B, we highlighted that NMR noise was white and Gaussian and we described Johnson and shot noise. Different NMR artefacts were shown, including transmitter spike, mirror image, phantoms, ghosts, t_1-noise, covariance noise, and band-specific noise. Signal-to-noise formulas were presented.

C. Signal processing in NMR

Signal processing is an extremely vast domain regrouping all electronic, mathematical, or computer operations, posterior to data acquisition (*70–72*), which can be provided by all kinds of captors. Despite raw data can not be modified, signal processing can strongly attenuate noise and artefacts. This branch of science has applications in industrial, medical or

[1] In NMR, standard deviation of noise is doubled for historical reasons.

military domains. We can cite for instance real-time error control in a chemical reactor, where flow rate, temperature, pressure and composition were regulated (*73*). Other examples are wearable sensors to monitor individual's state of health (*74*), parent-infant interaction (*75*), acoustic remote sensing (*76*), or millimetre wave multi-points communication (*77*). In NMR, signal processing involves many steps, such as reception chain (Section C.1), apodisation (Section C.2), Fourier transform (Section C.3) and phasing (Section C.4).

C.1. From probe to spectrum

NMR signal processing can be decomposed in many steps (*78*, *79*). Some of them are executed automatically by electronics on raw signal (Free Induction Decay, FID) out of spectrometer, from probe to disk recording. These operations are listed in Table I.2 in order of execution and are presented on Figure I.8. Once the FID has been recorded in electronics domain, the user needs to execute additional actions to exploit data, firstly in time domain, secondly in frequency domain. Some of them are optional. Significant hardware improvement has been achieved in terms of *SNR*, since *ca.* 1990 (Figure I.9). Hereafter are highlighted three essential steps of NMR signal processing: apodisation, Fourier transform and phasing.

Table I.2: Signal processing steps in NMR.

Electronics domain	**Time domain**	**Frequency domain**
Analog filtering	Oversampled points suppression	Phasing
Pre-amplification	Aberrant points removal	Baseline correction
Amplification	Apodisation	Chemical shift calibration
Phase Shift Quadrature (PSD)	Denoising	Peaks detection
Demodulation	Linear prediction	Integration
Oversampled digitalisation (ADC) (*80*)	Zero-filling	Spectrum saving
Digital filtering	Fourier Transform (FT)	Figure preparation
FID disk recording		

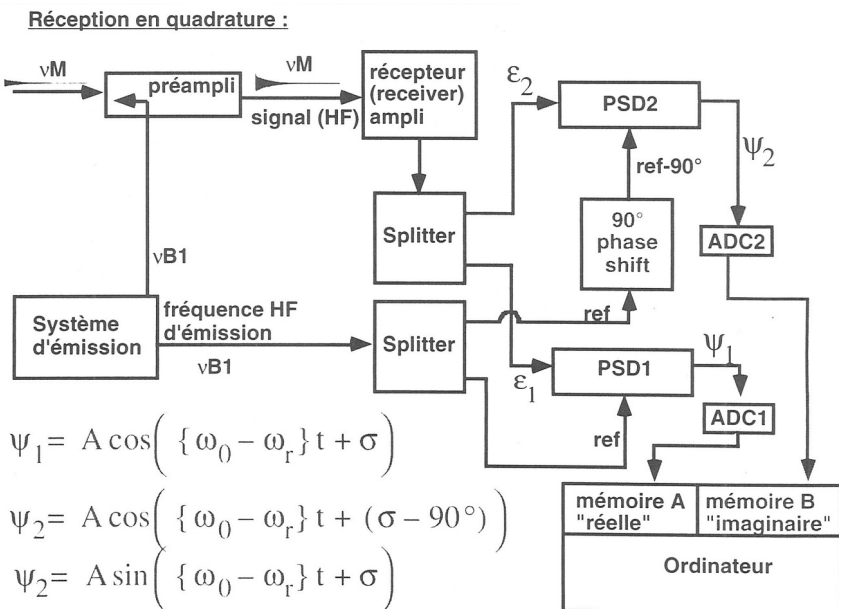


Figure I.8: Electronics signal processing during NMR FID acquisition (*81*).
ADC: Analog to Digital Converter; PSD: Phase Shift Detector.

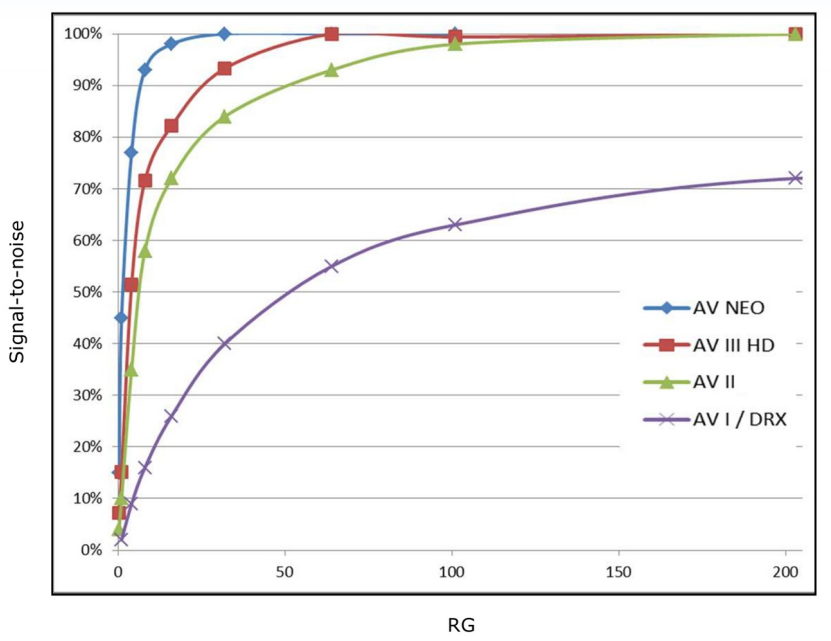


Figure I.9: Hardware improvement of *SNR* against receiver gain (*RG*), for Bruker NMR spectrometers released from *ca.* 1990 (Avance I) to 2017 (Avance Neo) (*82*).

C.2. Apodisation

An important step of processing is apodisation, which usually takes place in time domain. However, Spencer demonstrated the equivalence of this filter in both time and frequency domains (*83*). The first objective is to decrease truncation effects on FID, which

give cardinal sine oscillations, so-called *wiggles*. By multiplying FID with a decreasing function, these oscillations at peak foot are attenuated but peaks are broaden.

The second role of apodisation is to attenuate noise. While signal is preponderant at the beginning of the FID, noise is proportionally bigger at the end. Thereby, multiplying the FID by a mathematical function will have an important effect on *SNR* after FT. Several decreasing functions can be used: exponential, Gaussian, linear or cosine. Figure I.10 shows that exponential is convex, cosine is concave, and Gaussian is intermediate, as compared to linear curve. The optimal *SNR* can be obtained by multiplying the FID by a decreasing exponential with a constant equal to the natural decrease of the FID, the so-called *matched filter*, thus doubling the average peak width (*78*). However, better wiggles suppression is obtained using a cosine function, which is more abrupt at the end of the FID (*84*).

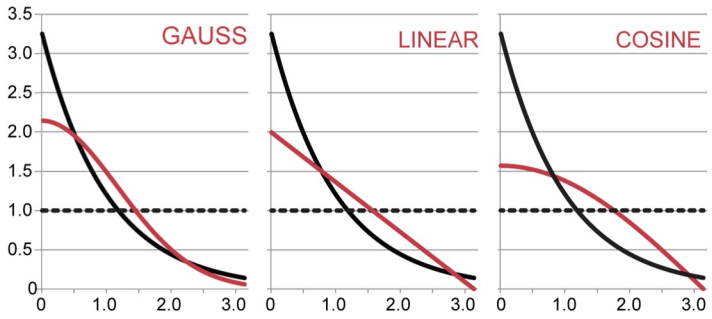


Figure I.10: Normalised apodisation density of Gaussian, linear and cosine (in red) as compared to a decreasing exponential (in black), adapted from (*85*).

Apodisation can also improve resolution, at the expense of *SNR*, by giving a higher weight to the middle of the FID, for instance using shifted sine bell (*86*) or Traficante (*87*) functions, or even Lorentzian/Gaussian conversion (*88*). The latter is the product of an increasing exponential function and a shifted decreasing Gaussian function.

C.3. Fourier transform

Another crucial step of signal processing is FT (*89*, *90*), which converts time t to frequency ν, *i.e.*, FID to spectrum (Equation I.12).

$$F(\nu) = \int_{-\infty}^{+\infty} f(t)\, e^{-i2\pi\nu t}\, dt \qquad \text{I.12}$$

Similarly, it converts reciprocal space to real space, as in Magnetic Resonance Imaging (MRI) or in X-Ray Diffraction (XRD). Continuous FT calculates the integral

according to points spacing, whereas discrete FT converts a series of *N* points, into another series of the same number of points (Equation I.13).

$$F(\nu_k)=\frac{1}{\sqrt{N}}\sum_{n=0}^{N-1} f(n)\,\mathrm{e}^{-i2\pi\nu_k\frac{n}{N}},\ k\in[0,N-1] \qquad \text{I.13}$$

Data unit is totally ignored during discrete FT, implying regularly spaced data. The formula presented is the normalised one with a $1/\sqrt{N}$ factor (*91*). Baseline distortions are avoided by halving the first point before FT. This compensates for the number of inter-points delays in the FID being one less than the number of points (*92*). Moreover, fast FT (FFT) has been developed with the Cooley-Tukey algorithm (*93*), which explains why a power of two is usually chosen for zero-filling. Finally, from a programming point of view, an additional inversion step of left and right side of spectrum is necessary, to reorder the spectrum after FFT. In order to not confuse discrete FT with Density Functional Theory (DFT), which we have not used, this acronym will be avoided.

C.4. Phasing

Phasing is the third important step, as it will influence spectrum quantification. Three kinds of phases are present: zero-, first- and second-order phase. While zero-order phase changes all peak shape simultaneously, first- and second-order phase distort baseline. Zero-order phase depends on phase cycling in pulse programs and on the entire hardware emission/reception signal chain. It can be influenced for instance by probe frequency tuning and impedance matching. First order phase corresponds to a delay, for instance the *dead-time* before signal acquisition. This delay of a few microseconds is necessary to let emission signal decrease, which is ranging from volts to hundreds of volts, before observing reception signals, of only microvolts. It can be corrected using Equation I.14 (*92*):

$$phase_1=-180*\frac{dead\ time}{dwell\ time} \qquad \text{I.14}$$

where $phase_1$ is the first order phase correction, and *dwell time* is the delay between two acquired points. As highlighted by Levitt, this is a first approximation and off-resonance effect or peak superposition will avoid exact first order phase correction (*6*). Second order phase is present for wideline experiments (*94*).

Signal processing investigated in Subchapter C is a crucial step to obtain usable data, with automatic operations done by the electronics, and manual operations done by the experimenter, such as apodisation, FT, and phasing.

D. Dissertation outline

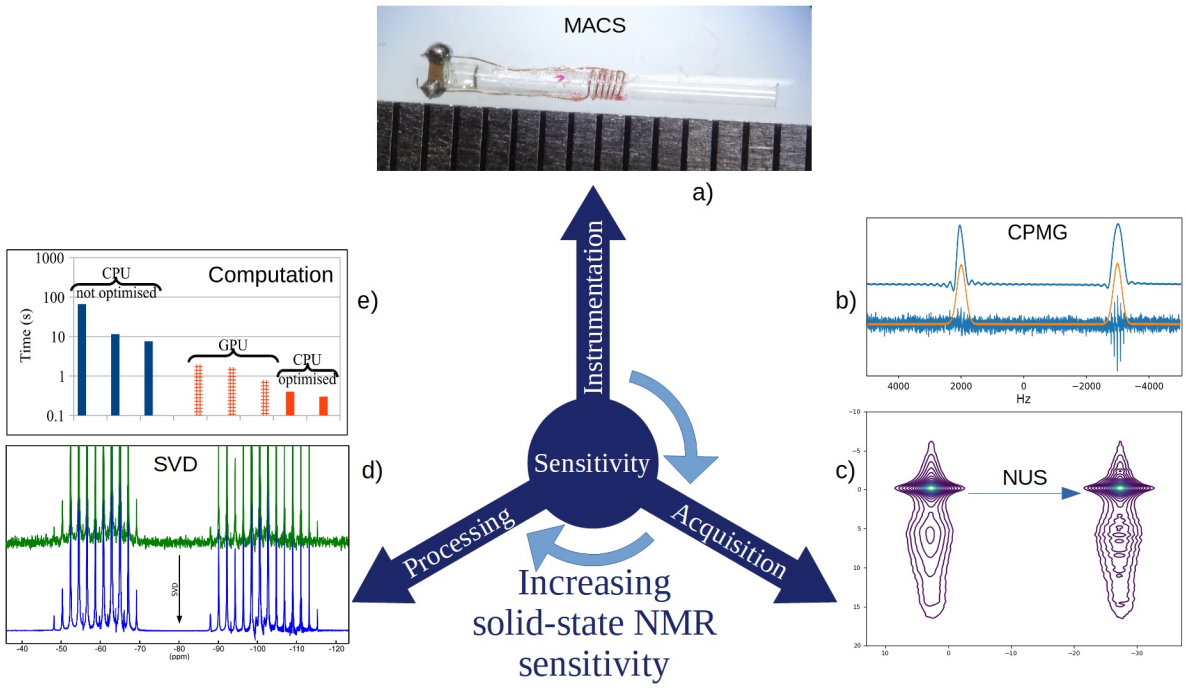


Figure I.11: Outline of this dissertation: a) microcoils with Magic Angle Coil Spinning (MACS), b) Carr-Purcell-Meiboom-Gill (CPMG) echoes, c) Non-Uniform Sampling (NUS), d) denoising with Singular Value Decomposition (SVD), e) computation time with Central Processing Unit (CPU) and Graphics Processing Unit (GPU).

Figure I.11 presents the outline of this dissertation. Three approaches were investigated to improve solid-state NMR sensitivity. Firstly, *instrumentation* was used to limit sample amount, with microcoils placed into the rotor (Magic Angle Coil Spinning, MACS, Figure I.11a, Chapter II). A time gain of ~ 5 was obtained. Secondly, the total *acquisition* duration was decreased with Carr-Purcell-Meiboom-Gill echoes (CPMG, Figure I.11b, Chapter III), and with Non-Uniform Sampling (NUS, Figure I.11c, Chapter IV). A time gain of up to ~ 100 was highlighted for CPMG for suitable samples. An improved reconstruction algorithm was provided. NUS revealed useful on multi-dimensional spectra, with hybrid sampling for broad peaks and a time gain of ~ 4 on 2D. Finally, *processing* was investigated with denoising by Singular Value Decomposition (SVD, Figure I.11d, Chapter V), leading to time decrease by a factor ~ 2.3. SVD computation was speed up by a factor 100 and compared

on Central Processing Unit (CPU) and Graphics Processing Unit (GPU) (Figure I.11e, Chapter VI). A thorough bibliography on developed subjects will be provided in each chapter.

E. Materials and methods

The main goal has been to optimize the sensitivity in several chemical contexts and for materials of highest interest. Three main categories of materials have been considered as illustrations of sensitivity enhancement. The first one is related to HydroxyApatite, HAp, an hydroxylated calcium phosphate which is the main inorganic component in mammal's bones and teeth. The complexity of biological HAp is of both chemical (multiple anionic - CO_3^{2-},… and cationic - Na^+,… substitutions) and structural (crystalline *vs* amorphous components) natures. HAp can also be used for depollution, through its high capability of ions substitution. The highest sensitivity is therefore necessary for optimal description of this complex material. Apart HAp, collagen corresponds to the most important protein associated to bones. Gelatin is obtained by denaturation of collagen. The second topic deals with the large family of hybrid materials containing organic and inorganic (mostly silica) components. Such hybrid materials are obtained as bulk or thin membranes by using sol gel (low temperature) chemistry. These hybrid membranes can have interesting applications as part of functional materials for energy. In this case, the measurement of the condensation rate of silica species is of paramount importance for the materials scientist using ^{29}Si MAS experiments. The highest NMR sensitivity is therefore needed for optimal time gain. A third category concerns bioactive glasses containing ^{87}Sr quadrupolar nuclei, which have to be well-characterised before considering using them for medical applications. For instance, strontium rich glasses can limit osteoporosis. The involved nuclei combine large anisotropy, low gamma and low natural abundance, hence the need to increase sensitivity.

In this section, we emphasize on the synthesis of the samples used here for NMR sensitivity purposes. During this work, we tested our methodology on already well-characterised samples, in order to not question ourselves about chemical shift assignments, but rather to focus on sensitivity issues.

E.1. ^{13}C enriched carbonated hydroxyapatite

^{13}C labelled nano-crystalline nonstoichiometric carbonated HAp were synthesized in order to characterize this particular anionic substitution of HAp (*95*). This sample exhibited

narrow lines in ^{1}H, ^{13}C and ^{31}P nuclei and was used for NUS experiments (Figure IV.24 of Chapter IV). It was prepared by wet chemical reaction between calcium nitrate tetrahydrate ($Ca(NO_3)_2 \cdot 4H_2O$) and diammonium hydrogenophosphate ($(NH_4)_2HPO_4$) at pH = 10 (adjusted by the addition of NH_4OH). 0.15 M of sodium bicarbonate ($NaH^{13}CO_3$, 99 % ^{13}C) was dissolved in 100 mL of the 0.3 M $(NH_4)_2HPO_4$ aqueous solution, so that the carbonate ions could be incorporated into the HAp lattice structure. The $(NH_4)_2HPO_4$ aqueous solution was added slowly to 100 mL of the 0.5 M $Ca(NO_3)_2.4H_2O$ solution at room temperature under stirring (24 h). Then, the calcium phosphates were separated from the suspension by centrifugation, washed with distilled water, and then dried at 105°C for 24 h. The calcium phosphate powders were heated at the temperature of 400°C for 1 h in an electric furnace in order to eliminate relatively unstable CO_3^{2-} sites near the surface.

E.2. Amorphous/crystalline hydroxyapatite sample

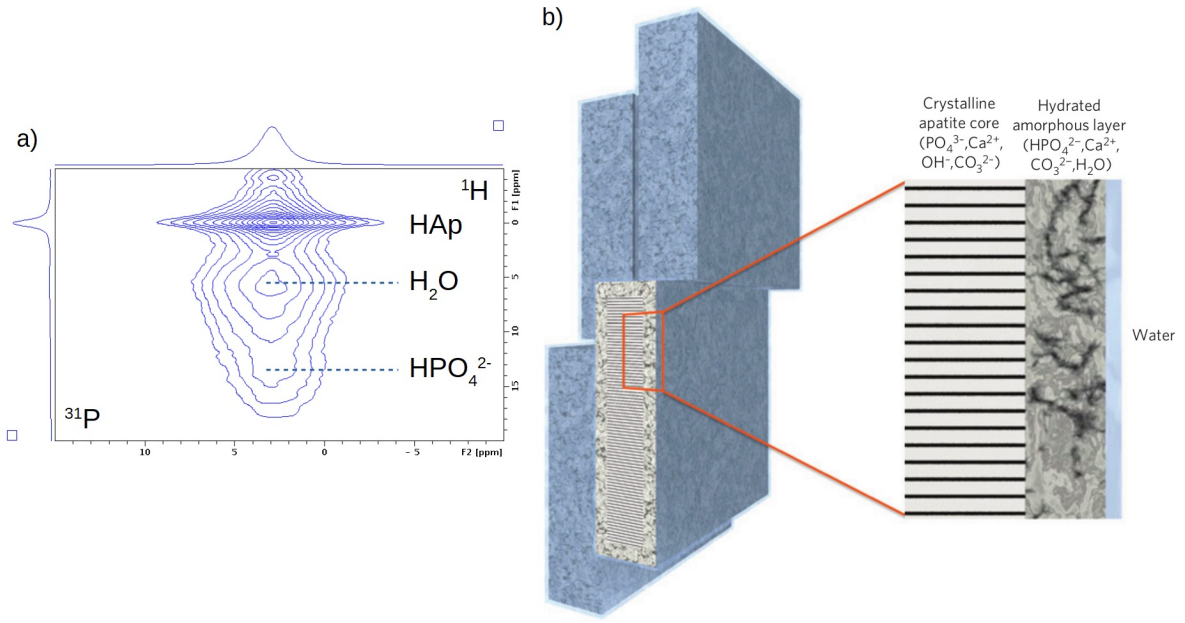


Figure I.12: Amorphous/crystalline HAp: a) $\{^{1}H\}$-^{31}P 2D CP HECTOR spectrum highlighting HAp crystalline peak and amorphous components constituted of adsorbed water and hydrogenophosphate; b) schematic drawing of HAp platelets in bone, with the crystalline HAp core and the hydrated amorphous layer (*96*). Parameters: B_0 = 7.0 T, *MAS* = 14 kHz, *RD* = 1 s, *NS* =16, $TD(^{1}H)$ = 170, t_{exp} = 0.75 h, t_{cp} = 1 ms, $\nu_{decoupling}(^{1}H)$ = 53 kHz SPINAL-64.

Another HAp synthesis is described in this section to obtain nanosized HAp particles especially interesting from the NMR point of view, as the chosen sample exhibit both a sharp line (due mainly to isolated ^{1}H species in HAp) and much broader ones (associated to disordered components exhibiting large distributions of chemical shift). This remark will be

of prime importance when implementing NUS (Chapter IV). A 2D ^{1}H-^{31}P spectrum is presented on Figure I.12a. The sharp line is easily visible at $\delta_{1H} = 0$ ppm, whereas the broad components are a sensitivity test, especially the HPO_4^{2-} one at $\delta_{1H} \sim 13$ ppm. This sample is a model of bone inorganic phase, where crystalline HAp platelets are surrounded by amorphous calcium phosphate (Figure I.12b) (*96*). HAp is also a constituent of Randall's plaque, the starting point of some kidney stones. Hence, HAp will act as well as a test compound for the set-up of MACS methodology (Chapter II).

$CaCl_2.2H_2O$ (calcium chloride dihydrate, ≥ 99 %), $NaH_2PO_4.H_2O$ (sodium phosphate monohydrate, ≥ 99.1 %), $NaHCO_3$ (sodium bicarbonate, ≥ 99.5 %), $Ca(NO_3)_2.4H_2O$ (calcium nitrate tetrahydrate, ≥ 99.0 %) and $(NH_4)_2HPO_4$ (diammonium hydrogenphosphate, ≥ 99.99 %) were purchased from Sigma-Aldrich. Sample was prepared following the HA-2 synthesis described by Nassif *et al.* (*97*). Experiments were carried out at room temperature (22 ± 1°C). A solution containing 110 mM $CaCl_2$, 33 mM NaH_2PO_4, and 33 mM $NaHCO_3$ was prepared in 500 mM acetic acid. The pH was adjusted to 2.2 with hydrochloric acid. Two flasks (35 mL, h = 50 mm) containing these solutions (20 mL) and covered by perforated Parafilm (to slow down the gas diffusion) were placed into a closed chamber (1000 cm^3). The HAp Precipitation was triggered via the slow increase of solution pH caused by the vapours of fresh ammonia aqueous solution (30 % w/w, 8 mL) in the chamber. A few hours after ammonia introduction, precipitation occurs in the solution. After 6 days the gas diffusion was considered complete (pH ~ 10-11), and the solids were washed and centrifuged (6000 rpm, 10 min) firstly in distilled water and then in ethanol to remove the non-precipitated salts. The recovered crystals were dried at 37°C for 7 days before characterization. The obtained chemical formula is $Ca_{10-x}(PO_4)_{6-x}(CO_3)_x(OH)_{2-x}$ with $0 \leq x \leq 2$.

E.3. Gelatin

Gelatin is the denatured form of collagen, a biopolymer present in many tissues, such as cornea, tendon, skin and bone (*98*). Gelatin was used as a solid-state sample with a high mobility for $\{^{1}H\}$-^{13}C INEPT 2D NUS analysis (Figure IV.22 of Chapter IV). Commercial gelatin A from porcine skin (Sigma-Aldrich, ref. G2625, 300 g bloom) was purified, lyophilised and solubilised at 150 $mg.mL^{-1}$ into 0.5 M acetic acid. The mixture was heated at 65°C under stirring and then kept 1 hour at 45°C to facilitate its solubilization. The solution was then stored at 4°C. Gelatin concentration in acidic solutions was determined by assessing the amount of hydroxyproline.

E.4. Sea urchin spines

Sea urchin are biomineralised marine organisms. Their spines have an axial symmetry promising to use with MACS (Chapter II). Both the body shell and the spines are calcium carbonate, although with different microstructures and magnesium content (*99*). During biomineralisation process, the mineral phase evolves from amorphous calcium carbonate to crystalline calcite. Under X-ray diffraction, mineral phase behaves as single crystals. The organic phase in skeletal elements is constituted from proteins, monosaccharides and pigments. Intercrystalline organics can be removed by chemical treatment or biochemical digestion. However, intracrystalline organics are entrapped into the mineral and heat treatment is necessary to degrade them.

Sea urchin specimens of *Paracentrotus lividus*, collected in the Atlantic Ocean, were received from the Roscoff marine station of Sorbonne Université (France). Large and small spines were cut off from three different sea urchins collected in the same place during the same period of the year. Small and large spines, being on average, respectively, 0.29 ± 0.09 cm and 1.17 ± 0.12 cm long, were taken from the same sea urchin specimens.

E.5. 50:50 MTEOS:TEOS sample

For ^{29}Si solid-state NMR studies on CPMG, SVD and computation (Chapters III, V and VI) we used a typical sol-gel sample, consisting of methyltriethoxysilane (MTEOS, T species) and tetraethylorthosilicate (TEOS, Q species) in a 50:50 MTEOS:TEOS molar ratio. Its synthesis will be described in Section B.2.a of Chapter V, as part of one of our articles. Its typical ^{29}Si Cross Polarisation (CP) MAS NMR non-quantitative spectrum is represented on Figure I.13. This isotope has a natural abundance of 4.7 %. ^{29}Si chemical shift is calibrated to TetraMethylSilane (TMS) at $\delta = 0$ ppm and is sensitive to the first and second neighbours. For tetravalent silicon atoms, if all their first neighbours are oxygen (Quadri-functional silicon, Q units), their position is at a chemical shift of $\delta \sim$ -100 ppm. If one oxygen is replaced by a carbon (Tri-functional silicon, T units), the chemical shift is $\delta \sim$ -60 ppm. Moreover, species are differentiated according to their second neighbour: for each Si-O-Si bond, the peak is shifted by $\delta \sim$ -10 ppm. The superscript of T and Q units corresponds to the number of condensed bonds (*100*).

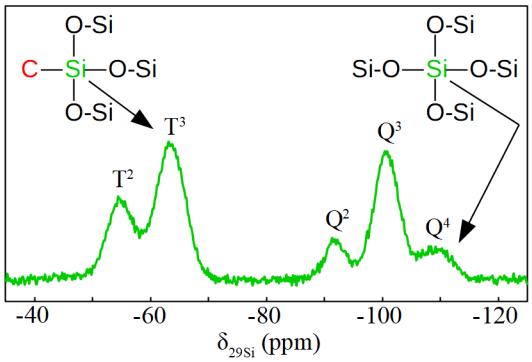


Figure I.13: Typical ^{29}Si CP MAS NMR spectrum with the attributed species. Parameters: B_0 = 7.0 T, *MAS* = 14 kHz, *RD* = 1 s, *NS* = 2048, t_{exp} = 0.57 h, t_{cp} = 5 ms, $\nu_{decoupling}(^1H)$ = 2.4 kHz SPINAL-64.

E.6. CSPTC:TEOS:PVDF-HFP sample

This sample was used for ^{29}Si solid-state NMR, as a limit case needing a longer acquisition time of a 1.5 day with CPMG (Figure III.13b of Chapter III). Silicon species were diluted into this sample as compared to the 50:50 MTEOS:TEOS one. Reactants are presented on Figure I.14a (*101*). PolyVinyliDeneFluoride-HexaFluoroPropylene (PVDF-HFP) was dissolved in N,N-Dimethylformamide (DMF) at 75°C and cooled down to room temperature. This PVDF-HFP/DMF solution was added and mixed into a hybrid membrane precursor containing TEOS and 2-(4-ChloroSulfonylPhenyl)-ethylTriChlorosilane (CSPTC) in a 2:1 molar ratio. Solutions was prepared at room temperature and stirred at 70°C for 3 h in a closed glass vial. After cooling down solution to room temperature, it was prepared by electrospinning. This technique consisted to inject a precursor solution into a syringe with a high voltage between the needle and a collector (Figure I.14b). A continuous fibre was obtained and formed a three dimensional interconnected network.

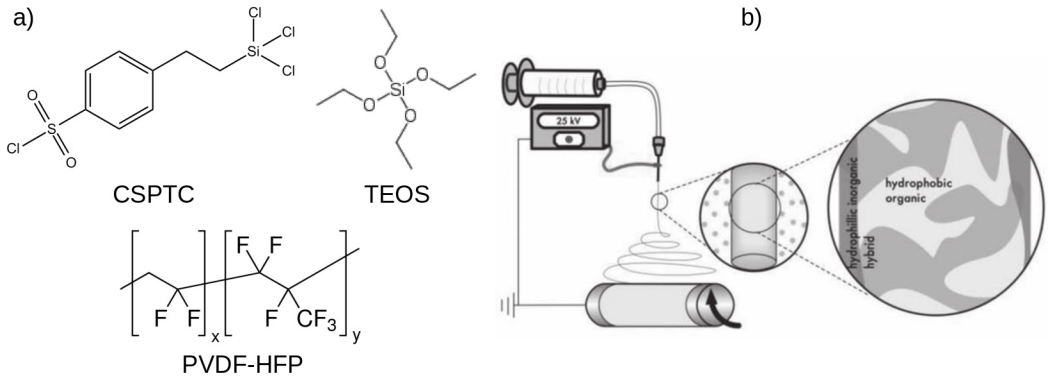


Figure I.14: a) Reactants and b) electrospinning method used to prepare the CSPTC:TEOS:PVDF-HFP sample (*101*).

E.7. Mesoporous silica SBA-15

SBA-15 materials belongs to the class of ordered mesoporous silicas, which can entrap organic molecules into their porosity, such as anti-inflammatory drugs, promoting control released into the body. They are also chemically inert, biocompatible and easily functionalised by grafting groups changing surface polarity (*102*). SBA-15 will be studied with SVD denoising (Figure V.12 of Chapter V). Tetraethoxysilane (TEOS, reagent grade, 98%), the surfactant P_{123} of molecular weight 5800 g.mol^{-1} and formula CH_3-(CH_2-CH_2-O)$_{20}$-(CH_3-CH-CH_2-O)$_{70}$-(CH_2-CH_2-O)$_{20}$-H were provided by Sigma Aldrich and were used as received. The HCl used was an ACS grade Carlo Erba, 35% in volume. SBA-15 was prepared in HCl aqueous solutions, then dried in air. TEOS over P_{123} molar ratio was 62.5, with a volume of 167.5 ± 2.5 mL of TEOS and a weight of 72 ± 0.5 g of P_{123}. The sample was obtained in HCl 0.3 mol.L^{-1} and washed with 1 L of distilled water for 50 g of solid. Calcination was performed in air at 500 °C (rate of 2 °C.min^{-1}) for 6 h.

E.8. Strontium bioactive glass

This sample was used for GPU denoising a dataset with a high number of points (Figure VI.2b of Chapter VI). Strontium is an important element for bone growth and to prevent osteoporosis. A 10 % strontium content was chosen for this bioactive glass to release Sr^{2+} ions at physiological concentration (*103*). The targeted glass composition of the glass named B75-Sr10 was 75.0SiO_2−15.0CaO−10.0SrO (wt %) (or 77.4SiO_2−16.6CaO−6.0SrO (mol %)). It was prepared by mixing 13.94 mL of TEOS, 3.158 g of $Ca(NO_3)_2{\cdot}4H_2O$, and 1.021 g of strontium nitrate $Sr(NO_3)_2$ in ethanol in the presence of deionized water and 2 N HCl under ambient pressure and temperature. The quantities of reactants were calculated using a H_2O:TEOS molar ratio of 12:1 and a H_2O:HCl volume ratio of 6:1. A 50 mL low viscosity sol was obtained and stirred for 1 h. The prepared sol was then transferred to airtight PTFE moulds in an oven at 60°C for gel formation and ageing. After 24 h, the obtained gels were heated at 125°C for another 24 h, ground to powder, and then stabilized at 700°C to eliminate nitrate and for further densification. Chemical content was measured by ICP-AES as: 74.9SiO2−16.0CaO−9.1SrO (wt %) (or 76.9SiO2−17.6CaO−5.5SrO (mol %)).

E.9. Zirconia ceramic

Zirconia ceramic has strong mechanical properties. It constitute MAS rotors material. ^{91}Zr quadrupolar nucleus (I = 5/2) is a low-gamma nuclei with a natural abundance of 11.22 %. This fast relaxing compound was suitable to tune VOCS-DFS-WURST-QCPMG. A single VOCS step was used in Chapter III (Figure III.13c).

E.10. NMR parameters

NMR acquisition parameters will be provided in figures legends, with magnetic field strength (B_0), MAS rate (*MAS*), recycle delay (*RD*), number of scans (*NS*), time dimension (*TD*), experiment time (t_{exp}), spectral width (*SW*), and contact time (t_{CP}).

F. Chapter bibliography

1. Bruker Corporation (2019), available at https://ir.bruker.com/press-releases/press-release-details/2019/Bruker-Announces-Worlds-First-12-GHz-High-Resolution-Protein-NMR-Data/default.aspx.
2. Z. Gan *et al.*, *J. Magn. Reson.* **284**, 125–136 (2017).
3. A. Samoson, *J. Magn. Reson.* **306**, 167–172 (2019).
4. L. B. Andreas *et al.*, *Proc. Natl. Acad. Sci.* **113**, 9187–9192 (2016).
5. H. W. Orton *et al.*, *Angew. Chem. Int. Ed.* **59**, 2380–2384 (2020).
6. M. H. Levitt, *Spin dynamics: basics of nuclear magnetic resonance* (John Wiley & Sons Ltd, Chichester, England, 2nd ed., 2008).
7. M. Pegeot, PhD dissertation (Université de Grenoble, 2014), available at https://tel.archives-ouvertes.fr/tel-01315172.
8. S. E. Ashbrook, J. M. Griffin, K. E. Johnston, *Annu. Rev. Anal. Chem.* **11**, 485–508 (2018).
9. L. Emsley, D. D. Laws, A. Pines, *Proc. Int. Sch. Phys. Enrico Fermi*. **139**, 45–210 (1999).
10. O. N. Antzutkin, Y. K. Lee, M. H. Levitt, *J. Magn. Reson.* **135**, 144–155 (1998).
11. D. D. Laws, H.-M. L. Bitter, A. Jerschow, *Angew. Chem. Int. Ed.* **41**, 3096–3129 (2002).
12. A. Medek, L. Frydman, *J. Braz. Chem. Soc.* **10**, 263–277 (1999).
13. G. E. Pake, *J. Chem. Phys.* **16**, 327–336 (1948).
14. K. Schmidt-Rohr, H. W. Spiess, *Multidimensional Solid-State NMR and Polymers* (Elsevier, 2012).
15. E. L. Hahn, B. Herzog, *Phys. Rev.* **93**, 639–640 (1954).
16. T. P. Das, A. K. Saha, *Phys. Rev.* **98**, 516–524 (1955).
17. T. Poumeyrol, PhD dissertation (Université d'Orléans, Orléans, France, 2013), available at http://www.theses.fr/2013ORLE2075.
18. P. M. J. Szell, D. L. Bryce, *Concepts Magn. Reson. Part A*. **45A**, e21412, 1–11 (2016).
19. E. R. Andrew, A. Bradbury, R. G. Eades, *Nature*. **182**, 1659–1659 (1958).
20. I. J. Lowe, *Phys. Rev. Lett.* **2**, 285–287 (1959).
21. J. Schaefer, E. O. Stejskal, *J. Am. Chem. Soc.* **98**, 1031–1032 (1976).
22. U. Sternberg *et al.*, *J. Magn. Reson.* **291**, 32–39 (2018).

23. D. I. Hoult, R. E. Richards, *J. Magn. Reson. 1969*. **24**, 71–85 (1976).
24. P. Styles *et al.*, *J. Magn. Reson. 1969*. **60**, 397–404 (1984).
25. P. Nikolaou, B. M. Goodson, E. Y. Chekmenev, *Chem. Weinh. Bergstr. Ger.* **21**, 3156–3166 (2015).
26. R. W. Adams *et al.*, *Science*. **323**, 1708–1711 (2009).
27. H. W. Long *et al.*, *J. Am. Chem. Soc.* **115**, 8491–8492 (1993).
28. J. H. Ardenkjær-Larsen *et al.*, *Proc. Natl. Acad. Sci.* **100**, 10158–10163 (2003).
29. M. Batel *et al.*, *Chem. Phys. Lett.* **554**, 72–76 (2012).
30. A. Comment, *J. Magn. Reson.* **264**, 39–48 (2016).
31. A. J. Rossini *et al.*, *J. Am. Chem. Soc.* **134**, 16899–16908 (2012).
32. D. Lee, S. Hediger, G. De Paëpe, *Solid State Nucl. Magn. Reson.* **66–67**, 6–20 (2015).
33. C. Leroy *et al.*, *Anal. Chem.* **89**, 10201–10207 (2017).
34. A. S. Lilly Thankamony, J. J. Wittmann, M. Kaushik, B. Corzilius, *Prog. Nucl. Magn. Reson. Spectrosc.* **102–103**, 120–195 (2017).
35. G. A. Morris, R. Freeman, *J. Am. Chem. Soc.* **101**, 760–762 (1979).
36. L. E. Overman, T. C. Malone, G. P. Meier, *J. Am. Chem. Soc.* **105**, 6993–6994 (1983).
37. S. R. Hartmann, E. L. Hahn, *Phys. Rev.* **128**, 2042–2053 (1962).
38. Y. Ishii, R. Tycko, *J. Magn. Reson.* **142**, 199–204 (2000).
39. A. Venkatesh *et al.*, *J. Phys. Chem. A*. **122**, 5635–5643 (2018).
40. A. Bax, S. Subramanian, *J. Magn. Reson. 1969*. **67**, 565–569 (1986).
41. W. F. Reynolds, R. G. Enríquez, *J. Nat. Prod.* **65**, 221–244 (2002).
42. C. M. Quinn, M. Wang, T. Polenova, in *Protein NMR* (Humana Press, New York, NY, 2018), *Methods in Molecular Biology*, pp. 1–35.
43. A. Lesage, D. Sakellariou, S. Steuernagel, L. Emsley, *J. Am. Chem. Soc.* **120**, 13194–13201 (1998).
44. Z. Gan, J. P. Amoureux, J. Trébosc, *Chem. Phys. Lett.* **435**, 163–169 (2007).
45. Y. Ishii *et al.*, *J. Magn. Reson.* **286**, 99–109 (2018).
46. R. R. Ernst, *J. Chem. Phys.* **45**, 3845–3861 (1966).
47. A. E. Bennett, C. M. Rienstra, M. Auger, K. V. Lakshmi, R. G. Griffin, *J. Chem. Phys.* **103**, 6951–6958 (1995).
48. P. K. Madhu, *Isr. J. Chem.* **54**, 25–38 (2014).
49. R. E. Hurd, *J. Magn. Reson. 1969*. **87**, 422–428 (1990).
50. M. Lee, W. I. Goldburg, *Phys. Rev.* **140**, A1261–A1271 (1965).
51. D. Sakellariou, A. Lesage, P. Hodgkinson, L. Emsley, *Chem. Phys. Lett.* **319**, 253–260 (2000).
52. A. Louis-Joseph, presented at the Réseau des électroniciens - Notions d'analyse et de traitement du bruit dans les systèmes électroniques et instrumentation scientifique (CNRS DR4, Île-de-France, France, 2017), available at http://www.electroniciens-dr4.u-psud.fr/download.php?file=BruitsAnalyseALJ_Public.pdf&dossier=Formations/2017%20Journee%20bruit.
53. O. M. Kvalheim, Frode. Brakstad, Yizeng. Liang, *Anal. Chem.* **66**, 43–51 (1994).
54. M. Lours, presented at the Réseau des électroniciens - Notions d'analyse et de traitement du bruit dans les systèmes électroniques et instrumentation scientifique (CNRS DR4, Île-de-France, France, 2017), available at http://www.electroniciens-dr4.u-psud.fr/download.php?file=Caracterisation%20du%20bruit%20en%20instrumentation%20Temps-Frequence.pdf&dossier=Formations/2017%20Journee%20bruit.
55. J. A. Doyle, A. C. Evans, (2018), available at http://arxiv.org/abs/1806.03704.
56. S. G. Hyberts, S. A. Robson, G. Wagner, *J. Biomol. NMR*. **55**, 167–178 (2013).
57. W. G. Clark, *Rev. Sci. Instrum.* **35**, 316–333 (1964).

58. D. V. Perepelitsa, (Department of Physics, MIT, MA, USA, 2006), available at http://web.mit.edu/dvp/Public/noise-paper.pdf.
59. H. Takahashi *et al.*, *Angew. Chem. Int. Ed.* **51**, 11766–11769 (2012).
60. D. Reichert, G. Hempel, *Concepts Magn. Reson.* **14**, 130–139 (2002).
61. D. I. Hoult, R. E. Richards, *Proc. R. Soc. Lond. Math. Phys. Eng. Sci.* **344**, 311–340 (1975).
62. G. Bodenhausen, R. Freeman, D. L. Turner, *J. Magn. Reson. 1969*. **27**, 511–514 (1977).
63. A. F. Mehlkopf, D. Korbee, T. A. Tiggelman, R. Freeman, *J. Magn. Reson. 1969*. **58**, 315–323 (1984).
64. F. A. Perras, M. Pruski, *J. Magn. Reson.* **298**, 31–34 (2019).
65. Y. Li *et al.*, *Analyst*. **138**, 2411–2419 (2013).
66. M. Zehfus, lecture notes (Chem 434, Black Hills State University, SD, USA, 2013), available at https://www.bhsu.edu/Portals/91/InstrumentalAnalysis/StudyHelp/LectureNotes/Chapter1.pdf.
67. N. Priyadarshani, S. Marsland, I. Castro, A. Punchihewa, *PLoS ONE*. **11**, e0146790, 1–26 (2016).
68. A. Shrivastava, V. B. Gupta, *Chron. Young Sci.* **2**, 21–25 (2011).
69. J. C. Hoch, presented at the Biomolecular NMR: modern tools for data processing and interpretation dynamics (Gothenburg, Sweden, 2017).
70. A. Boujrad, T. Legou, presented at the Réseau des électroniciens (Orsay, France, 2008), available at http://www.electroniciens-dr4.u-psud.fr/download.php?file=TNS.pdf&dossier=Formations/2010%20Traitement%20du%20signal/Presentation.
71. C. L. Byrne, lecture notes (Department of Mathematical Sciences, University of Massachusetts Lowell, USA, 2013), available at http://faculty.uml.edu/cbyrne/SP1text.pdf.
72. T. O'Haver, (Department of Chemistry and Biochemistry, University of Maryland, College Park, MD, USA, 2015), available at http://bit.ly/1TucWLf.
73. S. Yin, H. Luo, S. X. Ding, *IEEE Trans. Ind. Electron.* **61**, 2402–2411 (2014).
74. W. Gao *et al.*, *Nature*. **529**, 509–514 (2016).
75. M. Avril *et al.*, *Front. Psychol.* **5**, 1437, 1–14 (2014).
76. D. R. Dowling, K. G. Sabra, *Annu. Rev. Fluid Mech.* **47**, 221–243 (2015).
77. R. W. Heath, N. González-Prelcic, S. Rangan, W. Roh, A. M. Sayeed, *IEEE J. Sel. Top. Signal Process.* **10**, 436–453 (2016).
78. J. C. Lindon, A. G. Ferrige, *Prog. Nucl. Magn. Reson. Spectrosc.* **14**, 27–66 (1980).
79. D. N. Rutledge, *Signal treatment and signal analysis in NMR* (Elsevier Science, 1996), vol. 18 of *Data Handling in Science and Technology*.
80. M. A. Delsuc, J. Y. Lallemand, *J. Magn. Reson. 1969*. **69**, 504–507 (1986).
81. J. C. Beloeil, presented at the DNTS en méthodes physico-chimiques d'analyses (Orsay, France, 1999).
82. F. Decker, presented at the Bruker NMR Users' Meeting (Edinburgh, Scotland, UK, 2017).
83. R. G. Spencer, *Concepts Magn. Reson. Part A*. **36A**, 255–265 (2010).
84. P. Giraudeau, S. Akoka, *Magn. Reson. Chem.* **49**, 307–313 (2011).
85. M. R. Palmer *et al.*, *J. Phys. Chem. B*. **119**, 6502–6515 (2015).
86. R. G. Brereton, M. J. Garson, J. Staunton, *J. Magn. Reson. 1969*. **43**, 224–233 (1981).
87. D. D. Traficante, G. A. Nemeth, *J. Magn. Reson. 1969*. **71**, 237–245 (1987).
88. A. G. Ferrige, J. C. Lindon, *J. Magn. Reson. 1969*. **31**, 337–340 (1978).
89. E. O. Brigham, R. E. Morrow, *IEEE Spectr.* **4**, 63–70 (1967).

90. B. Osgood, lecture notes (Electrical engineering department, Stanford university, USA, 2007), available at https://see.stanford.edu/materials/lsoftaee261/book-fall-07.pdf.
91. J. O. *III* Smith (1993), available at https://www.dsprelated.com/freebooks/mdft/Normalized_DFT.html.
92. D. Marion, *Concepts Magn. Reson. Part A*. **40A**, 326–340 (2012).
93. J. W. Cooley, J. W. Tukey, *Math. Comput.* **19**, 297–301 (1965).
94. I. Hung, Z. Gan, *J. Magn. Reson.* **204**, 256–265 (2010).
95. F. Babonneau, C. Bonhomme, S. Hayakawa, A. Osaka, A. Osaka, in *MRS Online Proceedings Library* (2006), vol. 984, pp. MM06-05, 1–6.
96. Y. Wang *et al.*, *Nat. Mater.* **12**, 1144–1153 (2013).
97. N. Nassif *et al.*, *Chem. Mater.* **22**, 3653–3663 (2010).
98. F. Portier *et al.*, *Langmuir*. **33**, 12916–12925 (2017).
99. M. Albéric *et al.*, *Cryst. Growth Des.* **18**, 2189–2201 (2018).
100. R. A. Assink, B. D. Kay, *J. Non-Cryst. Solids*. **99**, 359–370 (1988).
101. V. Maneeratana *et al.*, *Adv. Funct. Mater.* **23**, 2872–2880 (2013).
102. B. Dragoi *et al.*, *J. Sol-Gel Sci. Technol.* **91**, 552–566 (2019).
103. C. Bonhomme *et al.*, *J. Am. Chem. Soc.* **134**, 12611–12628 (2012).

Chapter II. Microcoils

'Experience is simply the name we give our mistakes.'

Oscar Wilde (1854-1900), writer

A. Chapter introduction

In this Chapter, we tackle the question of NMR sensitivity when facing the problem of intrinsic limited number of spins (in other words, m < 100 µg). This situation is rather classical in biological NMR (bio-solids NMR for instance, where ^{13}C or ^{15}N enrichment is performed) but can occur in materials science as well. As an example, we consider first the field of thin silica-based organic-inorganic hybrid films obtained by sol-gel chemistry and various coating processes (*1–3*). A review of such Evaporation Induced Self-Assembly (EISA) techniques was proposed by Sanchez *et al*. (*4*). This field is of particular importance as such films may exhibit tailored and optimized properties such as hydrophobicity, anti-fogging, anti-fouling etc. (Figure II.1a) (*5*). We give now an order of magnitude of the usual corresponding mass: considering a 300 nm thick silica film deposited on a substrate of 2 cm^2, it corresponds to a volume of ~ 60 nL, or a weight of ~ 130 µg (the density of silica is taken here as ~ 2.2 $g.cm^{-3}$). As a consequence, and using *standard* NMR probes, the only option is to collect a large number of individual hybrid films, especially when targeting low sensitive nuclei such as ^{29}Si or ^{13}C. Such an approach is highly laborious and not satisfactory from an NMR point of view: the materials science community is clearly avid for more sensitive methodologies.

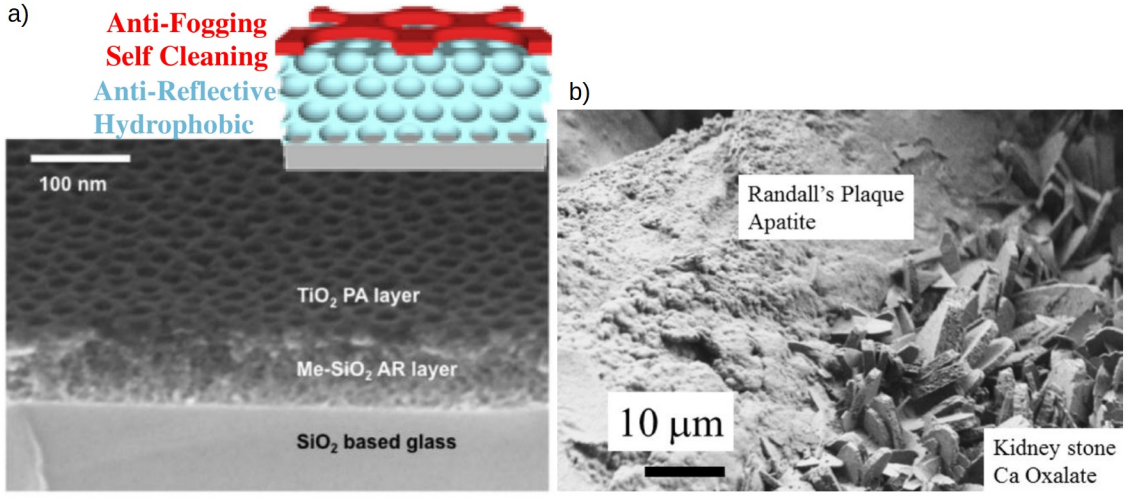


Figure II.1: Potential applications of microcoils: a) sol-gel organic-inorganic hybrid silica film deposited on a substrate (*5*), b) Randall's plaque (apatite) in a kidney stone (calcium oxalate) (*6*). In both cases, sample mass is lower than 100 µg.

In a drastically different context, the problem of mass limitation can be encountered in the case of natural samples. As an example, HydroxyApatite (HAp) is ubiquitous in human

bones and teeth (non-pathological calcification)… but also as a representative of pathological calcifications (or kidney stones). The chemical structure of kidney stones corresponds mainly to calcium oxalates at various level of hydration but a (small) nucleus of HAp is usually associated. Such an HAp nucleus is called the Randall's plaque and is considered as the architecture from which the growth of the calcium oxalate stones starts (Figure II.1b) (*6*). The mass of the Randall plaque is estimated to ~ 70-90 µg (this mass largely depends on the considered kidney stone). Currently, several groups conduct studies on the structure of the Randall plaque worldwide, including colleagues (biologists and nephrologists) from the Tenon hospital in Paris.

All in all, the problem of mass limited samples and the associated low NMR sensitivity is a great challenge. We have decided to focus on instrumentation developments involving mainly the Magic Angle Coil Spinning (MACS) first developed by Sakellariou *et al*. in 2007 (7). The key concept is to strongly enhance the filling factor by using microcoils, whose inner diameter should be smaller than 1 mm (*8*, *9*), keeping high resolution capabilities by fast spinning at the Magic Angle. The goal is clearly to analyze samples with restricted mass: our first example is related to sea urchin spine (calcium carbonate) which exhibits a cylindrical symmetry compatible with the geometry of a micro-solenoid. Secondly, solenoids and Transmission Line Resonators (TLR) will be considered and compared in terms of sensitivity by using HAp as a test sample. Due to the fragility of the micro-coils, low power 1D and 2D NMR pulse sequences were tested as well. A funding was provided by French Research National Agency (ANR) from 2015 to 2018, in the project 'Micro-detection by solid-state Nuclear Magnetic Resonance' (MicrogramNMR), in which I was implicated.

Available techniques suitable for the solid-state NMR studies of microquantities will be presented in Subchapter B. Electromagnetism background and microcoils manufacturing will be explained in Subchapters C and D, respectively. Subchapter E will be devoted to the obtained results.

B. Current methods for analysing microquantities by solid state NMR

In this subchapter, after investigating commercial probes capabilities (Section B.1), we will focus on microcoils (Section B.2) and their applications (Section B.3) before detailing Magic Angle Coil Spinning (MACS) concept (Section B.4).

B.1. Commercial solid-state NMR probes

In solid-state NMR, the sample is inserted into a zirconia rotor under Magic Angle Spinning (MAS), which averages anisotropic interactions and narrows lines (Section A.2 of Chapter I). Each manufacturer has his own rotors dimensions. In his pioneering work (*10*), Barbara, who was a member of Varian manufacturer team, misleadingly employed the term microprobe to study susceptibility averaging in 4 mm MAS rotors. These associated coils are too wide to be defined as microcoils. Bruker rotors outer diameter are ranging from 7.0 mm to 0.7 mm, with spinning frequencies from 7 to 111 kHz, respectively, and with sample volumes from 250 to 0.4 µL, respectively (Table II.1 and Figure II.3b) (*11*, *12*). Jeol manufacturer obtained a similar result with a 0.75 mm rotor containing 0.29 µL and spinning at 110 kHz (*13*). Up to now, the maximum achievable MAS frequency is 170 kHz with a 0.5 mm non-commercial probe (*14*). Spinning rate is limited by two factors: centrifugal forces applied on rotor walls, and speed of sound for the surrounding gas (*15*). The counterpart of spinning speed increase is sample volume decrease. If enough sample is available, the biggest rotor has to be chosen, in adequacy with the experiment requirements. Hence, for quantitative analysis of nuclei with low anisotropies, such as ^{29}Si with 4.7 % natural abundance, a moderate MAS rate of 5 kHz is enough at $B_0 = 7.0$ T for all silicates. For nuclei with much higher anisotropies, such as ^{207}Pb (chemical shift anisotropy), ^{27}Al (quadrupolar interaction) or ^{1}H (homonuclear broadening) as dedicated examples, faster spinning rotors have to be used, with the corresponding lost of sensitivity. According to Nishiyama (*16*), sample volume is decreasing as d^{3}, where d is the rotor diameter, whereas absolute Peak-Signal-to-Noise Ratio (*PSNR*) is decreasing approximately as $d^{3/2}$, for filled rotors. Hence, *PSNR* is decreasing slower than sample volume. The corresponding curves, for Bruker and Jeol rotors, are presented on Figure II.2.

Table II.1: Comparison of Bruker MAS rotors sizes, properties, and applications.
a: according to (*11*, *12*); b: according to cortecnet.com; c: Samoson's probe (*14*);
#N/A: not available.

Rotor size (mm)	7.0	4.0	3.2	2.5	1.9	1.3	0.7	0.5[c]	Capillary
MAS rate (kHz)	7	15	24	35	42	67	111	170	5-15
Volume (µL)[a]	250	70	30	14	10.4	1.5	0.4	0.2	0.2
Price (€)[b]	437	616	1230	1282	1427	2000	3274	#N/A	low
Designation	MAS	fast MAS		very fast MAS		ultra fast MAS			MACS

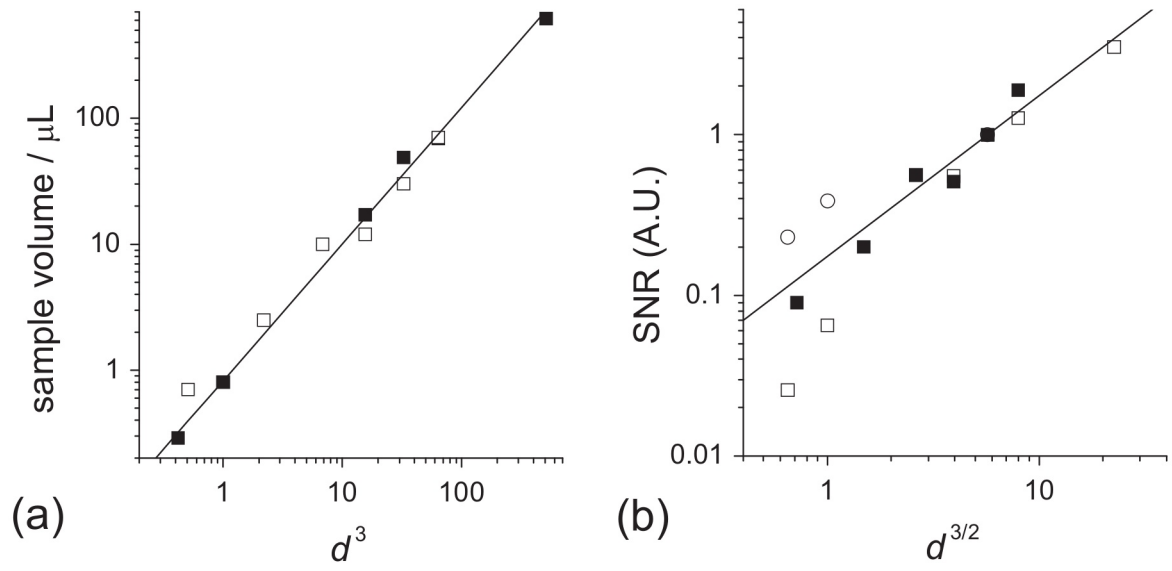


Figure II.2: a) Sample volume and b) *PSNR* function of rotor diameter, measured on ^{1}H (circles) or ^{13}C (squares). Empty shapes: Jeol; filled shapes: Bruker (*16*).

In order to further investigate commercial probes sensitivity, we checked the detection limit achievable for a two-dimensional (2D) {^{1}H}-^{31}P HETeronuclear CORrelation (HETCOR) spectrum of synthetic amorphous/crystalline HydroxyApatite (HAp, $Ca_{10}(PO4)_6(OH)_2$) powder (Section E.2 of Chapter I). Our samples mimic biological nano-apatites. The resulting spectrum presented three peaks: the crystalline phase at 0 ppm in ^{1}H dimension, adsorbed water around 5 ppm, and amorphous hydrogenophosphate HPO_4^{2-} around 13 ppm (Figure II.3a). We measured *PSNR* based on maximum of noise on this latter small peak, according to:

$$PSNR_{max} = \frac{H_{signal}}{h_{noise\ \ peak\ \ peak}/2} \qquad \text{II.1}$$

In this paragraph, we tried to determine the limit of detection available on standard equipment. With a low magnetic field B_0 = 7.0 T and a large 4.0 mm rotor containing a mass m ~ 100 mg, $PSNR_{max}$ = 28 was measured for a spectrum acquired in only 1 h. A similar 2D spectrum was obtained for a limited sample amount m = 2.5 mg at B_0 = 16.4 T with a 1.3 mm rotor, which is the smallest one available in our laboratory (Figure II.3c). This combination allowed an increased sensitivity by $B_0^{1.5}$ ~ 3.6 and an optimal coil filling factor. $PSNR_{max}$ = 3.2 was achieved in 16 h with the above set-up. According to Currie's definition this is just above quantification limit (Section B.3 of Chapter I) (*17*). Diluting sample by a factor 3.3 into KBr was attempted to decrease sample amount. The spectrum obtained with m ~ 750 µg at B_0 = 16.4 T in a 1.3 mm rotor is presented on Figure II.3d. $PSNR_{max}$ = 1.3 was just at detection

level in 64 h. When calculating sample weight needed to fill a 0.7 mm rotor, it would require a similar $m \sim 670$ µg. For materials like amorphous/crystalline HAp with broad and poorly intense species, a filled 0.7 mm rotor at high magnetic field on a modern spectrometer is probably the limit of detection in a reasonable time, still far from our objective to detect $m \sim 100$ µg. This highlights that commercial probes are – not yet – adequate to study mass-limited samples. However, Samoson developped a promising 0.5 mm MAS probe spinning at 170 kHz containing $m \sim 120$ µg of a protein for ^{1}H acquisition at ultra high magnetic field (*14*).

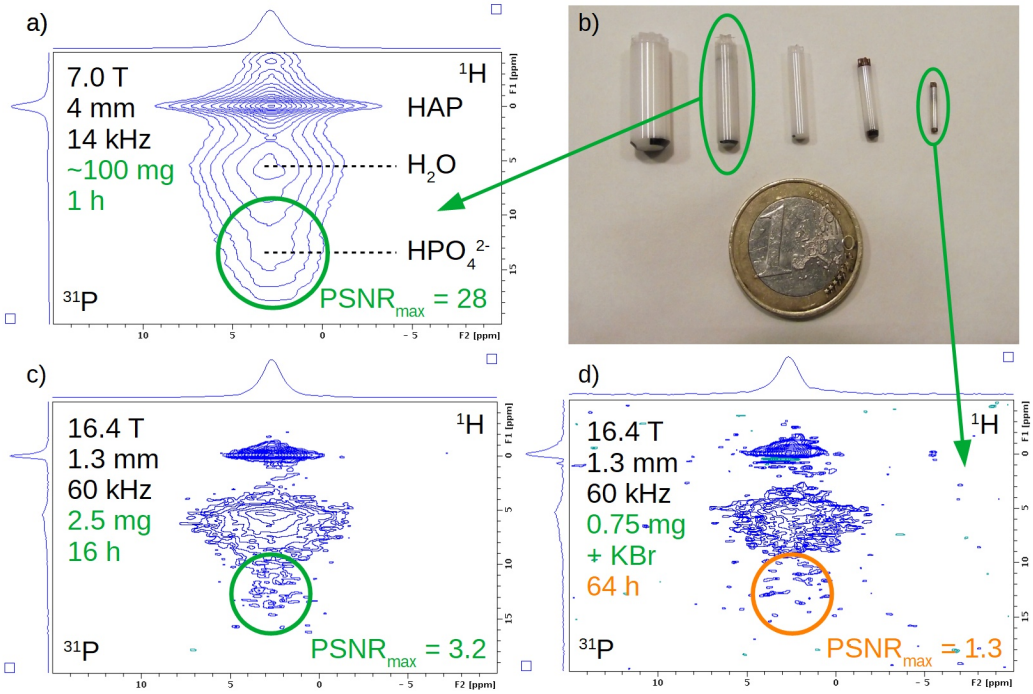


Figure II.3: Comparison of spectra of amorphous-crystalline HAp obtained for different sample volumes: a) $m \sim 100$ mg in a 4.0 mm rotor, c) $m = 2.5$ mg in a 1.3 mm rotor, d) $m \sim 0.75$ mg diluted in KBr in a 1.3 mm rotor; b) picture of rotors, 7.0, 4.0, 3.2, 2.5, and 1.3 mm, from left to right, respectively. $PSNR_{max}$ is indicated at bottom right of each spectrum, measured on HPO_4^{2-} peak. Parameters: a) $B_0 = 7.0$ T, $t_{exp} = 1$ h; c) $B_0 = 16.4$ T, $t_{exp} = 16$ h; d) $B_0 = 16.4$ T, $t_{exp} = 64$ h.

B.2. Microcoils

A solution to improve NMR sensitivity of samples available in small amounts is to maximise the coil filling factor η, *i.e.*, the ratio of the sample volume V_s to the coil volume V_c. For a solenoid coil, the filling factor is expressed according to Equation II.2,

$$\eta = \frac{V_s}{2V_c} \qquad \text{II.2}$$

where the factor 2 accounts for only half of the energy being stored inside the coil (*18*). This formula is not valid for other coils designs, such as saddle coil. For a maximum *PSNR*, η has to be as close as possible from 1. Hence, for sub-microlitre samples, coils with diameter lower than 1 mm, so-called *microcoils*, have to be preferred.

This can be illustrated by the reciprocity principle, developed by Hoult and Richards (Figure II.4a) (*18*). If point *A* is inside a coil, it experiences a strong magnetic field in presence of a circulating current in the loop. However, point *B* is outside the coil and feels a weaker magnetic field than in A. Reciprocally, a magnetic dipole rotating at point *A* induces a strong electric signal into the coil, whereas the same dipole at point *B* has a much xeaker effect. In other words, the longest is the distance between the coil and the sample, the less efficient are both excitation and signal detection. Accordingly, *PSNR* is proportional to the effective electromagnetic field B_1, for a given power *P* at Larmor frequency ω_0 (Equation II.3) (*19*).

$$PSNR \propto \frac{B_1}{\sqrt{P}} \qquad \text{II.3}$$

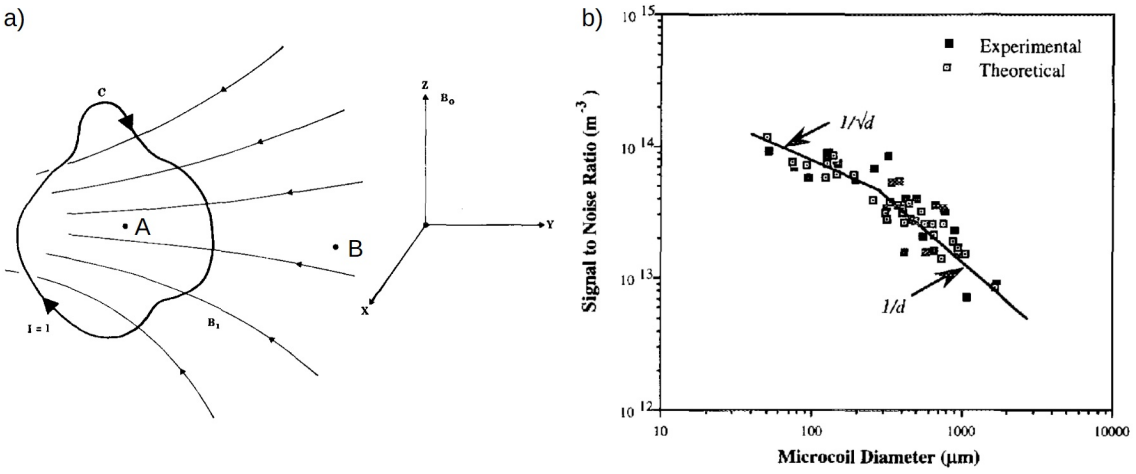


Figure II.4: a) Reciprocity principle, adapted from (*18*);
b) coil sensitivity against diameter (*20*).

In addition to filling factor, Peck highlighted that coil diameter has a strong impact on relative sensitivity of solenoid microcoils (Figure II.4b) (*20*). *PSNR* per volume is increasing for smaller coils. Two regimes are present, depending on coil wire diameter *d* in comparison to the Larmor frequency ω_0 (Equation II.4).

$$PSNR \propto \frac{\omega_0^{7/4}}{d}, \text{ skin effect regime}$$
$$PSNR \propto \frac{\omega_O^2}{\sqrt{nd}}, \text{ uniform current regime} \quad \text{II.4}$$

This is called the skin-effect and will be detailed in Subsection C.2. Skin effect regime is independent from the number of turns *n* of the coil.

B.3. Applications

The first known application of miniaturisation in continuous wave NMR was provided by Odeblad in 1966, by the observation of cervical secretions in microcoils ranging from 200 to 1000 µm (*21*). ^{13}C high resolution spectra of cholesterol in 1.7 mm coils were obtained in 1979, leading to a reduction of acquisition time by a factor 40 (*22*). A non-resonant 150 µm coil was used to enhance sensitivity of electron spin resonance in 1986 (*23*). Superfused mouse muscles were analysed by ^{31}P NMR in 2 mm capillaries surrounded by a small coil (*24*).

Moreover, NMR has been hyphenated with various chromatography techniques, either stopped-flow or continuous-flow (*25*). For instance, microcoils were connected to capillary electrophoresis by Wu *et al.* in 1994 (*26*). The limit of detection was investigated through ^{1}H-^{13}C Heteronuclear Multiple Quantum Correlation (HMQC), into a 850 µm ^{1}H solenoid coil with a neighbouring ^{13}C Alderman-Grant coil (*27*). An increase of sensitivity was provided by four parallel microcoils with RF switches (*28*). Microfluidic spectra were obtained by inductive coupling of a solenoid (*29*) and of a planar coil (*30*). An interesting study coupled flat coils and microfluidic into a miniaturized diagnostic magnetic resonance apparatus (*31*). A microfluidic on-flow reaction was monitored with photo-Chemically Induced Dynamic Nuclear Polarization (photo-CIDNP) (*32*). It should be noted that 1 µL microcoils are now commercially available for flow-through operation (*33*).

Microcoils revealed also useful for Magnetic Resonance Imaging (MRI), with 4 µm resolution on sub-millimetre objects (*34*), or with a 1 mm surface coil inductively coupled (*35*). In MRI, microcoils are often combined in arrays, up to 64 coils (*36*). ^{129}Xe flow imaging was performed by Hilty *et al.* (*37*). In solid state NMR, microcoils were used to increase Radio-Frequency (RF) field. Hence, as microcoils are concentrating electromagnetic field into a small volume, huge RF strength is achievable, such as 12.4 MHz at 256 W or even

24.6 MHz at 1 kW for a 127 µm coil tuned to ^{19}F at 377 MHz (*38*). For comparison, the same authors measured 195 kHz at 256 W on a standard 4 mm dual probe. This was used for instance to study ^{23}Na ions dynamics with stimulated echoes (*39*) or wideline ^{71}Ga spectra with an RF field of 600 kHz in a small coil of 1.6 mm (*40*). Different microcoils designs are presented on Figure II.5 (*41*). Gruber provided a practical guide for non-physicists on RF coils, especially for MRI (*42*). Furthermore, Lenz lenses were compared to LC resonators (*43*).

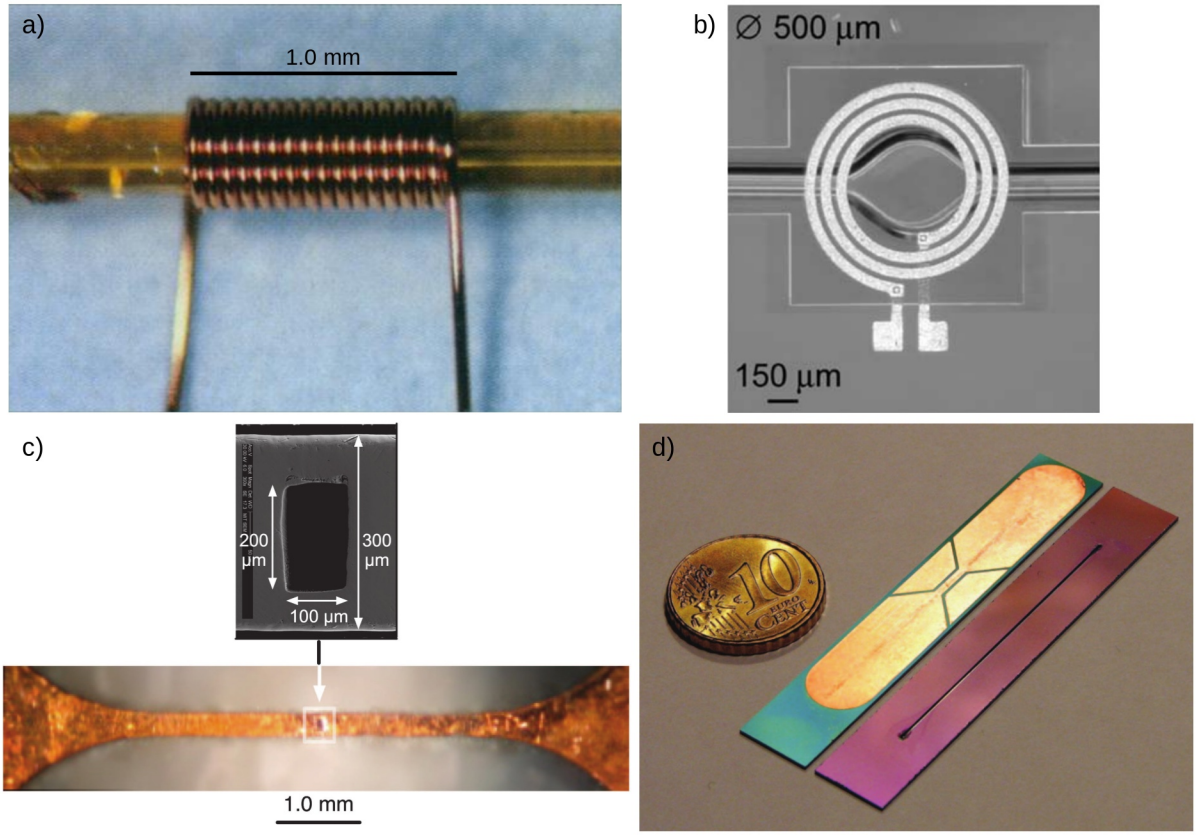


Figure II.5: Various NMR microcoils designs, adapted from (*41*): a) solenoid (*8*), b) planar spiral (*44*), c) microslot (*45*), and d) stripline (*9*).

When decreasing coil size, sample volume is more sensitive to magnetic inhomogeneities. As highlighted by Fucks *et al.* (*46*), any susceptibility effect introduced by coil materials and surroundings, has a strong effect on peaks shape and frequency. A list of measured susceptibilities on metals, dielectrics, and solvents, was provided in reference (*47*). A workaround is to match susceptibility of the coil surroundings, either using a perfluorinated fluid cell (*8*), or by bracketing the sample by a perfluoro-polymer, in conjunction with slow sample spinning (Figure II.6) (*48*).

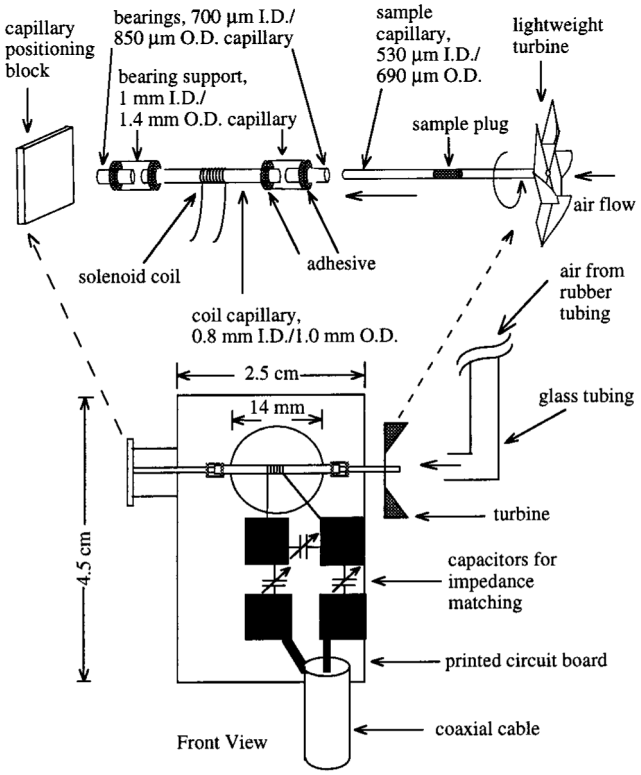


Figure II.6: Capillary spinning at 100-150 Hz to decrease susceptibility effect (*48*). The sample capillary is inserted into the microcoil capillary and connected to a lightweight turbine. The sample plug and the microcoil are both entrapped into a susceptibility matching fluid (Fluorinert FC-43, a mixture of C12 perfluoro-butylamines).

B.4. Magic angle coil spinning

In 2007, Sakellariou *et al.* combined microcoils with MAS, to give Magic Angle Coil Spinning (MACS) (*7*), which was patented (*49*). The idea was to introduce the sample into a solenoid microcoil, placed into a standard rotor, and to close the electronic circuit by a tuning chip capacitor. In this way, the microcoil was self-resonant. This is presented on Figure II.7a-b. The two coils are behaving like a power transformer, with a main circuit (coil 1 into the probe, Figure II.7c), and a secondary circuit (microcoil into the rotor). While the former is static, the latter is in rotation at the magic angle. By inductive coupling between the two coils, energy transfer occurs, enabling both excitation and signal detection of microquantities.

In the following years, further investigations were done by Sakellariou's group on MACS concept (*19*, *50*), on the effect of eddy currents (leading to extra heating of the sample, a drawback of MACS) (*51*, *52*), for application to biological samples (*53–56*), and to materials (*57*, *58*). Two other designs were tested in collaboration with Korvink *et al.* for

microfabricated coils (*59*) and with Ginefri *et al.* for Transmission Line Resonators (TLR) (*60*). After a common work on a static sample (*61*), Yamauchi, Kentgens *et al.* separately implemented a microMAS equipment (*62*, *63*) consisting of a microcoil entrapped into a flat capacitor, both positioned outside of the rotor, so-called *piggy-back* microcoil (Figure II.8a). Sample holder was fixed on a standard rotor. Kentgens *et al.* also studied heteronuclear decoupling (*9*, *64*), 1H homonuclear decoupling (*65*, *66*), and micrometre single crystals (*67*) in spinning microcoils.

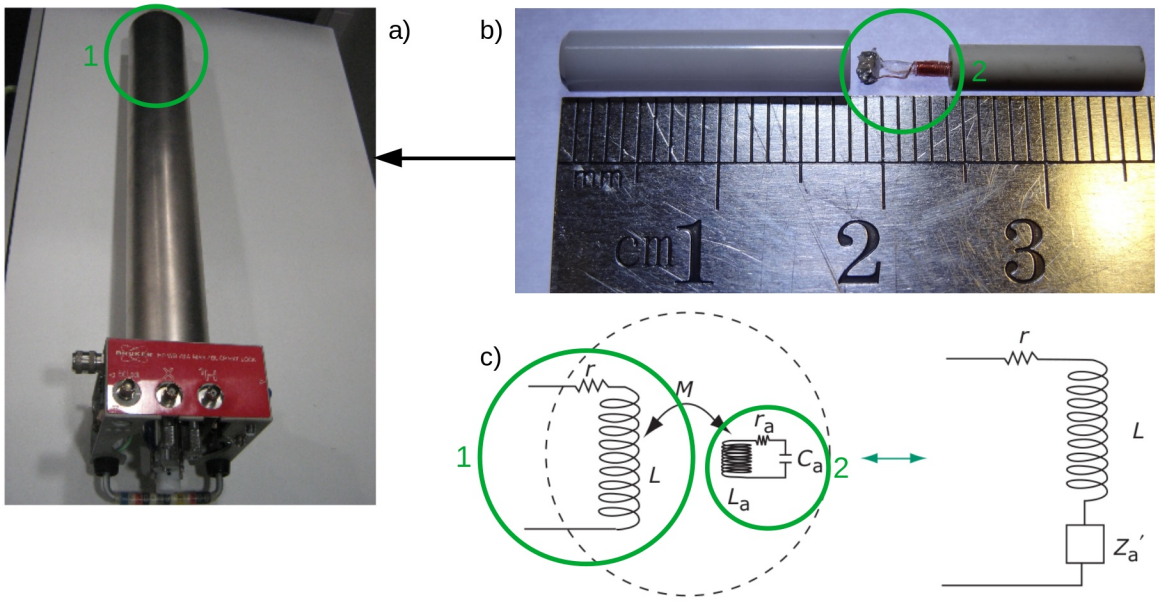


Figure II.7: MACS concept: a) standard MAS NMR probe (coil 1) containing b) a standard 4.0 mm rotor where a tuned microcoil (coil 2) is placed into a ceramic holder; c) by inductive coupling, the energy is transferred from coil 1 to coil 2, adapted from (*19*).

Takeda proposed three other microcoils designs: another *piggy-back* module attachable to a standard MAS probe (Figure II.8b) (*68*), a doubly tuned capillary coil inside a standard rotor (*69*), and a disk MAS for thin layers (*70*). These implementations were compared to each other (*71*), as well as the impact of the involved high RF field on pulse imperfections. Korvink's group designed Micro-Electro-Mechanical Systems (MEMS) (Figure II.8c) (*72*), which were placed into a MAS rotor (*59*). This was improved by new capacitor design (*73*) and recently applied to metabolomics (*74*). In this latter domain, Jeol manufacturer developed a high resolution 1 mm microMAS probe (*75*). Ginefri's group developed TLR for MRI (Figure II.8d) (*76*, *77*). A last design was provided by Leidich *et al.* with Helmholtz coils in a home-made probe, where a 330 µm capillary was spinning at magic angle (*78*).

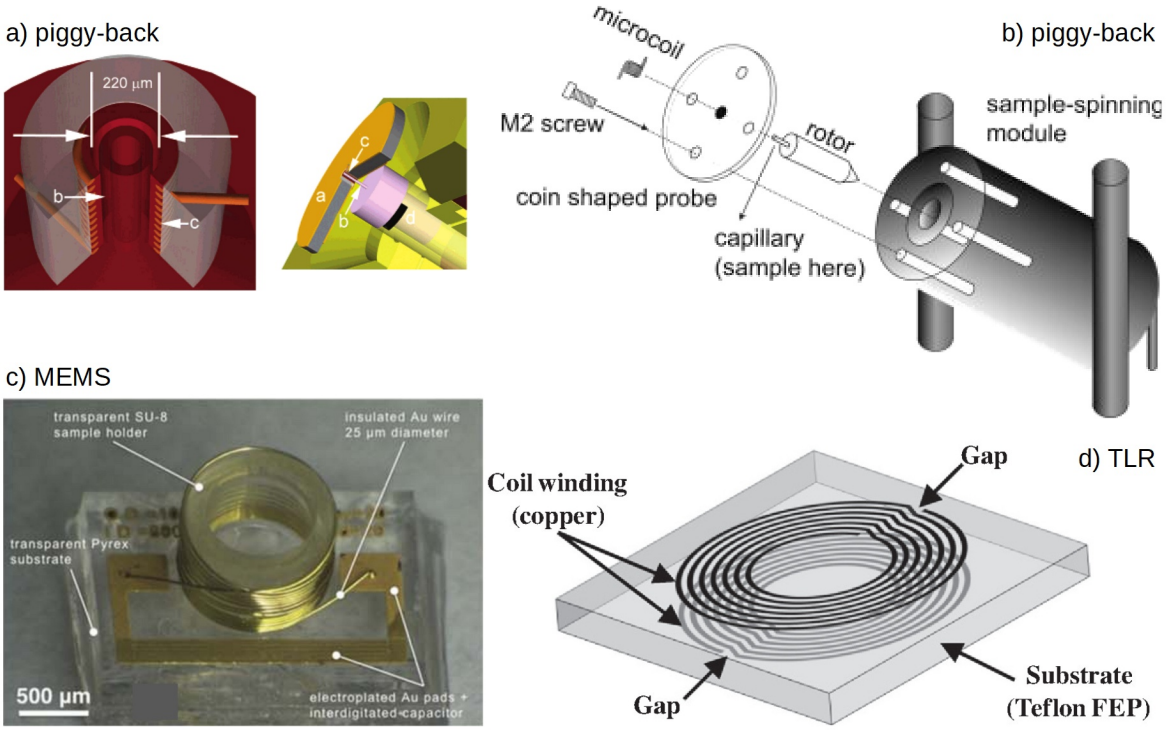


Figure II.8: MACS alternatives by a) Kentgens' group (2006) (*63*), b) Takeda's group (2008) (*68*), c) Korvink's group (2010) (*59*), d) Ginefri's group (2012) (*76*), initially implemented for imaging, contrary to other designs.

Piggy-back microcoils main disadvantage was the need to modify the probe, especially not only once, but each time one wanted to study a different nucleus (Figure II.8a-b). However, a high RF field was achievable. MEMS had the advantage to be highly reproducible (Figure II.8c). Nevertheless, they had a low aspect ratio (length to width), which decreased RF homogeneity and sensitivity. Moreover, materials used and clear lack of axial symmetry induced susceptibility broadening and limited MAS rate. Much progress has since been done and symmetric microcoils for solid-state NMR suitable for 100 µg samples are now commercially available (*33*). Hereafter, we will be interested in MACS and TLR designs (Figures II.7 and II.8d, respectively), which neither require a probe modification.

In Subchapter B, we highlighted that commercial probes are not yet suitable to analyse diluted species. For mass-limited samples, microcoils benefit from an optimised filling factor and an increased *PSNR* per unit of volume. Their applications were presented before detailing MACS concept and alternatives.

C. Some elements of electromagnetism

Electromagnetism is widely used to prepare radio-frequency coils, especially for radio transmitters (*79*). In order to design microcoils, a few basics are necessary, starting with resonance frequency, inductance and capacitance (Section C.1), skin and proximity effects (Section C.2), and resistance and quality factor (Section C.3). Eddy currents, radio-frequency homogeneity, and inductive coupling will be investigated in Sections C.4, C.5, and C.6, respectively.

C.1. Resonance frequency, inductance and capacitance

C.1.a. Solenoid coils

The solenoid microcoil form a serial RLC electronic circuit, whose frequency ν depends on inductance L and capacitance C:

$$\nu = \frac{1}{2\pi\sqrt{LC}} \qquad \text{II.5}$$

(ν: Frequency (Hz); L: Inductance (H); C: Capacitance (F))

For a solenoid coil, Lorenz gave a formula (*80*) based on infinitesimally thin conducting sheet (in orange on Figure II.9) (*79*). Nagaoka rewrote Lorenz's formula with a shape factor k_L, function of diameter d_{coil} and length l_{coil} of coil, and available as a table (*81*):

$$L = \frac{4\pi^2 N_{turns}^2 r_{coil}^2}{l_{coil}} k_L - 0.004\pi r_{coil} N_{turns} (k_S + k_M) \qquad \text{II.6}$$

(N_{turns}: Number of turns; r_{coil}: Coil radius (m); l_{coil}: Coil length (m); k_L: Nagaoka coefficient = f (d_{coil} / l_{coil}); k_S: Self inductance = f (p_{pitch} / r_{wire}); k_M: Mutual inductance = f (N_{turns}); Lorentz's sheet formula; Rosa's round wire corrections)

For a round wire solenoid coil (in blue on Figure II.9), Rosa provided two additional correcting factors (right part of Equation II.6) (*82*), function of pitch p_{pitch}, wire radius r_{wire} and of the number of turns N_{turns}. Once the inductance was calculated, the adequate capacitor value was chosen according to the commercially available ones. By brazing the capacitor to the solenoid, the circuit was closed and auto-resonating at the desired frequency. This is the key

point of MACS: the RLC circuit can be adjusted to almost any frequency *without* modifying the probe, simply by changing the capacitor value or the solenoid parameters.

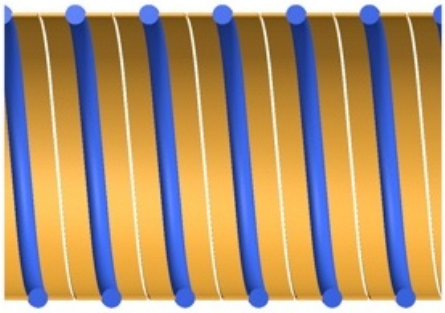


Figure II.9: Sheet layer (orange) and round wire (blue) solenoid coils, adapted from (*79*).

C.1.b. Transmission line resonators

Monolithic Transmission Line Resonators (TLR) are much more recent than solenoids (*83*). They consist of a low-loss dielectric substrate surrounded by a conductive layer on each side (Figure II.10) (*77*). Each layer is interrupted by one or many gaps, diagonally opposed from one side to the other one. Multi-turns and multi-gaps can be combined into a single TLR. Two current modes are simultaneously present on both TLR sides. The first mode is anti-symmetrical with equal magnitudes but opposite signs, responsible for the self-resonance frequency. The second mode is symmetrical with same magnitude and same sign and generates an out of plane RF magnetic field, whereas the electric field is concentrated around the gaps (*84*). TLR are prepared by photo-lithographic etching (*85*).

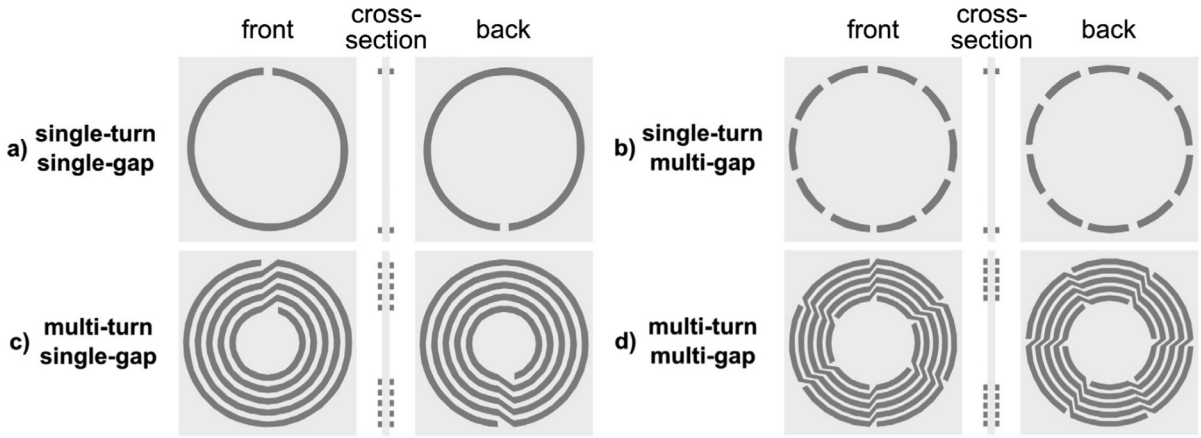


Figure II.10: TLR designs: a) single-turn single gap, b) single-turn multi-gaps, c) multi-turns single-gap, d) multi-turns multi-gaps, adapted from (*77*).

A complete mathematical description of TLR resonance frequency is available in (*86*). Briefly, an analytical formula was provided by Gonord *et al.* for single-turn coils (*84*) and extended to multi-turns designs by Serfaty *et al.* (*87*):

$$\frac{L_{tot}\,\omega}{4\,N_g\,Z}\tan\left(\frac{\omega\sqrt{\varepsilon}\,l_f}{4\,N_g\,c}\right)=1 \qquad \text{II.7}$$

where ω is the angular resonance frequency, N_g is the number of gaps per conductor, ε is the substrate dielectric constant, l_f is the length of one conducting band, and c is the speed vacuum of light. L_{tot} corresponds to the total TLR inductance, defined by the individual inductance L_i of each turn, the number of turns N, and their mutual inductances M_{ij}:

$$L_{tot}=\sum_{i=1}^{N}\left(L_i+2\sum_{j=1+1}^{N}M_{ij}\right) \qquad \text{II.8}$$

Z reflects the transmission line impedance, depending on the conductor width w, and the substrate thickness h. Depending on their ratio, the wide band approximation ($w > h$) or the narrow band approximation ($w < h$) is calculated.

C.2. Skin and proximity effects

When a Direct Current (DC) is circulating into a straight conductive wire, this current is uniformly distributed into the wire. On the contrary, with Alternating Current (AC), the current is limited to the surface, giving the so-called *skin effect*, due to eddy currents that tend to oppose to the current oscillations. Skin depth δ is presented on Figure II.11a (*88*) and defined according to Equation II.9 (*89*):

$$\delta=\sqrt{\frac{\rho}{\pi\,\nu\mu}} \qquad \text{II.9}$$

ρ: Resistivity (Ω.m); μ: Magnetic Permeability (H.m^{-1}); ν: RF Frequency (Hz)

involving RF frequency ν, resistivity ρ, and magnetic permeability μ, whose values are presented in Table II.2 for various materials (*90*). A unit-less scaling parameter can be defined as $z = d / \delta$, where d is the wire width. With $z \geq 8$, the system is in the skin effect regime. Minard and Wind rather defined the high frequency limit at $z \geq 5$ (*91*), the value we used. On the contrary, $z \leq 2$ corresponds to the uniform current regime. An intermediate regime is present for $2 < z < 5$ or 8.

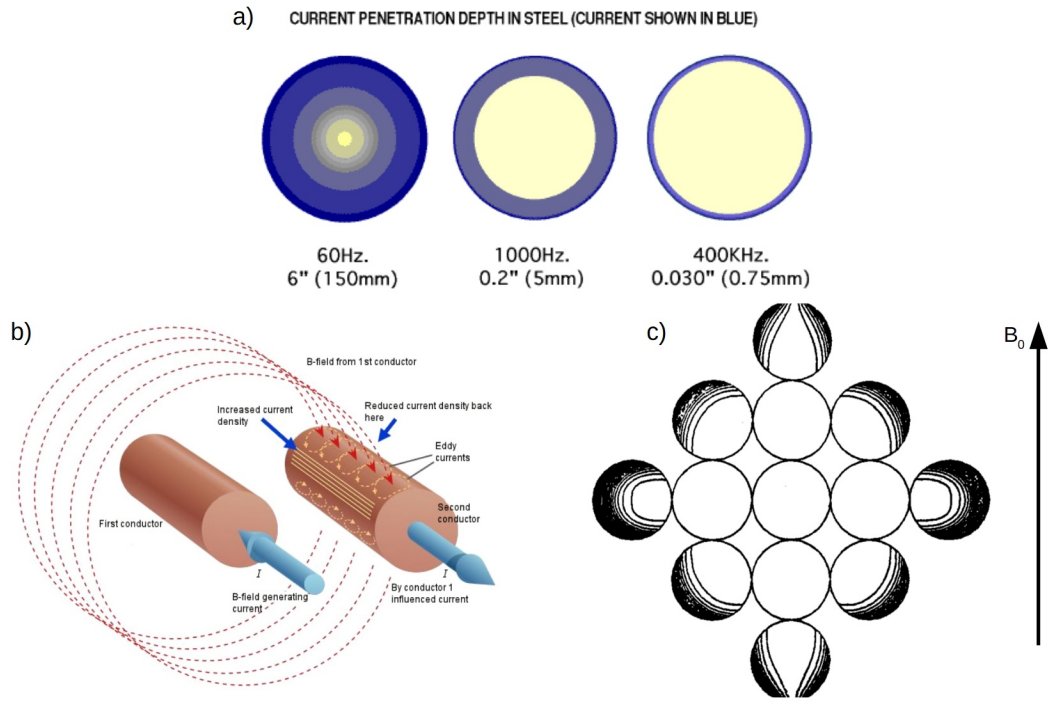


Figure II.11: Skin and proximity effect: a) skin depth as a function of AC frequency on steel cables (*88*); b) proximity effect applied from left wire to right one, on which current density increased on left part, whereas it decreased on right part (*92*); c) combination of skin and proximity effect, with a vertical magnetic field, adapted from (*93*).

Table II.2: Resistivity, magnetic permeability and skin depth of conductive materials.

Material	**Resistivity ρ (10^{-8} Ω.m)**	**Magnetic permeability μ (10^{-6} H.m^{-1})**	**Skin depth**			
			60 Hz	**1 kHz**	**1 MHz**	**1 GHz**
Lead	20.6	1.26	30 mm	7.2 mm	230 µm	7.2 µm
Tin	12.6	1.26	23 mm	5.7 mm	180 µm	5.7 µm
Aluminium	2.65	1.26	11 mm	2.6 mm	82 µm	2.6 µm
Gold	2.20	1.26	9.6 mm	2.4 mm	75 µm	2.4 µm
Copper	1.68	1.26	8.4 mm	2.1 mm	65 µm	2.1 µm
Silver	1.63	1.26	8.3 mm	2.0 mm	64 µm	2.0 µm
Iron	10.1	1260	0.65 mm	0.16 mm	5.1 µm	0.16 µm

Another important effect of current circulation is present when two wires are in close neighbouring, *i.e.*, the *proximity effect*, presented on Figure II.11b (*92*). The current circulating into the left wire induces a magnetic field, that crosses the right wire. This generates a second current, whose density is increased at proximity of the first conductor. The closer are both wires, the higher is the effect. Skin effect and proximity effect are combined on Figure II.11c (*93*). Each wire is symbolised by a white circle. As no insulation is present, all wires are connected to each other and skin effect concentrates current (in black) to the

global envelope. Proximity effect distorts the overall shape to fit the individual wires curvature.

C.3. Resistance and quality factor

Skin and proximity effects induce resistance into the conductive wire. For a straight wire, the global AC resistance R_{AC}^{wire} is given by:

$$R_{AC}^{wire} = \frac{\rho l}{\pi \delta (d - \delta)} \qquad \text{II.10}$$

where ρ is the resistivity, δ is the skin depth, and l and d are wire length and width, respectively. For a coil, proximity effect has to be taken into account with the factor ξ (*91*), experimentally tabulated by Medhurst (*94*), who revised Butterworth's work (*95*):

$$R_{AC}^{coil} = R_{AC}^{wire}\left[1 + (\xi - 1)\left(1 - \frac{1}{n_{turns}}\right)\right] \qquad \text{II.11}$$

ξ values depend on the length-to-diameter ratio of the coil and of the wire-diameter-to-inter-turn-spacing ratio. This factor was designed at high frequency limit $(z = d/\delta \geqslant 5)$ and for long coils $(n_{turns} \geq 30)$ and has to be corrected by $(1 - 1/n_{turns})$ for smaller coils (*96*). An improved formula taking into account low-frequency limit and correcting underestimation of R_{AC}^{coil} for $\xi = 1$ was provided by Knight (*96*). Resistance variation as a function of z parameter is depicted on Figure II.12a (*20*). The slope rupture for $2 < z < 5$ is clearly visible. In addition, circuit resistance impacts noise intensity during acquisition. Hence, *PSNR* slope is changing according to z (Figure II.4b).

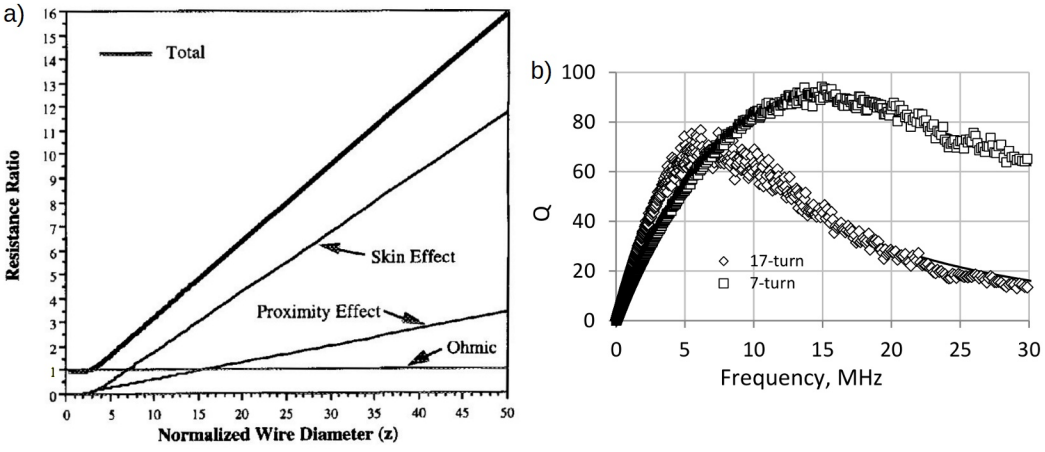


Figure II.12: a) Resistance as a function of normalised wire diameter (*20*); b) quality factor of two coils as a function of frequency (*97*).

Resistance R and frequency v are also involved into the quality factor Q (*42*):

$$Q = \frac{Maximum\ energy\ stored}{Average\ energy\ dissipated\ \ per\ cycle} = \frac{2\pi v L}{R} = \frac{1}{R}\sqrt{\frac{L}{C}} \qquad \text{II.12}$$

Variation of Q against frequency is presented on Figure II.12b (*97*). Similar results will be obtained at higher frequencies, simply by adjusting coil inductance. Typical Q values are between 10 and 100. As maximum of Q depends both on R and L, coil pitch is a crucial parameter, as it impacts both of them by the mutual inductance and the proximity effect. For a given wire diameter, the optimal sensitivity is obtained for a pitch of 1.5 times the wire diameter (*91*).

C.4. Eddy currents

Eddy currents are known since 1883 and were derived from Maxwell's electromagnetic theory (*98*). They can be simplified as follows: any movement from a conductive metal placed into a magnetic field induces a current that generates a magnetic field in the opposite direction. The same effect appears with a static metal but a moving magnetic field. This generates heating by Joule effect and undesired gradient magnetic fields. Schäfer and Heiden measured the current losses in full and hollow copper cylinders slowly spinning (Figure II.13a, black line and points) (*99*). They evidenced that these currents are proportional both to the height and to the thickness of copper cylinders. In other words, the higher volume of metal, the stronger are the eddy currents. Aubert *et al.* confirmed their model (Figure II.13a, superimposed red line) (*52*) and that two types of current trajectories are present, depending on cylinder shape ratio (Figure II.13b). They applied this analysis to MACS microcoils and highlighted that power dissipation P can be estimated according to Equation II.13:

$$\begin{gathered} P_{min} < P < 2P_{min} \\ P_{min} = n_{turns}\frac{2\pi\left(B_0\omega_r \sin\theta\right)^2}{3\rho}a^5\beta^3\varepsilon \\ \text{with } a = \frac{d_{capi} + w_{wire}}{2},\ \beta = \frac{h_{wire}/2}{a},\ \text{and } \varepsilon = \frac{w_{wire}}{a} \end{gathered} \qquad \text{II.13}$$

where n_{turns} is the number of wire turns, B_0 is the magnetic field, ω_r is the spinning angular frequency, θ is the spinning angle, ρ is the resistivity of the wire material, d_{capi} is the capillary diameter and h_{wire} and w_{wire} are the rectangular section dimensions of the wire, respectively.

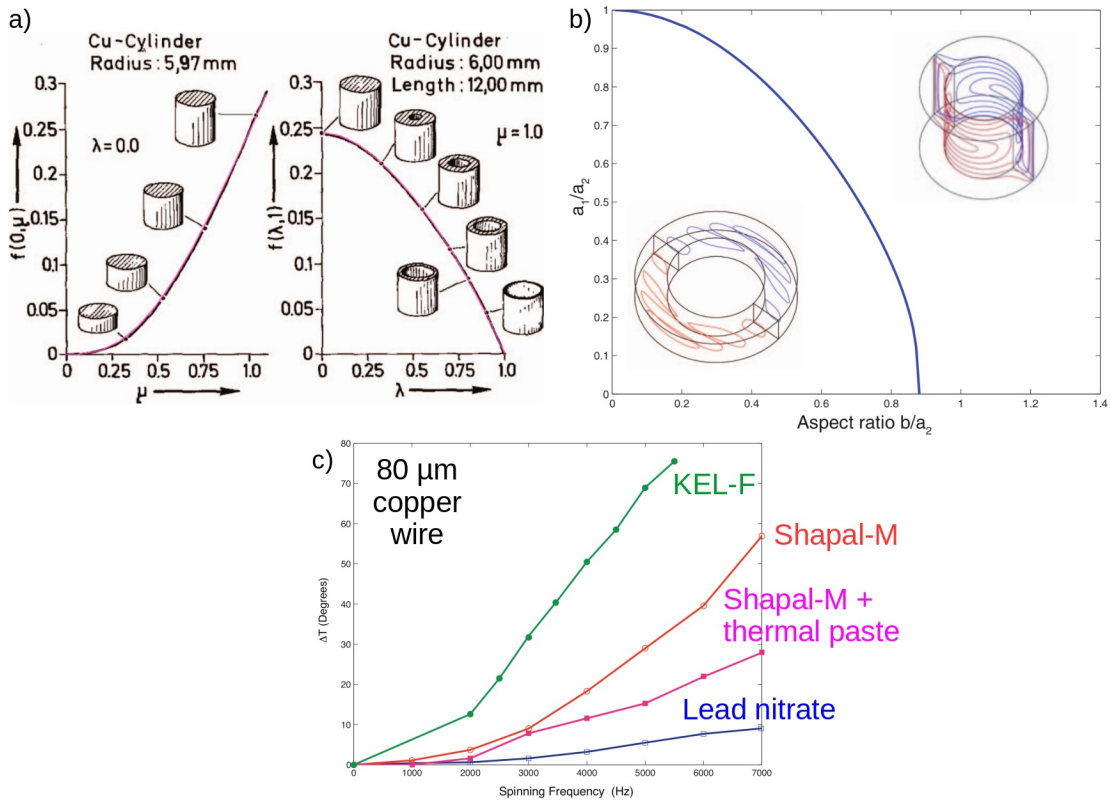


Figure II.13: a) Eddy currents for a full and hollow cylinders, with $\lambda = d_{capi}/(d_{capi}+w_{wire})$, $\mu = (h_{wire}/2)/(d_{capi}+w_{wire})$ and $f(\lambda,\mu) \propto P$ (*52*, *99*); b) current trajectories depending on shape factor, blue and red curves correspond to a mirror symmetry (*52*); c) temperature increase with MAS rate, with various insert materials, blue curve: without microcoil (*7*).

This heating impacts the sample entrapped into MACS microcoil, and has to be minimised. Temperature increase depends on the insert material holding the microcoil (Figure II.13c) (*7*). With a KEL-F insert (polychlorotrifluoroethylene), temperature increased by 75°C at *MAS* = 5500 Hz for a thick copper wire of 80 µm (green curve). KEL-F has a low thermal conductivity of ~ 0.21 $W.K^{-1}.m^{-1}$, whereas Shapal-M (machinable aluminium nitride) has a much higher one of ~ 80 $W.K^{-1}.m^{-1}$ (*51*). With Shapal-M, temperature increased only by 55°C at *MAS* = 7000 Hz (red curve), and was further reduced by adding thermal paste, reaching only 30°C at *MAS* = 7000 Hz (pink curve). Thermal paste disadvantage is the introduction of undesired NMR signals (metal particles and polymer as adjuvant), and of a change of magnetic susceptibility. Power dissipation of microcoil was calculated to be 100 mW in such conditions. Temperature was measured using lead nitrate ^{207}Pb NMR chemical shift as an internal thermometer (*100*). For comparison, 10°C heating was obtained on lead nitrate without microcoil (blue curve). Eddy currents heating were further minimised to 20°C with thin wires of 25 µm at *MAS* = 10,000 Hz (*51*). If fast spinning speeds are desired, the thinest usable wires have to be chosen.

C.5. Radio-frequency homogeneity

Radio-Frequency (RF) homogeneity is a necessary condition to manipulate the spin magnetisation present in a given volume. For instance, inhomogeneities prohibits the use of multi-pulses sequences, which are requested to provide the most accurate informations on the analysed sample. A way to characterize this RF homogeneity is to do a nutation curve, where intensity of signal is measured as a function of pulse duration. Figure II.14a highlights three different implementations: a solenoid (stars) with a regular nutation, a Helmholtz coil (squares) with an imperfect inversion and a left shift for second and third oscillations, and a planar coil (circles), with only a first maximum but no following inversion, demonstrating a very poor RF homogeneity (*101*). It should be noted that planar coils used here are not TLR, but rather a single spiral coil. Helmholtz coil combines two inductively coupled planar coils, with the sample in between.

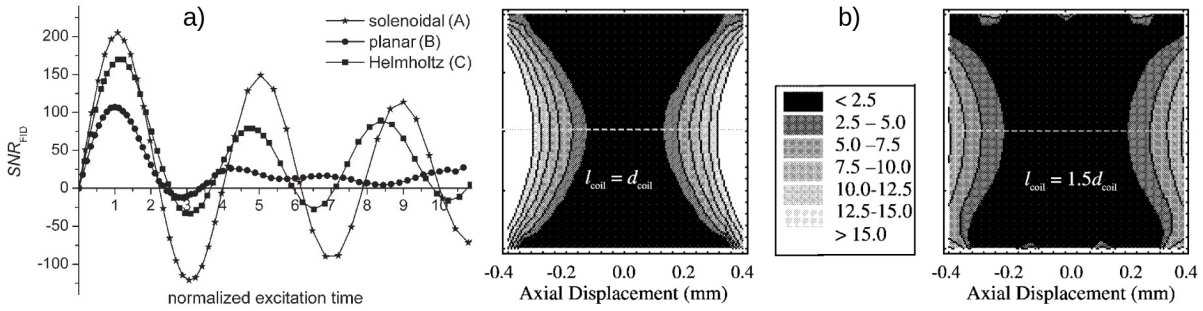


Figure II.14: RF homogeneity: a) comparison of different coil shapes (*101*), b) comparison of two solenoids with a different length, the grey scale and the values nearby indicate the percent of deviation of the RF field strength relatively to coil center (*102*).

Minard and Wind calculated RF field homogeneity for different coil shapes and highlighted that a solenoid microcoil with a length of 1.5 times its diameter gave much better results than with a relative length of 1 (Figure II.14b) (*102*). Even in this case, the sample has to occupy half of the coil volume, being centred into it. With this configuration, the minimum coil length is 1.2 times its diameter to have a RF homogeneity with a maximum deviation less than 10 %. Accordingly, the ratio $l_{sample} : d_{coil} : l_{coil}$ of 0.5 : 1 : 1.2 has to be preferred. Any oversized sample will have a poor RF filed homogeneity. An alternative is to use variable pitch coils, with a reduced pitch at each end of the solenoid, which increases the RF homogeneity at microcoil extremities (*103*).

C.6. Inductive coupling

When two coils with inductances L_1 and L_2 are close from each other, inductive coupling occurs, with a mutual inductance M (Figure II.7c). The coupling coefficient is defined as:

$$k = \frac{M}{\sqrt{L_1 L_2}} \qquad \text{II.14}$$

Three coupling regimes can be defined: under-coupling, critical-coupling and over-coupling (*86*). Under-coupling is present when the pick-up coil (primary circuit) is far away from the other coil (secondary circuit). Critical-coupling occurs when the same power is dissipated into the primary and secondary circuits. For two solenoid coils with the same shape ratio, critical coupling depends on their volume V_1 and V_2 and their quality factors Q_1 and Q_2 previously defined in Equation II.12:

$$Q_1 Q_2 V_2 / V_1 = 1 \qquad \text{II.15}$$

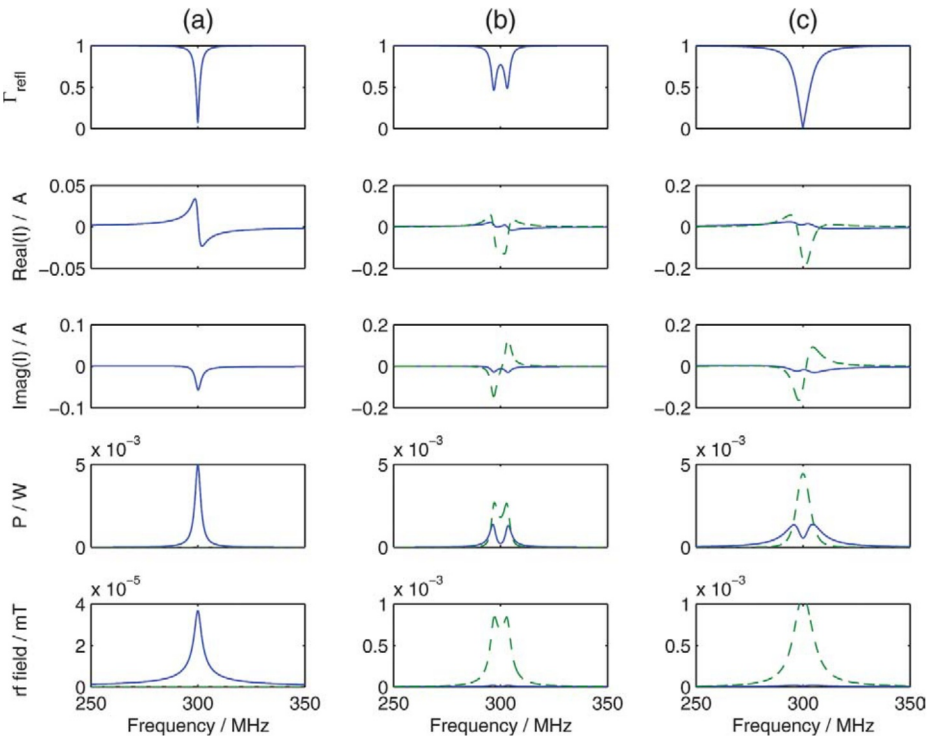


Figure II.15: Comparison of simulated frequency response (first row), real and imaginary parts of the current (second and third rows, respectively), power dissipation (fourth row), and RF magnetic field (last row) for a) probe coil, b) MACS coil inserted into probe coil, and c) after retuning and rematching to 50 Ω. Continuous and dashed curves correspond to probe coil and MACS microcoil responses, respectively (*50*).

Over-coupling corresponds to the regime where $Q_1Q_2V_2/V_1 >> 1$. Most of the power applied to the probe coil is transmitted to the microcoil. This is typical for MACS experiments. However, in the general case, the magnetic field B_1^{macs} into the microcoil and the one into the probe coil in absence of microcoil B_1^{ref} are related by:

$$\frac{B_1^{MACS}}{B_1^{ref}} = Q_2 \sqrt{\frac{1}{1+V_2 Q_1 Q_2 / V_1}} \qquad \text{II.16}$$

As a consequence of over-coupling, the frequency response is splitted into two resonances (first row of Figure II.15a-b) (*50*). This can be corrected by probe tuning and matching with the variable capacitors integrated into the probe, the ones that are daily used without MACS (Figure II.15c). When inspecting real and imaginary parts of the current (second and third rows), they are more important into the secondary coil (dashed curves), than into the primary coil (plain curves). Accordingly, the power is mostly dissipated into the secondary coil (fourth row), and RF magnetic field is much more intense into it (last row).

In Subchapter C, inductance formulae were provided to design solenoids and TLR microcoils. Two frequency regimes were highlighted, depending on skin-effect, and impacting circuit resistance. Eddy currents heating and RF homogeneity were investigated. Finally, inductive coupling was presented.

D. Microcoils manufacturing

Multiple steps are needed to manufacture microcoils. The first one is to calculate the adequate parameters (Section D.1). Miniaturisation will be highlighted in Section D.2. Coils winding and capacitor brazing, followed by frequency analysis will be detailed in Sections D.3, and D.4, respectively. Michromechanics will be investigated in Section D.5. Finally, a supplies list will be provided in Section D.6.

D.1. Parameters calculation

A considerable work has been undergone to implement all the above electromagnetism formulae into a single solenoid calculation LibreOffice sheet, available online (*104*). Page 1 is presented on Table II.3 and updated from values from all the other pages. By selecting desired nucleus and magnetic field strength into a list, the resonance frequency is found on page 2. The microcoil supplies are found on page 3, including wire material and diameter, capillary

inner and outer diameter, and available capacitor values. Inductance is defined on page 4, according to Nagaoka's factor (*81*) and Rosa's round-wire corrections (*82*), using macros provided by (*79*). Rosa's formula is valid only for pitch angles lower than 5°. If this condition is not respected, a warning is displayed on page 1 by conditional formatting in red of the corresponding cell. The inductance of straight wire leads whose length l_{leads} depends on coil position and capacitor size is also calculated (*105*):

$$L_{DC}^{wire} = 200\, l \ln\left(4\, l_{leads}/d_{wire} - 0.75\right), \text{ in nH}$$
$$L_{AC}^{wire} = 200\, l \ln\left(4\, l_{leads}/d_{wire} - 1\right), \text{ in nH, skin effect regime} \qquad \text{II.17}$$

Table II.3: LibreOffice sheet for solenoid calculation (*104*). Yellow, grey, green and red boxes are adjusted, calculated, valid and invalid values, respectively.

Parameter	**Units**					
Magnetic field	T	11,7				
Nucleus		31P				
Frequency	MHz	**202,46**				
Alllowed error	MHz	10				
Spinning Frequency	Hz	5 000				
Capillary out diam	µm	**870**				
Capillary in diam	µm	700				
Capacitor size		Micro				
Capacitor width	mm	0,5				
Wire material		Copper				
Resistivity	Ω.m	1,68E-8				
Permeability	H/m	1,26E-6				
Skin depth	µm	4,58				
Min wire diameter	µm	22,92		5	5*skin depth	
Wire diameter	µm	**50**				
Coil length	mm	**1,0**	optimum	1,1	1,2*Dcoil	Minard 2001
Coil position	mm	**5,0**				
Number of turns		**5,5**	optimum	14	maximum	16
Pitch	µm	182	optimum	75	1,5*Dwire	Minard 2001
Sample length	mm	**0,5**	optimum	0,46	0,5*Dcoil	Minard 2001
Sample Volume	nL	192,4				
Pitch Angle ψ	degree	3,81	smaller than	5		
Total inductance	µH	0,03423				
Total AC resistance	Ω	0,74	http://g3ynh.info/zdocs/magnetics/SolenoidZ.pdf			
Power dissipation	mW	**31,8**	maximum	200	Aubert 2012	
Capacity	pF	**15,0**	**18,0**	**22,0**		
Calc frequency	MHz	**222,1**	**202,8**	**183,4**		
Quality factor		**64,4**	**58,8**	**53,2**		

This formula is valid only if the wire is surrounded by air. A detailed analysis taking into account skin depth can be found at (*106*). Three capacitor values are suggested for the total inductance. If the calculated frequency is away from the desired one, its cell colour changes to red. Resistance and quality factor are extracted from Medhurst's table (*94*) on page 5 of the LibreOffice sheet. Eddy currents heating is obtained from page 6 according to Aubert *et al.* (*52*). Again a warning is displayed if heating exceed the fixed threshold.

The obtained resonance frequency was in accordance within an error of 10 MHz with the expected one. The imprecision arises typically from coil imperfections, in particular from an inexact number of turns, from coil length impacting pitch, and from leads length. A few rules of parameters choice are presented in Table II.4. Hence, the number of turns has a strong impact. On the contrary, coil position is typically a parameter to finely adjust resonance frequency, the ideal being to place the microcoil at the vertical center of the probe coil for the best inductive coupling.

Table II.4: Influence of parameters for microcoil design.

Parameter	Resonance frequency	Quality factor	Heating
↑ Spinning frequency	→	→	↑
↑ Capillary diameter	↓	↑	↑
↑ Wire diameter	↑	↑	↑
↑ Coil length	↑	↑	→
↑ Coil position	↓	↓	↑
↑ Number of turns	↓	↑	↑

TLR are more complicated to design. For the moment, one can use a 3-step Matlab (The MathWorks, Inc., Natick, MA, USA) script, provided by J. Lehmann-Horn (Bruker) with FastHenry program (*107*). An alternative is to use an electromagnetic simulation software, FEKO (Altair Hyperworks, Troy, MI, USA).

D.2. Miniaturisation

Wire diameter as a strong impact on eddy currents heating. Its width has to be five times higher than the skin depth (Section C.2). For instance, to study ^{31}P nucleus at 11.7 T, the wire should have a diameter thicker than 23 µm (Table II.3). The wires that we used were ranging from 30 to 80 µm. For comparison, hairs diameter is between 50 and 100 µm. Wire diameter is thus a compromise between eddy currents and ease of use. 50 and 80 µm wires are robust to manipulate, whereas 30 µm wire are easily broken.

My first successful solenoid is presented on Figure II.16a, with 80 µm wire, ‘huge’ capacitor, broken capillary, and irregular glue. After many trials, I succeeded to have a series of microcoils with 50 µm wire and small capacitor (Figure II.16b). A 50 µm wire microcoil is detailed on Figure II.16d, where one can see the capillary, the solenoid, and its leads connected to the capacitor. The different capacitors used are shown on Figure II.16c, ranging

from 1.5 mm (left) to 0.5 mm (right). The smallest ones gave a more stable rotor spinning. Their commercially available values covered the range from 1 to 330 pF. TLR with different diameters are depicted on Figure II.16e.

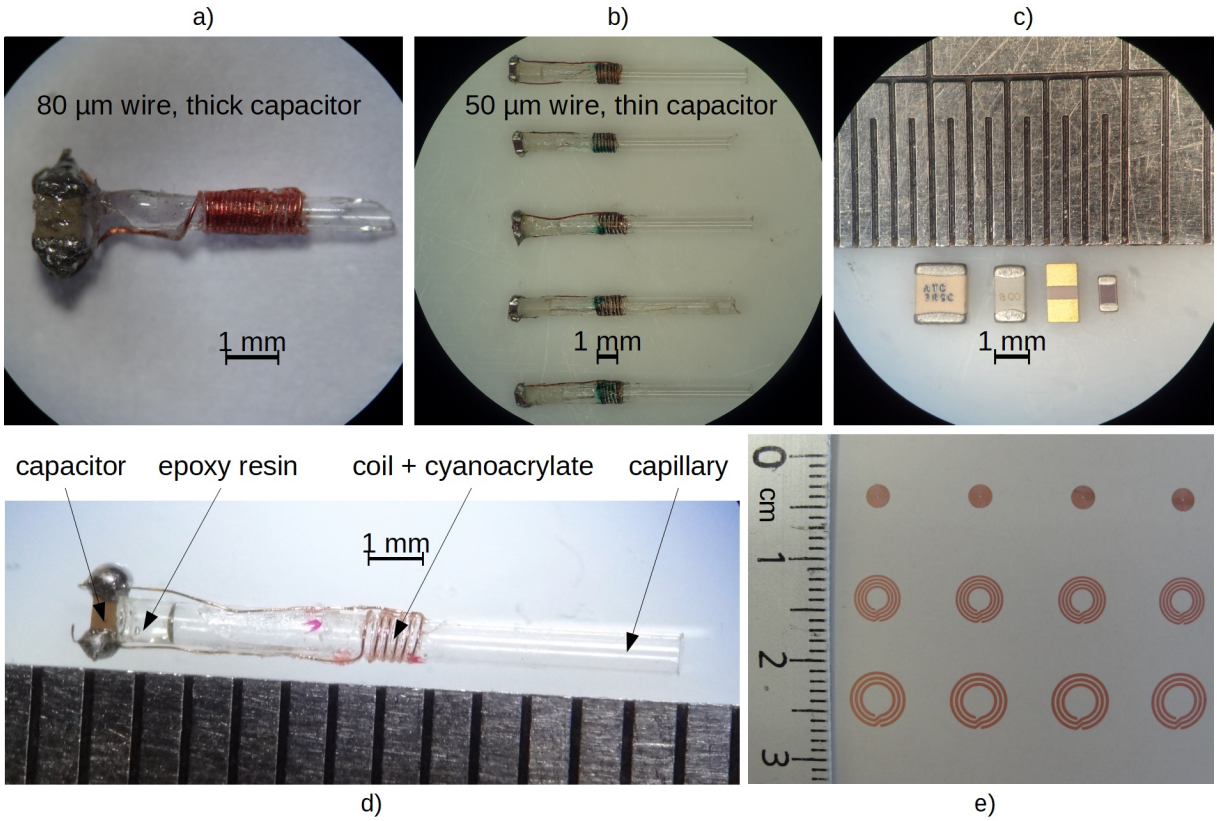


Figure II.16: Miniaturisation of coils, from a) 80 µm wire with thick capacitor to b) 50 µm wire with thin capacitor; c) different capacitors used; d) detail of a solenoid microcoil; e) some transmission line resonators (TLR).

D.3. Coil winding and capacitor brazing

Coil winding and capacitor brazing are mandatory steps in the manufacturing of solenoid microcoils. For TLR, the coil was etched by photo-lithography and capacitor was the substrate itself (*85*). Multiple techniques were tested to wind coils. Firstly, we used a winding apparatus (Figure II.17a). Unfortunately, this was oversized compared to the capillary and needed to apply strong mechanical force (Figure II.17b). Furthermore, it had multiple drawbacks: difficulty to clamp capillary without breaking it, difficulty to attach starting wire point, difficulty to apply adequate wire tension without breaking it, imprecision for coil position, and irregular winding for first turns. A second approach was to wind the coil around a calibrated screw (Figure II.17c). Such screws are described in mechanical engineering books (*108*) and various diameters are available, close from capillaries ones (500 to 870 µm).

The coil was very regular, however, it was destroyed when trying to enter a capillary into it. Finally, best results were obtained by hand winding under a binocular loupe (Figure II.17d).

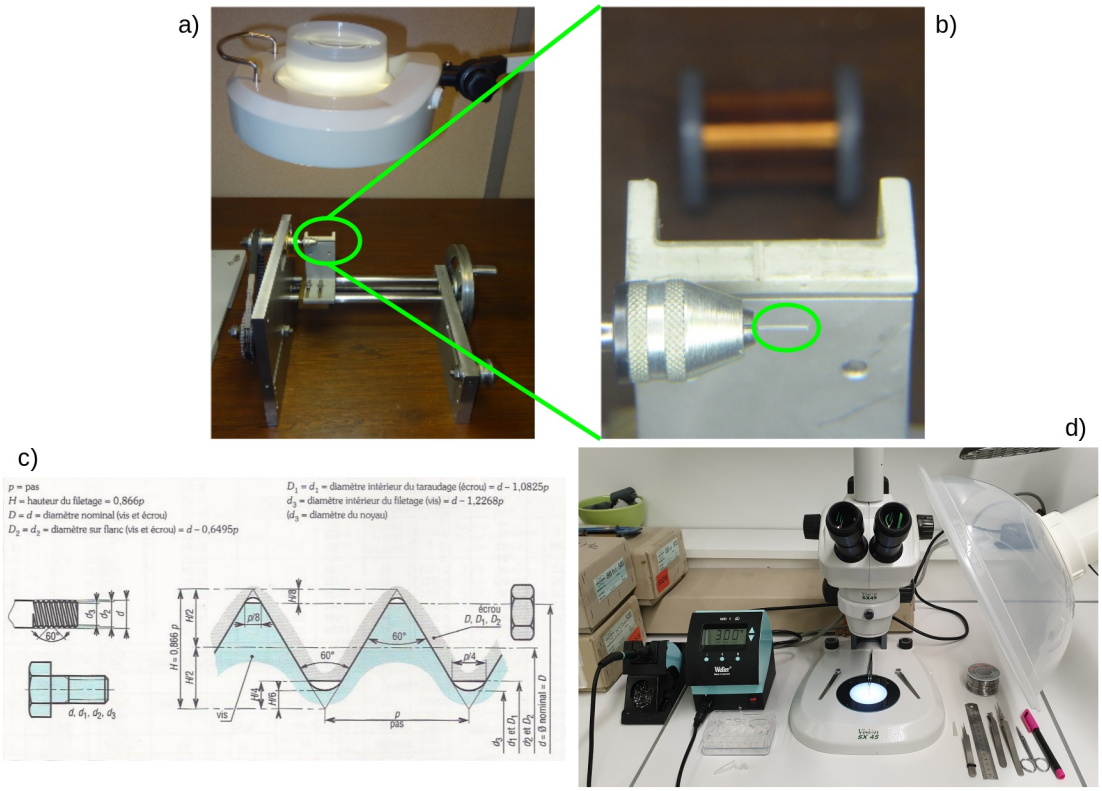


Figure II.17: Solenoid coil winding: a) mechanical tool, b) zoom to the capillary (green ellipses), c) calibrated screw (*108*), d) binocular loupe and brazing iron.

After winding, the solenoid was attached to the capillary by cyanoacrylate glue, which has a low viscosity, can easily be deposited around the coil, and dries in a few seconds. However, its drying needs water from air, and therefore it cannot be used inside the capillary to fix the capacitor. Epoxy resin has to be preferred for this purpose. This bi-component glue reacts by itself and has a higher viscosity. Fluorinated wax was also tried to fix the coil. It was melted by the brazing iron and solidified in a few seconds. Through this method, adjusting resonance frequency by changing pitch was quite easy. However, wax coating was thicker than cyanoacrylate coating and it had difficulties to enter into the Shapal-M insert. Moreover, despite having chosen wax with the highest available fluorinated content, strong ^{1}H signal were present, avoiding direct sample observation into the microcoil. ^{1}H signal of cyanoacrylate was much less visible. For completeness, cyanoacrylate ^{1}H and ^{13}C spectra can be found in (*109*).

Next step was to connect the capacitor to the coil through its leads. Soldering term is inadequately used for this operation, but brazing term should be preferred. *Soldering* is the addition of the same metal or fusion of the two parts. *Brazing* is the addition of a different material at low temperature (< 450°C), in our case brazing wire, composed of an alloy with tin, lead, silver and copper (*110*).

Figure II.18 evidenced three brazing qualities, from insufficient (left), through correct (middle), to excessive (right) (*111*). Insufficient brazing could break, whereas excessive brazing increased resistance and reduced microcoil quality factor. Preliminary step for brazing was to remove insulation from wire tips and to tin bare ends. Brazing quality strongly depended on insulation coating removal. We unsuccessfully tried chemical dissolution with chloroform or dichloromethane, that should dissolve polyurethane coating (*112*). New wire with polyamide coating was purchased, theoretically soluble in ethanol or acetone, but in practice it was insufficient to remove coating. Best results were obtained by wire scratching, which weakened it. In case of residual impurities, this gave a so-called cold brazing, with the wire entrapped into the brazing, but with a bad electric contact.

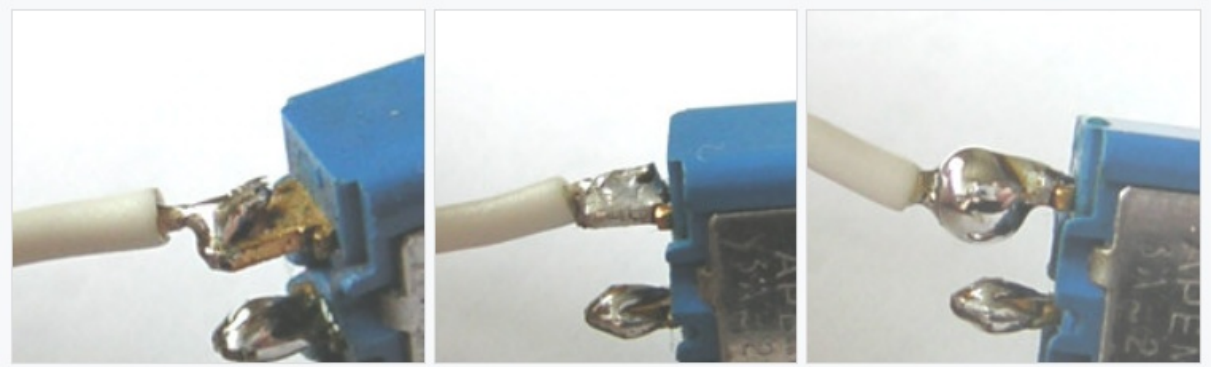


Figure II.18: Different brazing qualities: insufficient (left), correct (middle), and excessive (right) (*111*).

D.4. Frequency analysis

After preparing a microcoil, it was necessary to check its properties. The microcoil was inserted into a pick-up coil (Figure II.19a), connected to a frequency analyser (Figure II.19b), whose frequency was sweeping around the expected value. A dispersion signal on the screen reveals resonance frequency v, and the difference between positive and negative signals, Δv, is a measure of the quality factor Q:

$$Q = \frac{\nu}{\Delta\nu} \qquad \text{II.18}$$

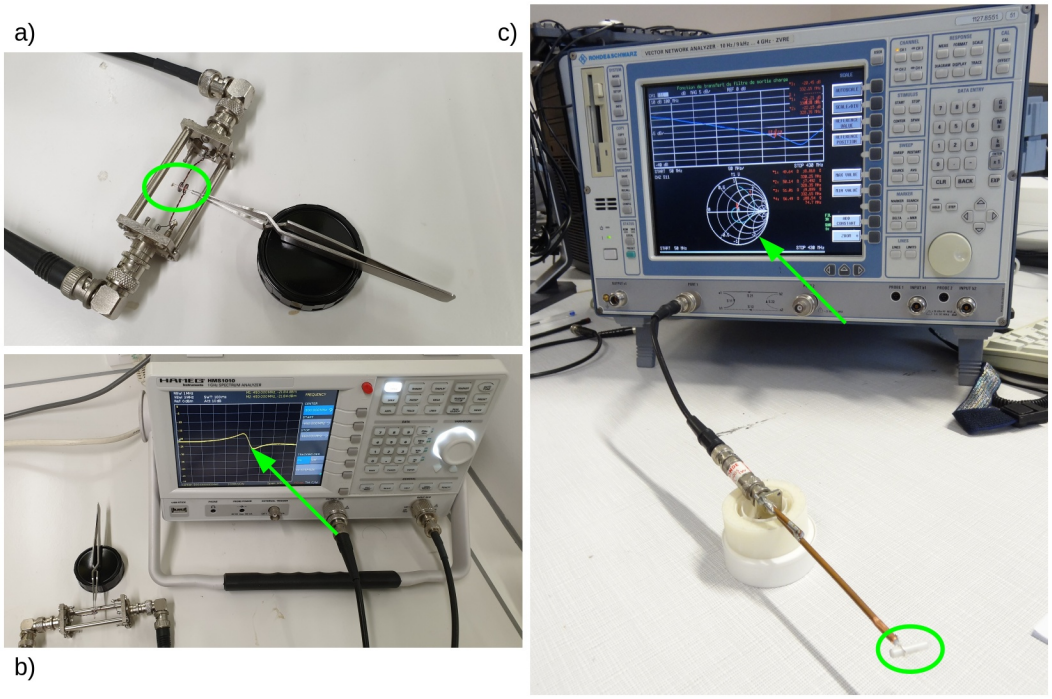


Figure II.19: Characterisation of microcoil: a) solenoid microcoil put into a pick-up coil, b) scalar, and c) vectorial frequency analyser. Green ellipse: microcoil; green arrows: frequency response.

Scalar frequency analyser (Figure II.19b) is for routine analysis and present only modulus, whereas vectorial frequency analyser (Figure II.19c) provides both modulus and phase, but is more expensive. The latter equipment could be helpful to obtain Smith's abacus (green arrow in Figure II.19c) and to optimise microcoil frequency response (*113*). Smith's abacus characterisations have been started in collaboration with A. Louis-Joseph at Polytechnique School (Palaiseau, France). However, this study is out of my knowledge expertise and has not been pushed further.

D.5. Micromechanics

The microcoil was placed into a Shapal-M insert, itself placed into a commercial rotor. Two designs were necessary, one for solenoids (Figure II.20 left), and the other one for TLR (Figure II.20 right). The insert had to be as close as possible from the cylindrical symmetry. Moreover, insert depended on rotor size and had to be sufficiently narrow to enter into and exit from the rotor only by gravity. However, if it was too narrow, spinning instabilities avoided high spinning frequencies. 10-20 µm was the adequate difference between the outer

diameter of the insert and the inner diameter of the rotor. As the inner diameter is slightly changing from one rotor to the other, an insert and a rotor form a pair that should not be separated. In mechanical engineering, this typically corresponds to the ISO norm H7 for the housing bore (rotor) and f7 for the shaft (insert) (*114*). Shapal-M is a ceramic that necessitates hard tools, and the necessary precision requires a strong mechanical expertise.

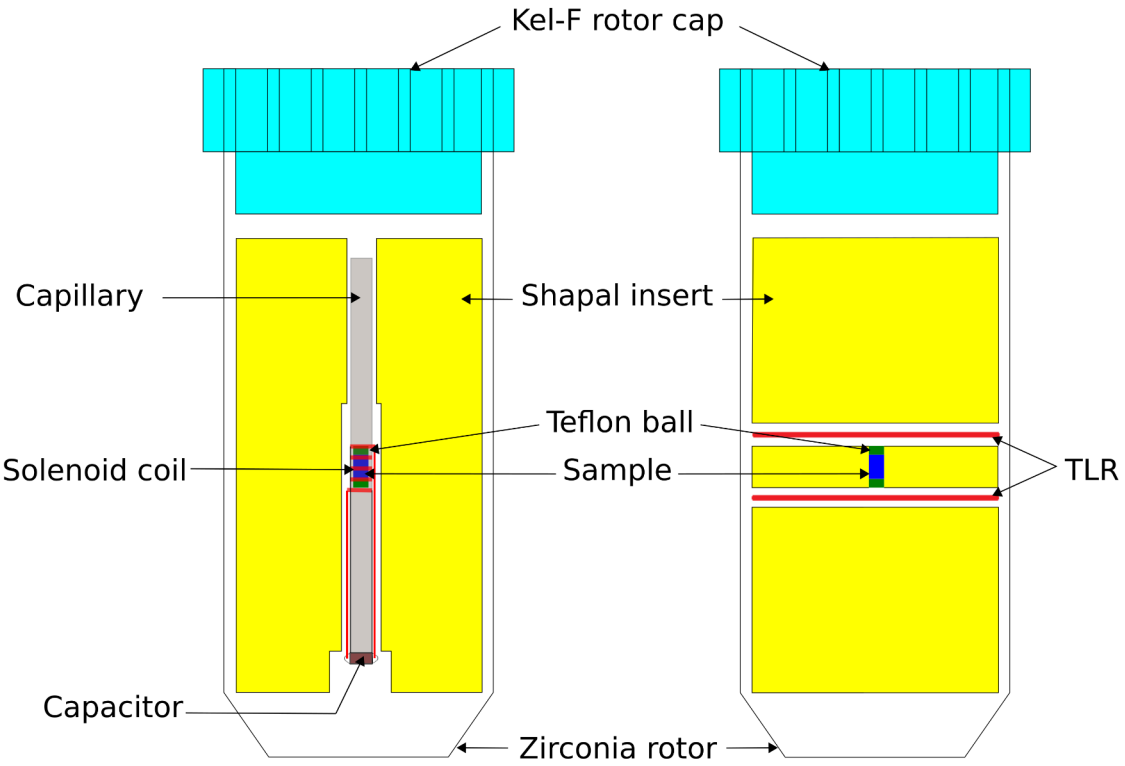


Figure II.20: Rotor inserts: solenoid design (left), TLR design (right).

A very precise protocol was defined to weight the amount of sample introduced into the microcoil. We tested three analytical balances displaying a precision of 10 µg, and kept the more precise one. First, a ball of Teflon ribbon was inserted into the capillary (solenoid) or into the holding disk (TLR). The capillary or the holding disk was then weighted three times, with a stable weight during 30 seconds and the balance doors closed. Between each measure, return to zero was checked and tare was adjusted if necessary. The weight was obtained by averaging the three measures. The same suite of operations was realised after adding the sample. Finally, a second Teflon ball was inserted to confine the sample. Using this procedure, a precision of ± 30 µg was obtained.

D.6. Supplies list

A list of supplies used for microcoils is available in Table II.5, including global equipment, and specific ones for solenoids or TLR.

Table II.5: Supplies list with provider, model, specifications and cost.
OD: outer diameter, ID: inner diameter, NA(): not available.

Designation	Brand / Provider	Model	Specifications	Cost excl. taxes
Global equipment				
Spectrum analyser	Hameg	HMS 1010	Max frequency 1 GHz	3050 €
Binocular loupe	Vision engineering	SX45	8x-50x magnifying	1350 €
Solenoids				
Copper wire	Elektrisola	FS15	Soluble polyamide coating, grade 1B, 0.05 and 0.03 mm	850 € for 3.15 kg
Quartz capillaries	Vitrocom / CM Scientific	Hollow round	OD :0.87, 0.84, 0.70 mm, ID : 0.70, 0.60, 0.50 mm	470 € for 200 units
Brazing station	Weller	WD1	Variable temperature 50-450°C	400 €
Capacitors	Vishay / Digi-Key	720-1223-ND 720-1224-ND	1.0-330 pF, 50 V, non-magnetic	180 € for 725 units
Brazing wire	NA()	NA()	0.25 mm, unknown composition	50 € for 250 g
Abrasive paper	Radiospares	797-5992	Extra-thin grain 1000	15 € for 25 units
Cyanoacrylate glue	Loctite / Radiospares	Super Glue 3	Liquid, fast drying	10 € for 3 g
Epoxy resin	Radiospares	850-940	Bi-component, fast drying	10 € for 32 g
TLR				
Printing	DB electronic (*85*)	NA()	Rogers RO3010 0.125 mm double side resist mask	2160 € for 500 units
Computer	Dell	Precision 7820	2*Intel Xeon silver 4110, 16C/32T, RAM 64 GB	2040 €
Simulations software	Altair	Feko	Academic licence	2015 € for 1 year

In Subchapter D, we provided a LibreOffice sheet allowing to calculate all necessary parameters for solenoid coil. Miniaturisation and microcoil preparation were detailed, before characterising microcoils, detailing the micromechanics and the adequate supplies list.

E. Results

First tests on sea urchin spines will be investigated in Section E.1. We will present RF homogeneity results obtained with MACS solenoids and TLR (Section E.2). Pulse sequences with increasing complexity will be used, starting with single pulse (Section E.3), and keeping on with PARIS / PISSARRO sequence (Section E.4), cross-polarisation (Section E.5), and TEDOR (Section E.6).

E.1. Sea urchin spine

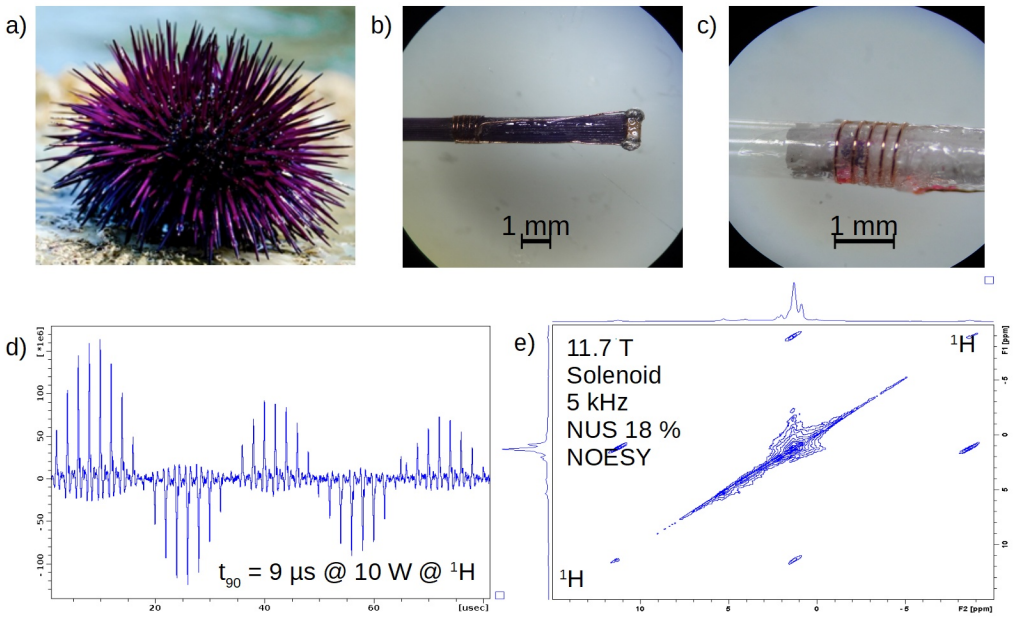


Figure II.21: a) Sea urchin, adapted from (*115*); b) microcoil winded around a sea urchin spine; c) piece of sea urchin spine into a capillary surrounded by a solenoid microcoil; d) nutation curve obtained with capillary microcoil; e) 2D NOESY spectrum, t_{mix} = 1 ms. Parameters: B_0 = 11.7 T, 4 mm rotor, *MAS* = 5 kHz, *RD* = 1 s.

Sea urchin spines (Figure II.21a) have the advantage to have a cylindrical symmetry, compatible with solenoid microcoil design. Hence, it was attempted to directly wind the microcoil around one spine of 10-15 mm long (Figure II.21b) collected on *Paracentrotus lividus* (Section E.4 of Chapter I) (*116*). However, RF homogeneity was very poor, in agreement to Section C.5, due to the important part of the sample outside of microcoil

volume. Better results were obtained with a piece of spine placed into a capillary surrounded by a solenoid microcoil (Figure II.21c and d). t_{90}^{1H} = 9 µs @ 10 W was measured, allowing to acquire a 2D Nuclear Overhauser Enhanced SpectroscopY (NOESY, (Figure II.21e). This spectrum presents mainly lipids signature and their spinning sidebands. Unfortunately, no correlation was visible on this spectrum, even at longer mixing times t_{mix}. Investigations will be pursued with samples chemically treated or heated, to focus on intercrystalline and intracrystalline organics.

E.2. RF homogeneity

Nutation curves allow to characterise RF homogeneity and to know the exact pulse length, in order to maximise signal to noise ratio (Section C.5).

E.2.a. Solenoid coils

For solenoids, amorphous/crystalline hydroxyapatite sample (Section E.2 of Chapter I) was introduced into the capillary and positioned into the microcoil. Without an adequate sample confinement, sample moved inside the capillary (right red circle on Figure II.22c) and no inversion was observed on ^{31}P nutation curve obtained without sample spinning of a mass m ~ 340 µg of hydroxyapatite (Figure II.22a). Indeed, during insertion or ejection of the rotor into the probe, a shock occurred at the end of these movements, shifting the sample. Furthermore, capacitor soldering was broken when spinning sample at MAS = 5 kHz (left red circle on Figure II.22c), due to centrifugal forces. To decrease breaking risks, the capacitor had to be blocked into the insert with Teflon ribbon. Similarly, Teflon can easily be stretched and inserted into the capillary in order to entrap the sample into the solenoid. Figure II.22b exhibit the ^{1}H nutation curve obtained with a slightly oversized adamantane sample (tricyclic $C_{10}H_{16}$, a reference compound used for ^{1}H chemical shift calibration) with a mass of m ~ 270 µg (Figure II.22d). The result is a partial inversion with curve distortion. By carefully restricting sample filling to the inner coil part, much better nutation curves were obtained (top insert on Figure II.22).

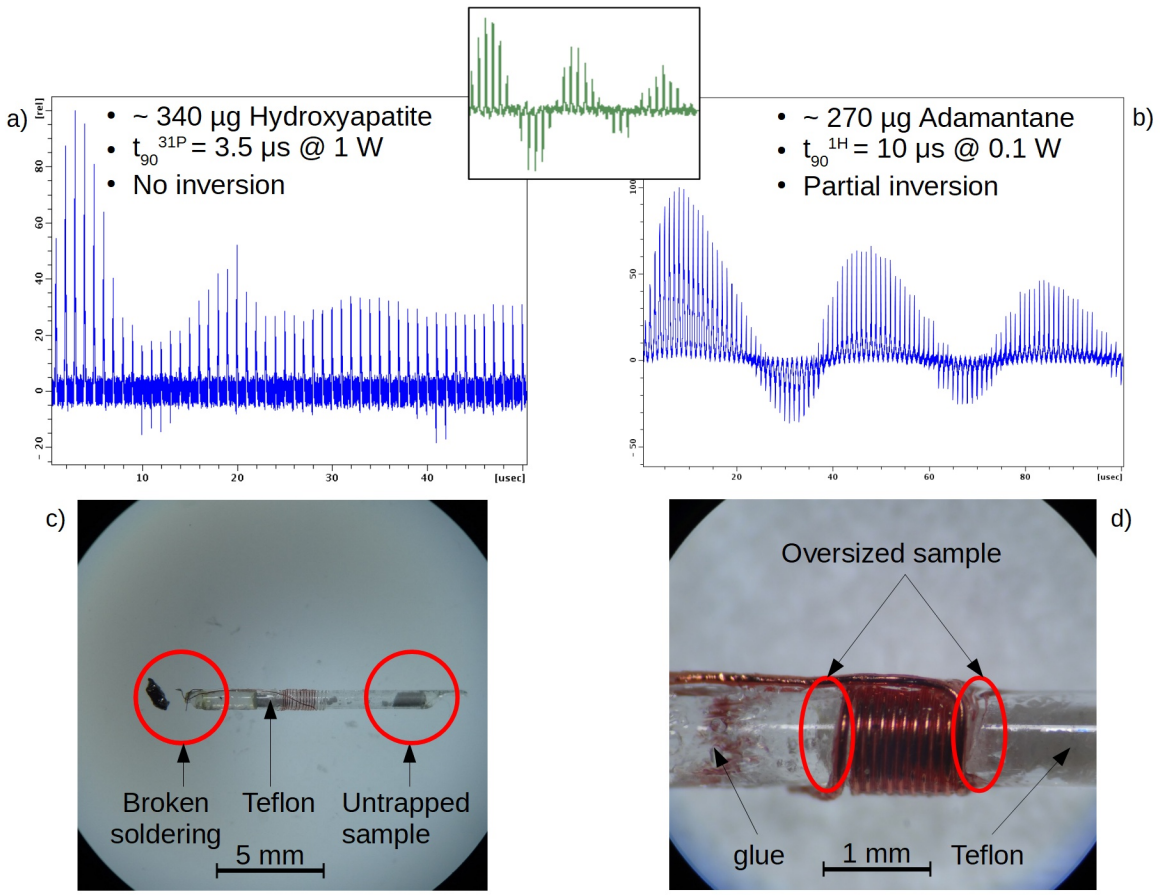


Figure II.22: Top insert: nutation curvewithout oversized sample; a) ^{31}P nutation curve obtained on hydroxyapatite; b) ^{1}H nutation curve obtained on adamantane; c) coil obtained after analysis of a); d) coil used for b). Red ellipses: defects of sample or of coil. Parameters: a) B_0 = 7.0 T, 4 mm rotor, MAS = 0 kHz, m ~ 340 µg; b) B_0 = 7.0 T, 4 mm rotor, MAS = 2.5 kHz, m ~ 270 µg

E.2.b. TLR

Confining a powder into TLR was harder than for MACS solenoids. For this purpose, two spacers were tested: 1.7 mm and 1.0 mm. The 1.7 mm spacer was a mini-crucible composed of a KEL-F ring containing sample, closed by a Teflon disk at one side, and a Teflon ball at the other side (Figure II.23a). The 1.0 mm spacer was a simplified version, with a second Teflon ball instead of the Teflon disk. Two TLR at the same resonance frequency were placed below and above. By combining two TLR, each of those being self-resonating, an increased RF field was obtained between them (Figure II.23b) (*60*).

For the 1.7 mm spacer, a correct ^{31}P nutation curve was obtained with m ~ 190 µg of hydroxyapatite (Figure II.23d), but the 90° pulse length was quite long with t_{90}^{31P} = 13.8 µs @ 10 W (MAS = 5 kHz). This value was very similar to the one without

microcoil, indicating that no enhancement was present. The reason was probably that sample was too far from the two TLR, due to spacer height. In order to circumvent this problem, the 1.0 mm spacer was placed between the two TLR with $m \sim 450$ µg of hydroxyapatite. A shorter value of $t_{90}^{31P} = 10.0$ µs @ 10 W was measured. According to the reciprocity principle (Section B.2), this shorter pulse length is equivalent to a gain in signal-to-noise ratio. In agreement, we measured a signal enhancement of ~ 40 % compared to the same spacer without TLR, *i.e.*, a gain in time of 2 (not shown). Further tests are in progress to calibrate the adequate distance for a maximum enhancement. A too short distance will increase mutual inductance and decrease resonance frequency. In such a case, TLR design has to be adjusted to compensate the resonance drift.

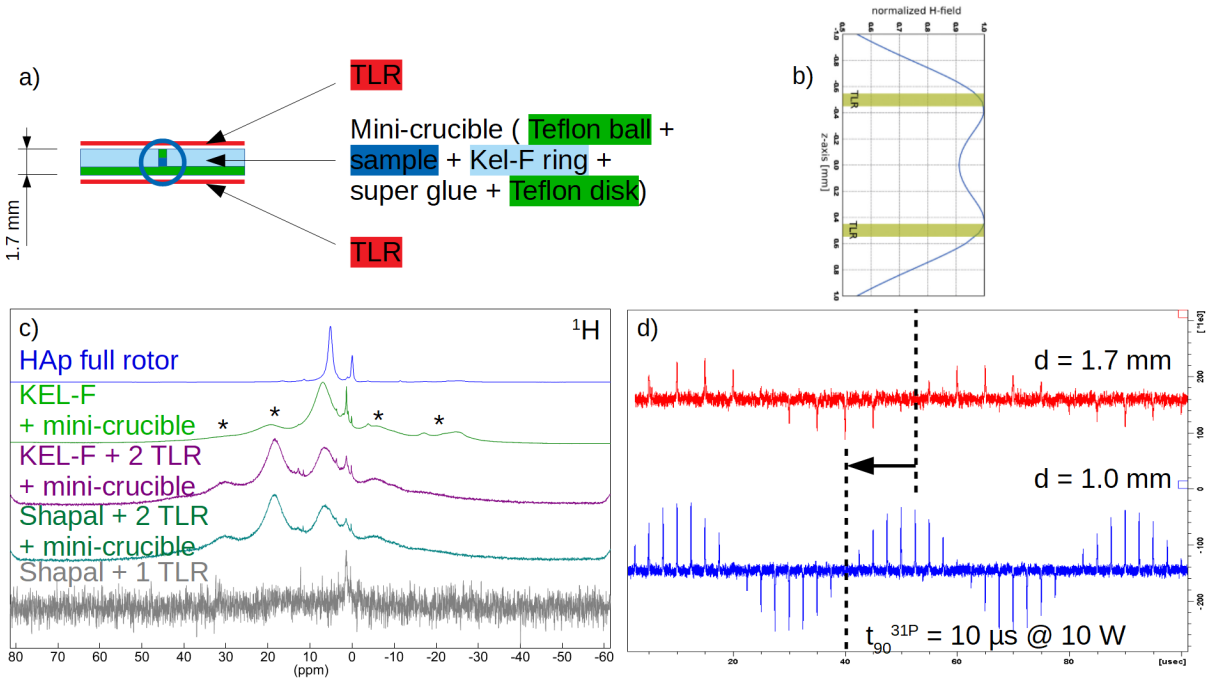


Figure II.23: a) Scheme of the mini-crucible; b) magnetic field between two TLR along their common axis, adapted from (*60*); c) ^{1}H residual signals with spacers and TLR, *: spinning side bands; d) ^{31}P nutation curve of amorphous/crystalline HAp as a function of TLR distance. Parameters: c) $B_0 = 16.4$ T, 4 mm rotor, $MAS = 8$ kHz; d) $B_0 = 11.7$ T, 7 mm rotor, $MAS = 5$ kHz, in red: $m \sim 190$ µg, $NS = 512$, $d_{spacer} = 1.7$ mm; in blue: $m \sim 450$ µg, $NS = 384$, $d_{spacer} = 1.0$ mm.

A drawback of the KEL-F spacer was the presence of unwanted ^{1}H signals (light green spectrum on Figure II.23c), which masked the signal of interest (blue spectrum), and was prohibitive for direct ^{1}H acquisition and calibration. In this case, the spacer as a whole was acting as a protonated sample. An interesting feature appeared in presence of TLR, with the KEL-F spacer filling the gap between the two TLR: spinning sidebands strongly increased

(indicated by stars), in particular the first left one at 20 ppm (purple and dark green spectra). A similar effect was observed by Tekely and Goldman on samples exceeding probe coil size (*117*) and was attributed to radial field. Such an explanation is compatible with TLR design. This effect is promising to investigate Magnetic Resonance Imaging (MRI) through MACS microcoils. As no Shapal-M spacer was available, one TLR was placed between top and bottom Shapal-M inserts. The obtained ^{1}H signal negligible (grey spectrum). New spacers will be manufactured in this material, which has also the advantage of favouring heat transfer and to avoid magnetic susceptibility discontinuities.

E.3. Single pulse

In order to compare solenoid and TLR designs, we first recorded their ^{31}P nutation curve on the same probe (Figure II.24a). At a power of 10 W, t_{90}^{31P} = 14.75 µs was measured in a full 7.0 mm rotor without microcoil (blue curve). With TLR and a spacer of 1.0 mm, a smaller value of t_{90}^{31P} = 10 µs was obtained (red curve). Solenoids gave the best results with t_{90}^{31P} = 3.25 µs (green curve) and damped regular oscillations. This corresponded to 77 kHz at only 10 W. The figure of merit of the resonator $B_1/\sqrt{P}$ was improved from 0.311 to 0.458 and 1.41 mT.W$^{-1/2}$ for probe alone, TLR and solenoid, respectively. Hence a gain of 1.5 was obtained with TLR and of 4.5 with solenoid. For comparison with previous studies, 3.7 was obtained on TLR (*60*) and 15.1 on solenoids (*7*). The reason of our lower enhancement is not clear but may be due to sample confinement between TLR or to coil shape for solenoid.

A similar quantity of hydroxyapatite was analysed with TLR and solenoids, with $m \sim 220 \pm 30$ and 250 ± 30 µg, respectively. A single pulse experiment was undergone, to better characterise the gain obtained by the two microcoil designs (Figure II.24b). This experiment is the most simple one and is well suited to test sensitivity, without risking to damage the microcoil by too much RF. Microcoils spectra were compared to their microcoil-free counterparts with the exact same amount of sample. Hence, the capacitor was unbrazed and capillary kept for solenoid, whereas coils were removed and spacer kept for TLR. Due to weighting incertitude, the microcoil-free measured intensities were not exactly the same and the capillary spectrum was scaled down by 0.75 to match the spacer spectrum. The same factor was applied on solenoid spectrum. A signal gain of 2.0 and 2.6, *i.e.*, a time gain of 4.0 and 6.8 were obtained for TLR and solenoids against no microcoil, respectively, highlighting that a better sensitivity improvement was obtained for solenoids.

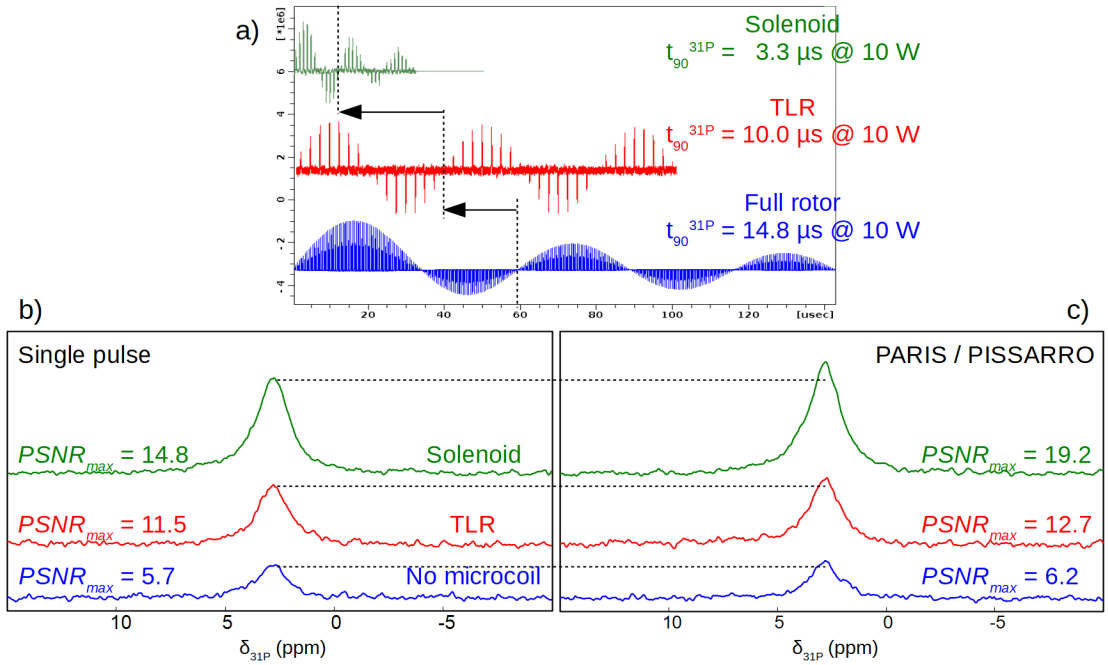


Figure II.24: Comparison of TLR (red curve, $m \sim 220 \pm 30$ µg) and solenoid microcoils (green curve, $m \sim 250 \pm 30$ µg) on amorphous/crystalline HAp: a) ^{31}P nutation, b) single pulse experiment, c) PARIS / PISSARRO experiment. $PSNR_{max}$ is indicated near each spectrum. Intensity and $PSNR_{max}$ value of solenoid have been scaled down by 0.75 to be coherent between TLR spacer (blue curve) and capillary (not shown), both without microcoils. Parameters: B_0 = 11.7 T, 7 mm rotor, MAS = 5 kHz, RD = 10 s.

E.4. PARIS and PISSARRO

To go a step further without damaging microcoils, we used the low-power multipulses sequences Phase-Alternated Recoupling Irradiation Scheme (PARIS) (*118*) and Phase-Inverted Supercycled Sequence for Attenuation of Rotary ResOnance (PISSARRO) (*119*). These sequences were initially designed for very fast MAS = 30 kHz.

PARIS is presented on Figure II.25a, where τ_p is the pulse duration and τ_r is the MAS period. It was preceded by a saturation step (typically a train of pulses with decreasing inter pulses delays) to ensure relaxation stability. By low-power irradiation on ^{1}H nucleus during relaxation delay, broadening of peaks occurs on the other nucleus (Figure II.25b), allowing its spin diffusion and enhancing relaxation of non-protonated species. Figure II.25d presents the ^{13}C relaxation of L-alanine COOH group in absence of irradiation (circles), with Continuous Wave (CW) irradiation (triangles) and with PARIS recoupling (squares) (*120*). In the latter case, a much faster relaxation was observed. PARIS has been improved to $PARIS_{xy}$ with additional phase cycling (*121*). Moreover, another similar experiment is available, with

different phases: Second-order Hamiltonian among Analogous Nuclei Generated by Hetero-nuclear Assistance Irradiation (SHANGAI) (*122*).

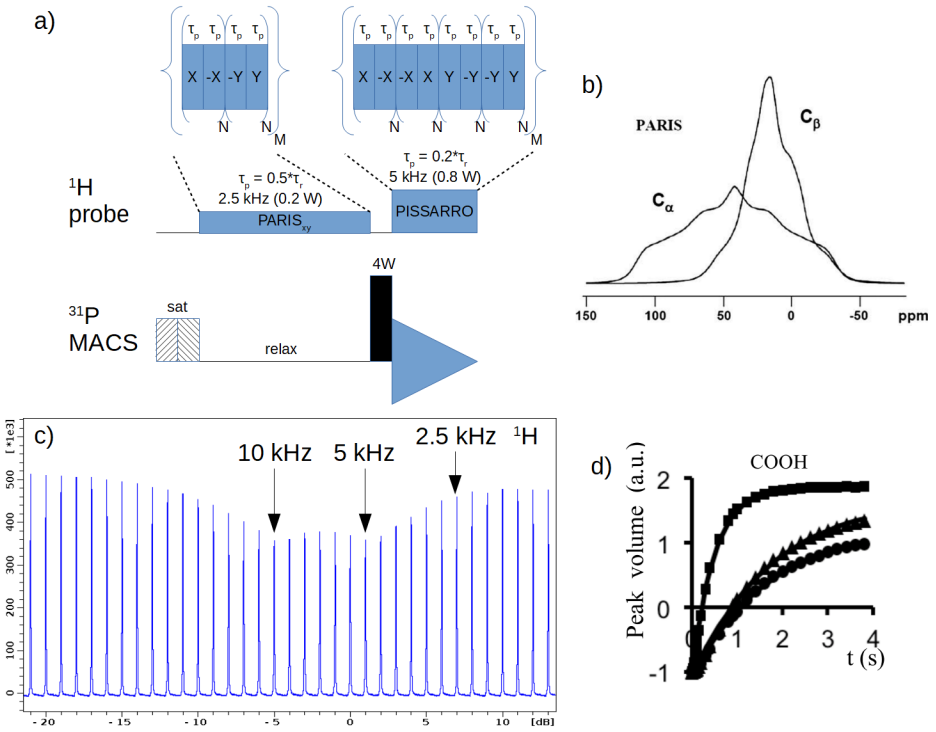


Figure II.25: a) PARIS$_{xy}$ / PISSARRO pulse sequence; b) broadening of L-alanine ^{13}C peaks simulated in presence of PARIS (*118*); c) illustration of R^3 conditions on ^{31}P spectrum of amorphous/crystalline HAp with *MAS* = 5 kHz and ^{1}H CW decoupling power ranging from 126 W to 0.05 W; d) L-alanine ^{13}C T_1 measurements with no recoupling (circles), continuous wave irradiation (triangles) or PARIS recoupling (squares) (*120*).

In addition, PISSARRO was applied for low-power decoupling during acquisition (Figure II.25a). PARIS and PISSARRO are very similar pulse sequences: a pulse is applied during a time τ_p, then a phase alternation from X to -X is repeated *N* times, and finally the phase is incremented. The overall cycle is repeated *M* times, to fit the necessary duration. The main difference between PARIS and PISSARRO is that τ_p and *RF* values are different, with $\tau_p = 0.5 * \tau_r$ and $RF = MAS / 2$ for PARIS and $\tau_p = 0.2 * \tau_r$ and $RF = MAS$ or $MAS * 2$ for PISSARRO. Both are using Rotary Resonance Recoupling (R^3) conditions (*123*). This effect appears when RF irradiation frequency is a (sub-)multiple *n* of MAS frequency. Chemical shift anisotropy and hetero-nuclear dipolar interactions are refocused for $n = 2$ or 1, whereas homonuclear dipolar interaction is reintroduced for $n = 1$ or ½. Figure II.25c evidences this effect for $RF = 10$, 5, and 2.5 kHz (black arrows) and $MAS = 5$ kHz on ^{31}P intensity of HydroxyApatite (HAp) in presence of ^{1}H CW decoupling with power ranging from 126 W to 0.05 W. According to its structure (*124*, *125*), crystalline HAp is not very sensitive to

decoupling as ^{1}H-^{31}P heteronuclear dipolar interaction ($D_{1H\text{-}31P}$ = -853 Hz, 3.85 Å) and ^{1}H-^{1}H homonuclear dipolar interaction ($D_{1H\text{-}1H}$ = -2951 Hz, 3.44 Å) are easily averaged by slow MAS. The visible effect is probably coming rather from sample's amorphous part, which has stronger dipolar interactions. A better decoupling effect could be obtained for phoshines with direct H-P bond.

On Figure II.26a, CW, PISSARRO and SPINAL-64 (*126*) decouplings pulse sequences were compared to the absence of decoupling for a full rotor of amorphous/crystalline HAp. PISSARRO gain at RF = 5 kHz was marginal, lower than the standard SPINAL-64 decoupling at RF = 56 kHz, whereas signal intensity decreased with CW decoupling at RF = 2.5 kHz, highlighting unexpected peak broadening. In addition, homonuclear recoupling was applied during the entire relaxation delay RD = 10 s (Figure II.26b). A similar intensity gain of 18-20 % was obtained for all tested pulse sequences at RF = 2.5 kHz (CW, PARIS$_{xy}$ with N = 1 or 2 cycles, and SHANGAI). On HAp sample, this sensitivity gain could also be a consequence of heteronuclear Overhauser effect during ^{1}H irradiation (*127*). Further investigations would be needed to check this hypothesis. The effect of combined PARIS$_{xy}$ and PISSARRO pulse sequences on microcoils is presented on Figure II.24c and confirmed the above results with an average gain of 16 %. No degradation of microcoils, neither TLR nor solenoids, was observed with these low-power multipulses experiments. Consequently, if microcoil is tuned to another nucleus than ^{1}H, these experiments are suitable to enhance the second channel signal, *e.g.*, for ^{31}P pulse calibration.

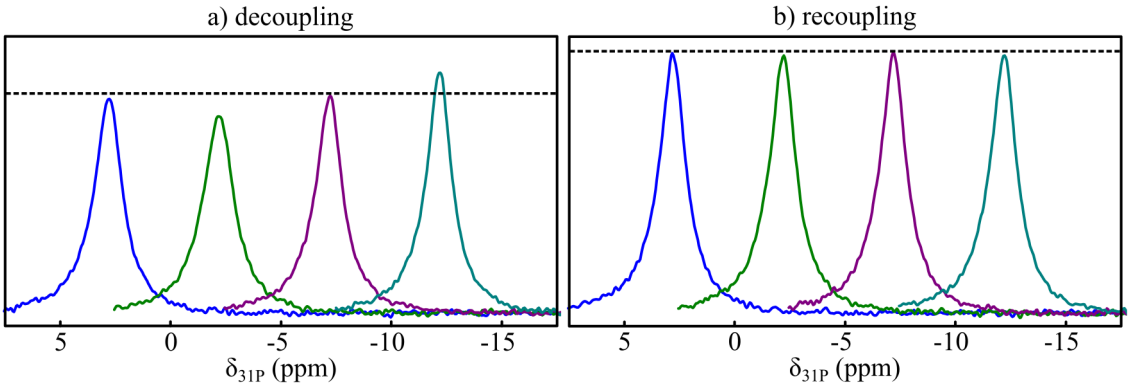


Figure II.26: a) Comparison of decoupling pulse sequences, from left to right: no decoupling, CW (RF = 2.5 kHz), PISSARRO (RF = 5 kHz), SPINAL-64 (RF = 56 kHz). b) Comparison of recoupling sequences during RD, in presence of PISSARRO decoupling (RF = 5 kHz) during acquisition, from left to right: CW (RF = 2.5 kHz), PARIS$_{xy}$ (N = 1, RF = 2.5 kHz), PARIS$_{xy}$ (N = 2, RF = 2.5 kHz), SHANGAI (RF = 2.5 kHz). Parameters: amorphous/crystalline HAp, B_0 = 11.7 T, 4 mm rotor, MAS = 5 kHz, RD = 10 s. Dashed lines highlight a) PISSARRO and b) PARIS$_{xy}$ (N = 2) intensities.

E.5. Cross-polarisation

Cross-Polarisation (CP) experiment is presented on (Figure II.27a). By varying CP delay, distance and / or relative mobility can be checked. Two dimensional spectra can also easily be obtained, highlighting chemical correlations. {^{1}H}-^{31}P CP was tested on a pair of ^{31}P tuned TLR. Low power at 1-10 W was applied on ^{31}P channel, whereas a power at 190 W was necessary on ^{1}H nucleus, corresponding to probe coil RF field. After only a few scans, probe detuning occurred, indicating damaging of microcoil. After one night of acquisition, microcoils were burned (red circle on Figure II.27b). Not only a hole was present on TLR substrate, but also copper track was severely darkened. None of the two TLR present into the rotor during the experiment were reusable, and they burned at the exactly same point, corresponding to the inner extremity of copper track. An explanation could be the intense electric field on ^{1}H channel, that led to substrate breakdown, despite the low ^{31}P power.

Power on microcoil channel is not problematic, gratefully to the RF field enhancement, and 10 W will be enough in many cases. However, for double resonance experiments, the probe coil channel requires a much higher power for the same RF field. It is not uncommon to need 200 W on the main coil for CP experiments, depending on the probe used. As presented in former paragraph, such power is unsustainable for TLR and may also damage solenoids. Nevertheless, we would like to highlight in the following paragraph that CP could be undergone on microcoils, if power dissipation on both channel is carefully controlled.

A probe with a higher quality factor was successfully used to test CP experiments on a solenoid coil fixed with fluorinated wax rather than with cyanoacrylate glue, and containing $m \sim 110 \pm 30$ µg of HAp. In the present case, a moderate power at 33 W was used on ^{31}P probe coil channel, whereas the microcoil was tuned to ^{1}H with 0.9 W. It was even possible to implement double CP, with acquisition on ^{1}H nucleus (Figure II.27c). This experiment proved useful to study ^{1}H interfaces with an increased sensitivity and resolution (*128*, *129*). With direct acquisition (green spectrum on Figure II.27d), strong residual ^{1}H signals were visible, typical of a long unsaturated carbon chain, attributed to wax, and HAp signal was only present as a shoulder. With ^{1}H-^{31}P CP filtered experiment (blue curve), the wax was no longer visible and HAp peak was enhanced. This demonstrated that CP or double CP experiments are conceivable with MACS, providing that probe quality factor is high enough to use moderate power on probe coil channel.

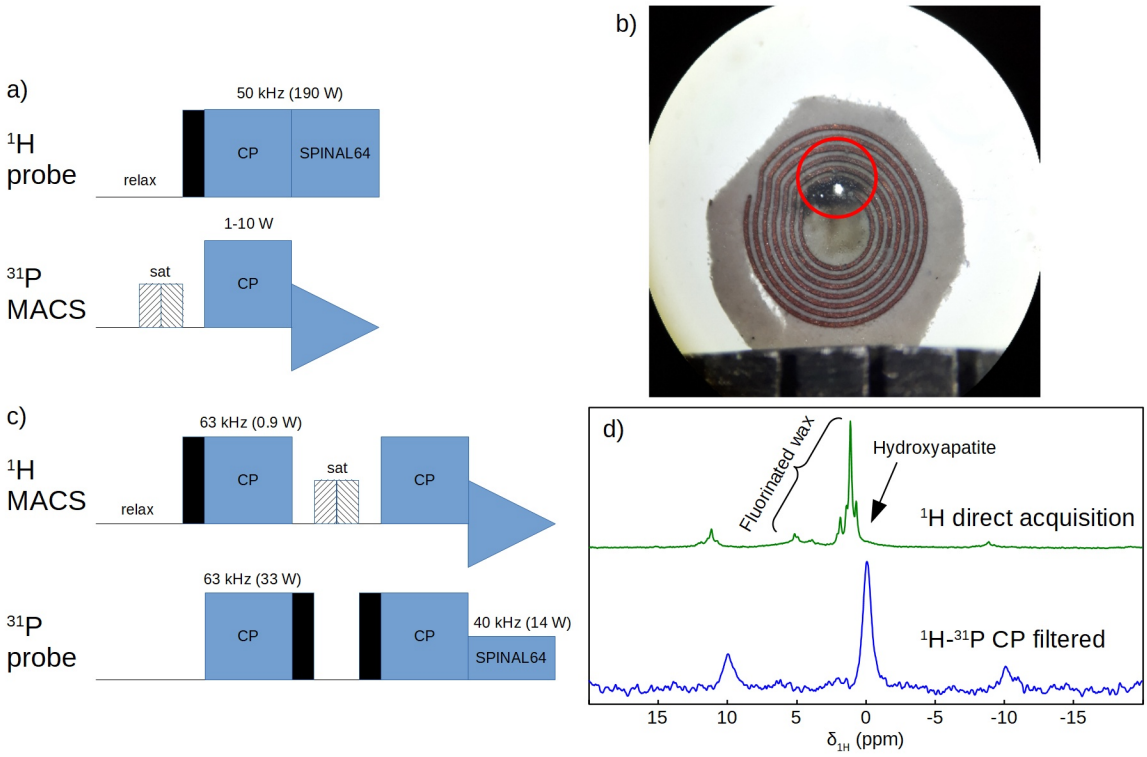


Figure II.27: a) CP experiment; b) burned TLR during CP; c) ^{1}H-{^{31}P} CP filtered experiment; d) spectra obtained on solenoid with ^{1}H direct acquisition (green, $NS = 4$) and ^{1}H-{^{31}P} CP filtered experiment (blue, $NS = 2048$, $t_{exp} = 0.6$ h, $t_{cp1} = t_{cp2} = 10$ ms). Parameters: b) $B_0 = 11.7$ T, 7 mm rotor, $MAS = 5$ kHz, $RD = 1$ s; d) $B_0 = 7.0$ T, 4 mm rotor, $MAS = 3$ kHz, $RD = 1$ s, $m \sim 110 \pm 30$ µg of HAp.

A similar ^{1}H-^{43}Ca double CP 2D experiment was previously published on a ^{43}Ca labelled $Ca(OH)_2$ sample entrapped into MACS solenoid (*58*). In this case, sample amount was limited by cost of ^{43}Ca enrichment. As ^{43}Ca is a quadrupolar nucleus ($I = 7/2$), a very limited RF field of 4 kHz was requested to excite the central transition with CP. We estimated the corresponding power to 0.1 W on ^{43}Ca probe channel, whereas ^{1}H microcoil channel power was adjusted to 10 W.

Hereafter, we limited ourself to 10 W on both channels, which proved to be safe to not destroy microcoils. A value of 50 W could have been an alternative, but it has not been thoroughly tested.

E.6. TEDOR

In this section we tried to implement a low power alternative to CP experiments, compatible with microcoils. As microcoils were not well suited to develop pulse sequences, due to the low sample amount and its inherent lack of signal, full rotor with a reference

compound was preferred. In the continuity of previous section, we tried to correlate ^{1}H to ^{31}P, but with a lower power, including on probes with a poor quality factor. A list of available sequences to transfer magnetisation, depending on the interaction chosen, were presented in (*130*).

Rotational Echo Double Resonance (REDOR) (*131*) is a sequence which refocusses heteronuclear dipolar interaction through concerted manipulation of spin magnetisation and sample spinning. A spin echo is applied to one nucleus (^{1}H), while a train of 180° pulses, two per rotor period, is applied to the other nucleus (^{31}P). The local dipolar field is reversed at each 180° pulse on ^{31}P (*132*). As a consequence, the dipolar interaction is no longer averaged to zero by MAS. XY-8 phase cycling proved to be fruitful to decrease transversal relaxation during REDOR recoupling (*133*). Transferred-Echo DOuble-Resonance (TEDOR) (*134*) is based on REDOR, but with initial excitation on ^{1}H and signal acquisition on ^{31}P. TEDOR can be used either with rotor synchronised acquisition to obtain a dipolar coupling curve, or to directly obtain a spectrum, similarly to Insensitive Nuclei Enhanced by Polarization Transfer (INEPT) (*135*), which is used as a building block into many liquid-state experiments.

INEPT and TEDOR are presented on Figure II.28a and b, respectively. Firstly, ^{1}H magnetisation is prepared during evolution step, secondly it's transferred to the other nucleus, finally it's phased during refocussing step. In TEDOR, the 180° pulses at the middle of INEPT evolution and refocussing steps are replaced by REDOR recoupling. Magnetisation is transferred through bonds during INEPT (scalar interaction J), whereas it's transferred through space during TEDOR (dipolar interaction D). TEDOR can be generalised as D-INEPT, whereas J-INEPT term is used for differencing scalar pulse sequence. J-INEPT can be used for mobile species such as those entrapped into mesoporous materials (*136*) or for crystalline materials (*137*, *138*), both having long apparent transversal relaxation times T_2^*. On the contrary, amorphous samples have shorter T_2^* and are therefore best analysed with D interaction. As D is stronger than J, shorter evolution and refocussing delays are requested. Moreover, D-INEPT can be used to investigate through space non-covalent bonds. Other D-INEPT implementations exists but TEDOR has the advantage to use discrete pulses instead of continuous ones, which can limit RF electric field and microcoils degradation. Dipolar Heteronuclear Multiple Quantum Correlation (D-HMQC) is another alternative (*139*, *140*), with acquisition on ^{1}H rather than on ^{31}P, which was not investigated here.

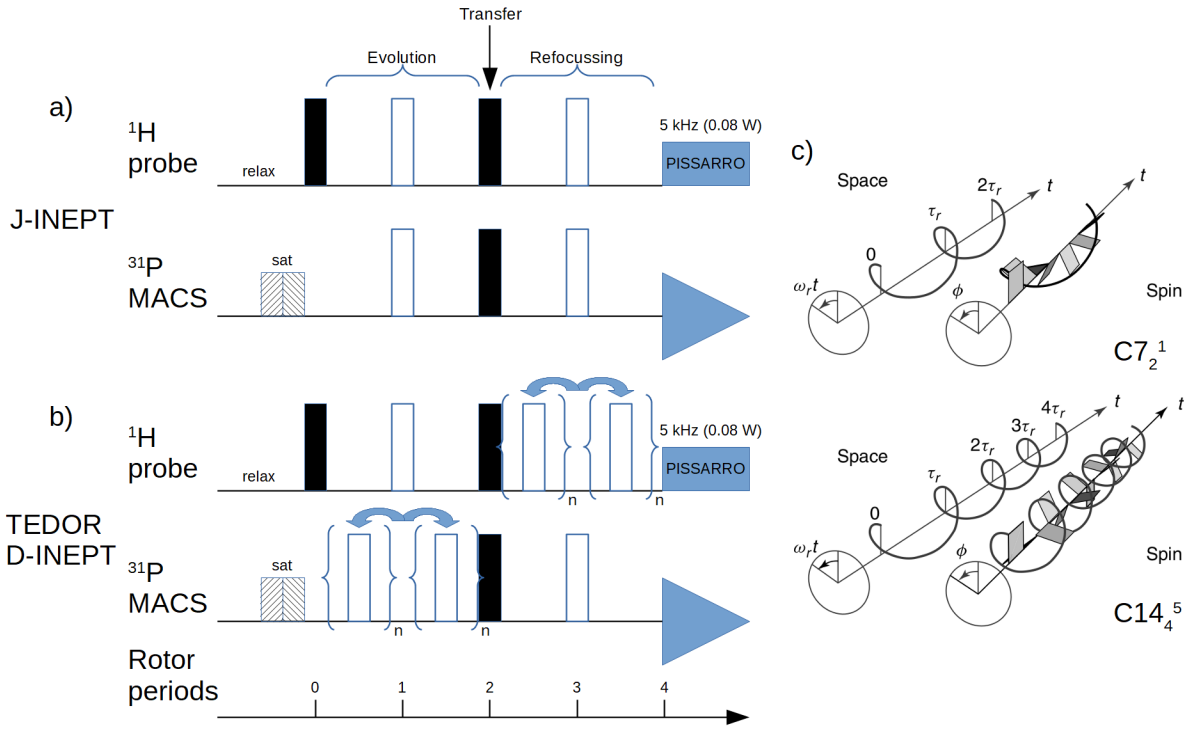


Figure II.28: Comparison of a) J-INEPT and b) TEDOR / D-INEPT experiments;
c) illustration of concerted manipulation of spin magnetisation and sample spinning (*141*).

Before using D-INEPT sequences, it is necessary to define Levitt's notation, used into dipolar recoupling blocks, such as $R2_1^1$ and $SR4_1^2$. Levitt highlighted that rotor synchronised pulse sequences can be used to reintroduce or on the contrary to suppress various interactions, leading to homo- or hetero-nuclear de- or re-coupling (*141*). This was summarised into two families: CN_n^v and RN_n^v, where *N* impulsions are phase incremented to follow *v* spin rotation during *n* rotor periods. This is illustrated on Figure II.28c with $C7_2^1$ pulse sequence, consisting of 7 phase increments forming 1 spin rotation during 2 rotor periods. The more complex $C14_4^5$ is also represented. The difference between CN_n^v and RN_n^v is that each (composite-) impulsion consists of a multiple of 360° in CN_n^v, whereas 180°/-180° cycles are subsequently phased in RN_n^v. The amount of power needed depends on the chosen pulse sequence. According to Levitt's nomenclature, REDOR corresponds to $R4_2^1$ (*141*). S letter in $SR4_1^2$ indicates a supercycle.

Whether REDOR irradiation is placed on ^{1}H channel (HH), or on ^{31}P channel (XX), or even crossed (HX or XH), is questionable and different choices were tested in the literature (*142*, *143*). ^{1}H-^{31}P TEDOR with HH, HX, XH, and XX irradiation, as well as another D-INEPT implementation with $R2_1^1$ applied on HH were tested on amorphous/crystalline HAp (Figure II.29a), with broad (unfavourable T_2^*) and narrow (favourable T_2^*) species. TEDOR

pulse sequence with HH irradiation seemed to gave the best results. Despite negligible J-coupling was present into this sample, J-INEPT was applied to check that no resiual magnetisation was passing through the phase cycling.

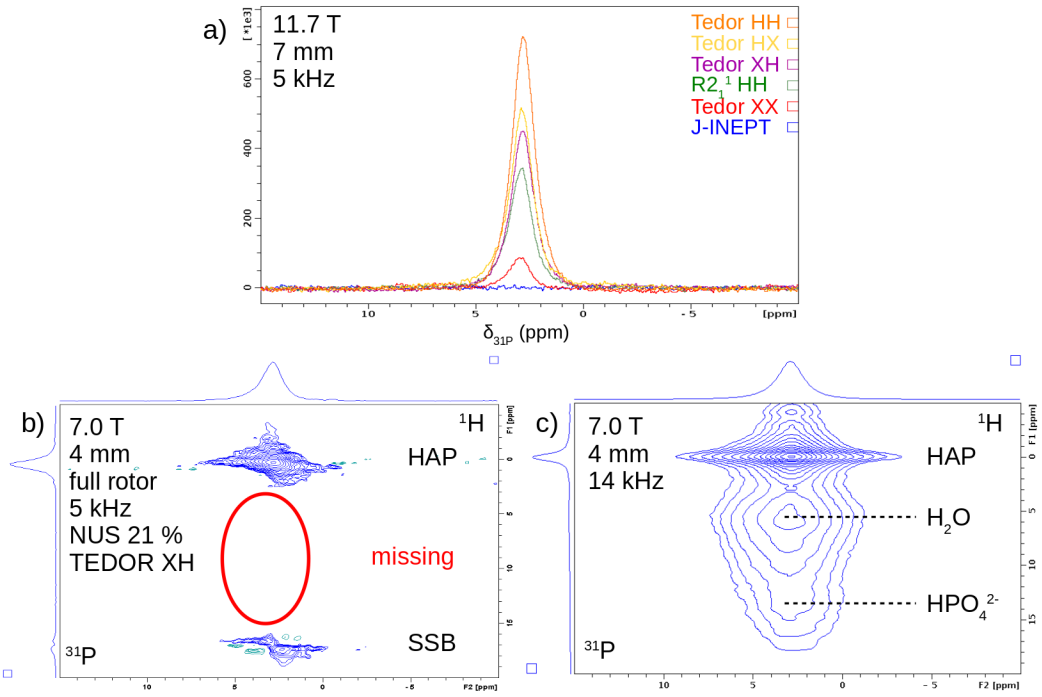


Figure II.29: a) Comparison of different D-INEPT pulse sequences; b) 2D TEDOR XH experiment; c) 2D CP HETCOR experiment. Parameters: a) B_0 = 11.7 T, 7 mm rotor, *MAS* = 5 kHz, *RD* = 1 s; b) B_0 = 7.0 T, 4 mm rotor, *MAS* = 5 kHz, *RD* = 1 s, 21% hybrid sampling NUS; c) B_0 = 7.0 T, 4 mm rotor, *MAS* = 5 kHz, *RD* = 1 s. SSB: Spinning SideBands.

TEDOR can be implemented as a two-dimensional (2D) experiment, by adding a frequency coding evolution after the first ^{1}H pulse (*144*). Figure II.29b presents the obtained 2D TEDOR spectrum of HAp. For comparison, the same sample analysed by 2D CP HETCOR is presented on Figure II.29c. One can notice that the broad peak between 3 and 18 ppm, attributed to adsorbed water and amorphous hydrogenophosphate species, was missing with 2D TEDOR. This was not a consequence of hybrid Non-Uniform Sampling (NUS) applied during acquisition, whose conditions were carefully tested (Section E.5 of Chapter IV), but was rather due to short transverse relaxation of the broad peak.

Surprisingly, TEDOR signal-to-noise ratio was much poorer than with CP (not shown), despite Christiansen *et al.* highlighted that {1H}→^{15}N→^{13}C TEDOR and CP efficiency could be similar on doubly labelled ^{13}C-^{15}N glycine (*145*). A possible cause could be an unoptimised phase cycling or rotor synchronisation. The absence of the central 180° on

same channel than REDOR cycles, as denoted in an early publication (*144*) but in contradiction with many other ones, could be another explanation. This could also explain lineshapes distortions on 2D TEDOR spectrum. Additional tests will be undergone with rotary resonance recoupling, with $RF = MAS * 2$, or with $SR4_1^2$, both simultaneously refocusing heteronuclear dipolar recoupling, while decoupling the homonuclear proton dipolar interactions (*146*).

In Subchapter E, we confirmed that perfect sample confinement into the microcoil was necessary for proper RF homogeneity. Better enhancement was obtained for solenoid than for TLR. PISSARRO pulse sequence improved *PSNR* by ~ 16 % on amorphous/crystalline HAp. CP destroyed microcoil unless low power was used on both channels. TEDOR and D-INEPT were investigated to transfer magnetisation from ^{1}H to ^{31}P.

F. Chapter conclusion

In this chapter, we focussed on small sample amounts. First, we highlighted that microcoils are more sensitive than commercial probes to study microquantities (Subchapter B). Basics of electromagnetism were provided in Subchapter C. In particular, wire diameter has to be a compromise between skin depth, ease of manipulation and eddy currents. Best quality factor is obtained with a pitch equal to 1.5 times wire diameter. Moreover, necessary formulae to design microcoils were given. Optimal RF homogeneity requires $l_{sample} : d_{coil} : l_{coil}$ dimensions of 0.5 : 1 : 1.2. Practical details to manufacture microcoils were provided in Subchapter D, with warnings on capacitor size and insert dimensions. The latter needed to be 10-20 µm smaller than inner rotor diameter. All necessary supplies were listed. Results were presented in Subchapter E, starting with preliminary tests were on sea urchin spines and RF homogeneity tests. With TLR microcoils, dissymmetric spinning sidebands were obtained without sample confinement. This effect is promising for imaging. TLR and Solenoids were compared using nutation, single pulse and PARIS /PISSARRO experiments. A time gain of 4.0 and 6.8 was obtained for TLR and solenoids, respectively. PARIS / PISSARRO was useful to gain 16 % of signal during pulse calibration. CP was applied on both microcoils, but the necessary high power burned TLR. TEDOR sequence was thus implemented for low power magnetisation transfer between nuclei.

We compared solenoid and fat coils in Table II.6. Solenoids are easy to design, have a high RF homogeneity and support quite large RF power. However, they are easily breakable,

especially at capacitor soldering. Sample spinning may be difficult due to the lack of cylindrical symmetry. On the contrary, TLR are easy to spin, but harder to design and support less power. Sample spacer has to be lower than 1 mm.

Table II.6: Comparison of solenoids and TLR.

Characteristics	Solenoid coil	TLR
Design	☺	☹
Manipulation	😐	☺
Mechanical tolerance	😐	😐
Ease of spin	☹	☺
RF homogeneity	☺	😐
RF power	☺	😐
Breaking point	Capacitor soldering	Substrate

G. Chapter bibliography

1. L. Znaidi, G. J. A. A. Soler Illia, S. Benyahia, C. Sanchez, A. V. Kanaev, *Thin Solid Films*. **428**, 257–262 (2003).
2. A. Demessence *et al.*, *Chem. Commun.*, 7149–7151 (2009).
3. B. Fousseret *et al.*, *Chem. Mater.* **22**, 3875–3883 (2010).
4. C. Sanchez, C. Boissière, D. Grosso, C. Laberty, L. Nicole, *Chem. Mater.* **20**, 682–737 (2008).
5. M. Faustini, D. Grosso, C. Boissière, R. Backov, C. Sanchez, *J. Sol-Gel Sci. Technol.* **70**, 216–226 (2014).
6. D. Bazin, M. Daudon, *J. Phys. Appl. Phys.* **45**, 383001, 1–10 (2012).
7. D. Sakellariou, G. Le Goff, J.-F. Jacquinot, *Nature*. **447**, 694–697 (2007).
8. D. L. Olson, T. L. Peck, A. G. Webb, R. L. Magin, J. V. Sweedler, *Science*. **270**, 1967–1970 (1995).
9. A. P. M. Kentgens *et al.*, *J. Chem. Phys.* **128**, 052202, 1–17 (2008).
10. T. M. Barbara, *J. Magn. Reson. A*. **109**, 265–269 (1994).
11. J.-P. Demers, V. Chevelkov, A. Lange, *Solid State Nucl. Magn. Reson.* **40**, 101–113 (2011).
12. Y. Ishii *et al.*, *J. Magn. Reson.* **286**, 99–109 (2018).
13. T. Kobayashi *et al.*, *Angew. Chem. Int. Ed.* **52**, 14108–14111 (2013).
14. A. Samoson, *J. Magn. Reson.* **306**, 167–172 (2019).
15. M. Deschamps, in *Annual Reports on NMR Spectroscopy*, G. A. Webb, Ed. (Academic Press, 2014), vol. 81, pp. 109–144.
16. Y. Nishiyama, *Solid State Nucl. Magn. Reson.* **78**, 24–36 (2016).
17. L. A. Currie, *Anal. Chem.* **40**, 586–593 (1968).
18. D. I. Hoult, R. E. Richards, *J. Magn. Reson. 1969*. **24**, 71–85 (1976).
19. J.-F. Jacquinot, D. Sakellariou, *Tech. Ing. Innov. En Anal. Mes.*, IN85, 1–9 (2008).
20. T. L. Peck, R. L. Magin, P. C. Lauterbur, *J. Magn. Reson. B*. **108**, 114–124 (1995).
21. E. Odeblad, *Acta Obstet. Gynecol. Scand.* **45**, 7–188 (1966).
22. J. Shoolery, in *Topics in Carbon-13 Nuclear Magnetic Resonance Spectroscopy*, G. C. Levy, Ed. (John Wiley & Sons Inc, New York u.a, 1979), vol. 3, pp. 28–38.
23. H. Mahdjour, W. G. Clark, K. Baberschke, *Rev. Sci. Instrum.* **57**, 1100–1106 (1986).

24. R. W. Wiseman, T. S. Moerland, M. J. Kushmerick, *NMR Biomed.* **6**, 153–156 (1993).
25. M. E. Lacey, R. Subramanian, D. L. Olson, A. G. Webb, J. V. Sweedler, *Chem. Rev.* **99**, 3133–3152 (1999).
26. N. Wu, T. L. Peck, A. G. Webb, R. L. Magin, J. V. Sweedler, *Anal. Chem.* **66**, 3849–3857 (1994).
27. R. Subramanian, A. G. Webb, *Anal. Chem.* **70**, 2454–2458 (1998).
28. Y. Li, A. M. Wolters, P. V. Malawey, J. V. Sweedler, A. G. Webb, *Anal. Chem.* **71**, 4815–4820 (1999).
29. M. Utz, R. Monazami, *J. Magn. Reson.* **198**, 132–136 (2009).
30. H. Ryan, S.-H. Song, A. Zaß, J. Korvink, M. Utz, *Anal. Chem.* **84**, 3696–3702 (2012).
31. H. Lee, E. Sun, D. Ham, R. Weissleder, *Nat. Med.* **14**, 869–874 (2008).
32. M. Mompeán *et al.*, *Nat. Commun.* **9**, 1–8 (2018).
33. Voxalytic GmbH, available at https://www.voxalytic.com.
34. Z. H. Cho *et al.*, *Med. Phys.* **15**, 815–824 (1988).
35. P. M. Glover, R. W. Bowtell, G. D. Brown, P. Mansfield, *Magn. Reson. Med.* **31**, 423–428 (1994).
36. M. P. McDougall, S. M. Wright, *Magn. Reson. Med.* **54**, 386–392 (2005).
37. C. Hilty *et al.*, *Proc. Natl. Acad. Sci.* **102**, 14960–14963 (2005).
38. E. W. Hagaman, J. Jiao, T. Moore, *J. Magn. Reson.* **193**, 150–152 (2008).
39. T. Jessat, M. Adjei-Acheamfour, M. Storek, R. Böhmer, *Solid State Nucl. Magn. Reson.* **82–83**, 16–21 (2017).
40. J. A. Tang *et al.*, *Chem. Phys. Lett.* **466**, 227–234 (2008).
41. A. G. Webb, *J. Magn. Reson.* **229**, 55–66 (2013).
42. B. Gruber, M. Froeling, T. Leiner, D. W. J. Klomp, *J. Magn. Reson. Imaging*. **48**, 590–604 (2018).
43. M. Jouda *et al.*, *Concepts Magn. Reson. Part B Magn. Reson. Eng.* **47B**, e21357, 1–10 (2017).
44. C. Massin *et al.*, *J. Magn. Reson.* **164**, 242–255 (2003).
45. Y. Maguire, I. L. Chuang, S. Zhang, N. Gershenfeld, *Proc. Natl. Acad. Sci.* **104**, 9198–9203 (2007).
46. L. F. Fuks, F. S. C. Huang, C. M. Carter, W. A. Edelstein, P. B. Roemer, *J. Magn. Reson. 1969*. **100**, 229–242 (1992).
47. F. D. Doty, G. Entzminger, Y. A. Yang, *Concepts Magn. Reson.* **10**, 133–156 (1998).
48. B. Behnia, A. G. Webb, *Anal. Chem.* **70**, 5326–5331 (1998).
49. D. Sakellariou, J.-F. Jacquinot, patent, WO2007003218, A3 (2007), available at https://worldwide.espacenet.com/publicationDetails/biblio?FT=D&CC=WO&NR=2007003218A1#.
50. J.-F. Jacquinot, D. Sakellariou, *Concepts Magn. Reson. Part A*. **38A**, 33–51 (2011).
51. P. M. Aguiar, J.-F. Jacquinot, D. Sakellariou, *J. Magn. Reson.* **200**, 6–14 (2009).
52. G. Aubert, J.-F. Jacquinot, D. Sakellariou, *J. Chem. Phys.* **137**, 154201, 1–14 (2012).
53. A. Wong, P. M. Aguiar, D. Sakellariou, *Magn. Reson. Med.* **63**, 269–274 (2010).
54. A. Wong *et al.*, *Anal. Chem.* **84**, 3843–3848 (2012).
55. A. Wong, X. Li, D. Sakellariou, *Anal. Chem.* **85**, 2021–2026 (2013).
56. A. Wong *et al.*, *Anal. Chem.* **86**, 6064–6070 (2014).
57. P. M. Aguiar, J.-F. Jacquinot, D. Sakellariou, *Chem. Commun.* **47**, 2119–2121 (2011).
58. A. Wong, P. M. Aguiar, T. Charpentier, D. Sakellariou, *Chem. Sci.* **2**, 815–818 (2011).
59. V. Badilita *et al.*, *PLoS ONE*. **7**, e42848, 1–8 (2012).
60. J. A. Lehmann-Horn, J.-F. Jacquinot, J. C. Ginefri, C. Bonhomme, D. Sakellariou, *J. Magn. Reson.* **271**, 46–51 (2016).
61. K. Yamauchi, J. W. G. Janssen, A. P. M. Kentgens, *J. Magn. Reson.* **167**, 87–96 (2004).

62. K. Yamauchi, T. Asakura, *Chem. Lett.* **35**, 426–427 (2006).
63. H. Janssen, A. Brinkmann, E. R. H. van Eck, P. J. M. van Bentum, A. P. M. Kentgens, *J. Am. Chem. Soc.* **128**, 8722–8723 (2006).
64. I. Frantsuzov *et al.*, *ChemPhysChem*. **18**, 394–405 (2017).
65. A. Brinkmann, S. K. Vasa, H. Janssen, A. P. M. Kentgens, *Chem. Phys. Lett.* **485**, 275–280 (2010).
66. D. L. A. G. Grimminck, S. K. Vasa, W. L. Meerts, A. P. M. Kentgens, A. Brinkmann, *Chem. Phys. Lett.* **509**, 186–191 (2011).
67. S. K. Vasa, E. R. H. van Eck, J. W. G. Janssen, A. P. M. Kentgens, *Phys. Chem. Chem. Phys.* **12**, 4813–4820 (2010).
68. M. Inukai, K. Takeda, *Concepts Magn. Reson. Part B Magn. Reson. Eng.* **33B**, 115–123 (2008).
69. M. Inukai, K. Takeda, *J. Magn. Reson.* **202**, 274–278 (2010).
70. M. Inukai, Y. Noda, K. Takeda, *J. Magn. Reson.* **213**, 192–195 (2011).
71. K. Takeda, *Solid State Nucl. Magn. Reson.* **47–48**, 1–9 (2012).
72. K. Kratt, V. Badilita, T. Burger, J. G. Korvink, U. Wallrabe, *J. Micromechanics Microengineering*. **20**, 015021, 1–11 (2010).
73. S. S. Adhikari, U. Wallrabe, V. Badilita, J. G. Korvink, *Concepts Magn. Reson. Part B-Magn. Reson. Eng.* **47B**, e21362, 1–13 (2017).
74. S. S. Adhikari, L. Zhao, T. Dickmeis, J. G. Korvink, V. Badilita, *Analyst*. **144**, 7192–7199 (2019).
75. Y. Nishiyama, Y. Endo, T. Nemoto, A.-K. Bouzier-Sore, A. Wong, *Analyst*. **140**, 8097–8100 (2015).
76. J.-C. Ginefri *et al.*, *J. Magn. Reson.* **224**, 61–70 (2012).
77. R. Frass-Kriegl *et al.*, *J. Magn. Reson.* **273**, 65–72 (2016).
78. S. Leidich, M. Braun, T. Gessner, T. Riemer, *Concepts Magn. Reson. Part B Magn. Reson. Eng.* **35B**, 11–22 (2009).
79. R. Weaver (2016), available at http://electronbunker.ca/eb/CalcMethods.html.
80. E. B. Rosa, F. W. Grover, *Bull. Bur. Stand.* **8**, 1–237 (1916).
81. H. Nagaoka, *J. Coll. Sci. Imp. Univ. Tokyo Jpn.* **27**, 1–33 (1909).
82. E. B. Rosa, *Bull. Bur. Stand.* **2**, 161–187 (1906).
83. H. J. Zabel, R. Bader, J. Gehrig, W. J. Lorenz, *Radiology*. **165**, 857–859 (1987).
84. P. Gonord, S. Kan, A. Leroy-Willig, C. Wary, *Rev. Sci. Instrum.* **65**, 3363–3366 (1994).
85. D. Boeck SAS, available at https://db-electronic.com.
86. R. Kriegl, PhD dissertation (Université Paris Sud - Paris XI, 2014), available at https://tel.archives-ouvertes.fr/tel-01127308.
87. S. Serfaty, N. Haziza, L. Darrasse, S. Kan, *Magn. Reson. Med.* **38**, 687–689 (1997).
88. R. Rosales (2014), available at https://www.ecured.cu/Efecto_pelicular.
89. H. A. Wheeler, *Proc. IRE*. **30**, 412–424 (1942).
90. Z. Popovic, B. D. Popovic, in *Introductory Electromagnetics* (Prentice Hall, Upper Saddle River, NJ, USA, 1st edition., 1999).
91. K. R. Minard, R. A. Wind, *Concepts Magn. Reson.* **13**, 190–210 (2001).
92. Vistapointe, available at https://vistapointe.net/images/proximity-effect-wallpaper-9.jpg.
93. A. W. Lotfi, P. M. Gradzki, F. C. Lee, *IEEE Trans. Magn.* **28**, 2169–2171 (1992).
94. R. G. Medhurst, *Wirel. Eng.*, 35–43 (1947).
95. S. Butterworth, *Exp. Wirel. Wirel. Eng.* **3**, 203–210 (1926).
96. D. W. Knight, (Ottery St Mary, Devon, England, 2016), available at http://g3ynh.info/zdocs/magnetics/SolenoidZ.pdf.
97. G. K. Felic, D. Ng, E. Skafidas, *IEEE Trans. Magn.* **49**, 1353–1360 (2013).

98. H. Lamb, J. W. L. Glaisher, *Philos. Trans. R. Soc. Lond.* **174**, 519–549 (1883).
99. R. Schäfer, C. Heiden, *Appl. Phys.* **9**, 121–125 (1976).
100. A. Bielecki, D. P. Burum, *J. Magn. Reson. A*. **116**, 215–220 (1995).
101. K. Ehrmann *et al.*, *Lab. Chip*. **7**, 373–380 (2007).
102. K. R. Minard, R. A. Wind, *Concepts Magn. Reson.* **13**, 128–142 (2001).
103. S. Idziak, U. Haeberlen, *J. Magn. Reson. 1969*. **50**, 281–288 (1982).
104. G. Laurent, C. Bonhomme (2020), available at http://doi.org/10.5281/zenodo.3598539.
105. E. B. Rosa, *Bull. Natl. Bur. Stand.* **4**, 301–344 (1908).
106. D. W. Knight (2014), available at http://www.g3ynh.info/zdocs/comps/part_2.html.
107. FastFieldSolvers (2019), available at https://www.fastfieldsolvers.com/fasthenry2.htm.
108. J.-L. Fanchon, *Guide des sciences et technologies industrielles: dessin industriel et graphes, matériaux, éléments de construction, économie et organisation d'entreprise, automatisme* (Afnor ; Nathan, Paris, France, 1994).
109. D. R. Robello, T. D. Eldridge, F. M. Michaels, *J. Polym. Sci. Part Polym. Chem.* **37**, 2219–2224 (1999).
110. H. Silve, available at http://herve.silve.pagesperso-orange.fr/brasage.htm.
111. Sonelec Musique (2015), available at http://www.sonelec-musique.com/electronique_bases_tutoriel_soudure.html.
112. Spilltech, available at https://www.spilltech.com/wcsstore/SpillTechUSCatalogAssetStore/Attachment/documents/ccg/POLYURETHANE.pdf.
113. G. Villemaud, (INSA Lyon, CITI Laboratory, 2013), available at http://perso.citi.insa-lyon.fr/gvillemaud/Docs/Propag/Abaque%20de%20Smith%20%28notion%204%29.ppt.
114. A. Chevalier, *Guide du dessinateur industriel : Pour maîtriser la communication technique* (Hachette, 1995), *Guides industriels*.
115. A. Fiore, available at https://www.eyeem.com/p/127737296.
116. M. Albéric *et al.*, *Cryst. Growth Des.* **18**, 2189–2201 (2018).
117. P. Tekely, M. Goldman, *J. Magn. Reson.* **148**, 135–141 (2001).
118. M. Weingarth, D. E. Demco, G. Bodenhausen, P. Tekely, *Chem. Phys. Lett.* **469**, 342–348 (2009).
119. M. Weingarth, P. Tekely, G. Bodenhausen, *Chem. Phys. Lett.* **466**, 247–251 (2008).
120. R. N. Purusottam, PhD dissertation (Sorbonne Université, 2015), available at https://tel.archives-ouvertes.fr/tel-01228506.
121. M. Weingarth, G. Bodenhausen, P. Tekely, *Chem. Phys. Lett.* **488**, 10–16 (2010).
122. B. Hu, O. Lafon, J. Trébosc, Q. Chen, J.-P. Amoureux, *J. Magn. Reson.* **212**, 320–329 (2011).
123. T. G. Oas, R. G. Griffin, M. H. Levitt, *J. Chem. Phys.* **89**, 692–695 (1988).
124. K. Sudarsanan, R. A. Young, *Acta Crystallogr. B*. **25**, 1534–1543 (1969).
125. C. Leroy, PhD dissertation (Paris 6, 2016), available at http://www.theses.fr/2016PA066256.
126. B. M. Fung, A. K. Khitrin, K. Ermolaev, *J. Magn. Reson.* **142**, 97–101 (2000).
127. Y. Sun *et al.*, *Sci. Rep.* **6**, 30756 (2016).
128. M. Wilhelm, H. Feng, U. Tracht, H. W. Spiess, *J. Magn. Reson.* **134**, 255–260 (1998).
129. N. Baccile, G. Laurent, C. Bonhomme, P. Innocenzi, F. Babonneau, *Chem. Mater.* **19**, 1343–1354 (2007).
130. O. Lafon *et al.*, *J. Phys. Chem. A*. **113**, 12864–12878 (2009).
131. T. Gullion, J. Schaefer, *J. Magn. Reson. 1969*. **81**, 196–200 (1989).
132. T. Gullion, *Concepts Magn. Reson.* **10**, 277–289 (1998).
133. T. Gullion, D. B. Baker, M. S. Conradi, *J. Magn. Reson. 1969*. **89**, 479–484 (1990).

134. A. W. Hing, S. Vega, J. Schaefer, *J. Magn. Reson. 1969*. **96**, 205–209 (1992).
135. G. A. Morris, R. Freeman, *J. Am. Chem. Soc.* **101**, 760–762 (1979).
136. T. Azaïs, G. Hartmeyer, S. Quignard, G. Laurent, F. Babonneau, *J. Phys. Chem. C*. **114**, 8884–8891 (2010).
137. C. Coelho, T. Azaïs, L. Bonhomme-Coury, G. Laurent, C. Bonhomme, *Inorg. Chem.* **46**, 1379–1387 (2007).
138. C. Coelho *et al.*, *Comptes Rendus Chim.* **11**, 387–397 (2008).
139. Z. Gan, *J. Magn. Reson.* **184**, 39–43 (2007).
140. B. Hu, J. Trébosc, J. P. Amoureux, *J. Magn. Reson.* **192**, 112–122 (2008).
141. M. H. Levitt, in *Encyclopedia of Nuclear Magnetic Resonance*, D. M. Grant, R. K. Harris, Eds. (John Wiley & Sons, Ltd, 2002), vol. 9, pp. 165–196.
142. J. P. Amoureux, J. Trébosc, G. Tricot, *Magn. Reson. Chem.* **45**, S187–S191 (2007).
143. S. Xin *et al.*, *Solid State Nucl. Magn. Reson.* **84**, 103–110 (2017).
144. C. A. Fyfe, K. T. Mueller, H. Grondey, K. C. Wong-Moon, *Chem. Phys. Lett.* **199**, 198–204 (1992).
145. S. C. Christiansen *et al.*, *Solid State Nucl. Magn. Reson.* **29**, 170–182 (2006).
146. A. Brinkmann, A. P. M. Kentgens, *J. Am. Chem. Soc.* **128**, 14758–14759 (2006).

Chapter III. Carr-Purcell-Meiboom-Gill echoes

'What is history? An echo of the past in the future; a reflex from the future on the past.'

Victor Hugo (1802-1885), writer

A. Chapter introduction

A first way to increase NMR sensitivity during acquisition is to use Carr-Purcell-Meiboom-Gill (CPMG) echoes (*1*, *2*). This technique focusses on one-dimensional experiments or on direct dimension of higher-dimensional experiments. This work started from the need to study electrospinning films (Section E.6 of Chapter I) by ^{29}Si solid-state NMR. ^{29}Si is 4.7 % abundant and need very long longitudinal relaxation time (hundreds of seconds). These hybrid membranes were used for efficient proton transport placed into batteries and contained a low molar ratio of silicon atom. Moreover, these samples were particularly soft and ductile, very similar to a Teflon ribbon. Even if enough sample was synthesised to fill a 7 mm rotor, spinning was almost impossible due to weight heterogeneity. However a filled 4 mm rotor spun easily, but with four times less sample weight inside it, which decreased sensitivity. As a consequence, Peak Signal-to-Noise Ratio based on maximum of noise ($PSNR_{max}$, Equation III.1)

$$PSNR_{max} = \frac{H_{signal}}{h_{noise\_peak\_peak}/2} \qquad \text{III.1}$$

was very low and no quantification was possible, even after a week-end of acquisition, thus the need to improve sensitivity with CPMG. In order to process spectra, I developed a Python software (*3*). In this chapter, we will first focus on theory of CPMG (Subchapter B) before detailing our simulations (Subchapter C) and applying this technique to materials (Subchapter D).

B. Theory

Historically, CPMG was first used to study solution-state dynamics (Section B.1) before being applied to solid-state NMR (Section B.2) and as a tool to increase sensitivity (Section B.3). It requires specific processing (Section B.4) and has an impact on quantification (Section B.5).

B.1. Solution-state dynamics

In 1950, Hahn observed spin echoes (SE) when a π pulse was applied after initial signal decrease (*4*). In 1954, Carr and Purcell suggested rather to use a train of π pulses, so-called CPSE, in order to minimize diffusion effect (*1*). However, cumulative radio-frequency

homogeneity errors were present and Meiboom and Gill proposed in 1958 to minimise them by shifting the first pulse phase by 90°, keeping all other pulses with the original phase, thus giving the CPMG pulse sequence (*2*). A variant was developed in 1966 by Mansfield, Waugh *et al.* (MW-4) with π/2 instead of π pulses (*5*, *6*). These experiments revealed useful to measure spin-spin relaxation time of molecules of interest.

In the continuity, Gutowsky *et al.* evidenced the impact of interpulses delay on chemical exchange (*7*), and Carver and Richards highlighted the influence of static magnetic field strength and temperature (*8*). However, a major drawback of CPMG for relaxation studies is the modulation of echoes intensity by scalar couplings (*9*). Tošner *et al.* shown that, with such a modulation, the measured true T_2 relaxation rates can vary by up to 50 % (*10*). As a solution, Aguilar *et al.* designed a pulse sequence, Periodic Refocusing of J Evolution by Coherence Transfer (PROJECT), to suppress scalar modulation, by alternating π and π/2 pulses (*11*). In addition, Loria *et al.* proposed a relaxation compensated CPMG method (*12*).

CPMG found many applications, such as millisecond timescale motions (*13*) and proteins exchange dynamics (*14*). Palmer *et al.* focussed on high energy states of proteins, with CPMG relaxation dispersion (*15*). Greener *et al.* analysed hydration of white cement paste (*16*). CPMG revealed also fruitful for well logging with shaped pulses (*17*). In Magnetic Resonance Imaging (MRI) fast spin echoes were obtained (*18*), Mulkern and Spencer observed diffusion imaging (*19*) and Constable compared MRI contrast factors (*20*).

B.2. Solid-state NMR experiments

CPMG was first applied to static solids by Garroway in 1977 (*21*). In order to maximise echoes intensity, Henrichs and Nicely synchronised the echo delay on MAS rate (*22*). Hung and Gan raised some important warnings about timings and digitisation to avoid distortions in CPMG (Figure III.1a) and Quadrupolar CPMG (QCPMG, Figure III.1b) experiments. On these figures, *M* and *N* correspond to integer numbers of echoes and rotor periods, respectively (*23*). QCPMG is equivalent to CPMG but with differently written timings, with modified pulse lengths to excite only central transition of quadrupolar nuclei (*24*) and with an additional phase cycling (*25*). The involved delays are defined according to Equation III.2.

$$\begin{gathered}\tau_1=(\tau_a-p_1)/2,\ \tau_a=2N\tau_r-2\tau_2-p_2 \\ \tau_1'=\tau_2'=\tau_a/2+\tau_3',\ \tau_3'=\tau_4'\end{gathered} \qquad \text{III.2}$$

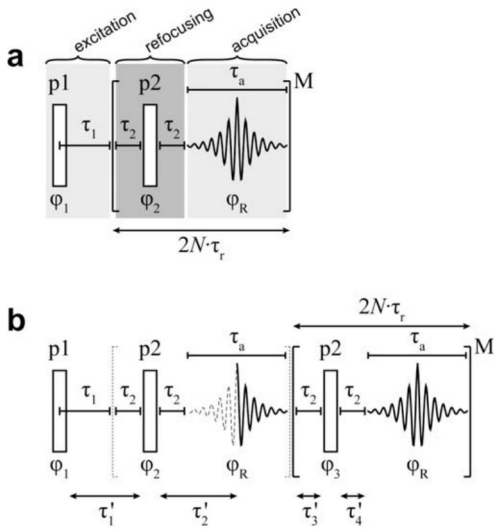


Figure III.1: a) CPMG and b) QCPMG pulse sequences (*23*).

Furthermore, in between spikes artefacts can appear, in particular from imperfect π pulses through stimulated echoes, *i.e.*, from zero-quantum coherence pathway (*26*). Pratum evidenced that 0-180° phase alternation of π pulses induced cumulative errors due to finite pulse length effects (*27*). XY-8 phase cycling was developed to improve magnetisation longevity from one echo to another (*28*). By replacing the first refocusing π by π/2, improved true T_2 were measured, favouring narrowed lines (*29*). A detailed pathway analysis for π and π/2 refocusing pulses highlighted the importance of odd and even echoes (*30*). Noticeable decrease of artefacts and odd/even pathway selection, was obtained with a pulse sequence named Phase Incremented Echo Train Acquisition (PIETA), that continuously increased phase shift (*31*). However, this required additional processing steps.

Garroway depicted that (Q)CPMG experiment are similar to a spin-lock, and that multiple echoes are present, with reduced homonuclear damping in time domain (FID), leading to a comb of narrowed lines in frequency domain (SPC), while keeping constant the global shape (*21*). Contrary to MAS spinning side bands, (Q)CPMG spikes are not at multiples of isotropic shifts but rather at multiples of window centre. That explains why (Q)CPMG cannot discriminate close isotropic chemical shifts (*32*).

According to Siegel *et al.* (*30*), the overall (Q)CPMG envelope is driven by inhomogeneous interactions while individuals spikelet shape reflects homogeneous interactions. These terms where defined by Maricq and Waugh (*33*), whether 'the eigenstates of the spin Hamiltonian […] are unaffected by the rotation' (*inhomogeneous interaction*), or are affected (*homogeneous interaction*). Hence, an inhomogeneous Hamiltonian commutates

with itself at a later time. The former category consists of chemical shift anisotropy, first order quadrupolar interaction, heteronuclear dipolar coupling, and scalar coupling, whereas the latter category contains homonuclear dipolar coupling and more complex combinations.

Concerning the dynamics aspect, Swanson *et al.* studied the mobility of polyvinyl acetate in function of humidity and temperature (*34*). Bank *et al.* applied CPMG technique to ^{113}Cd ions adsorbed on montmorillonite clay with temperature variation (*35*). ^{13}C bond lengths were determined on polycrystalline acetic acid, through Pake doublet splitting reduction by CPMG, with the help of cross-polarisation (CP) to enhance signal-to-noise ratio (*36*). ^{19}F dipolar splitting was studied on oriented systems (*37*), and adsorbed ^{87}Rb ions were distinguished according to their T_2 relaxation (*38*).

B.3. Sensitivity

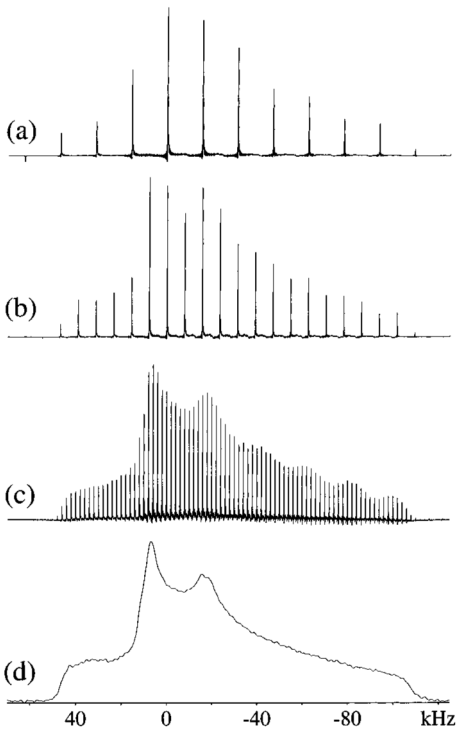


Figure III.2: Sensitivity improvement from d) Hahn Echo to a-c) QCPMG, with decreasing echoes frequency ν_{CPMG} = 15.6, 7.8, and 2 kHz, from top to bottom, corresponding to enhancement factors of 46, 21, and 8, respectively. ^{87}Rb solid-state NMR of a static sample of $RbVO_3$; B_0 = 9.4 T (*25*).

In contrast to previous approach for dynamics, (Q)CPMG can be used to improve sensitivity by discretising the peak area in frequency domain. This processing will be detailed in the following section. As peak area is proportional to the number of involved spins, the

intensity is concentrated into the remaining peaks, thus improving $PSNR_{max}$. This is illustrated on Figure III.2, where a $PSNR_{max}$ gain of 46 was obtained from the Hahn echo experiment to the QCPMG experiment with an echo frequency of $\nu_{CPMG} = 15.6$ kHz (*25*). However, a compromise has to be made between resolution and sensitivity. When increasing echoes rate, truncation in time domain can occur, degrading peak shape, as highlighted on Figure III.2a.

The first known usage of CPMG on quadrupolar nuclei was in 1987 to gain a factor 2.8 on signal of ^{17}O, when measuring ^{17}O-^{13}C distances in CO adsorbed on Pd metal (*39*). QCPMG was also useful for $^{35/37}$Cl NMR (*40*) and for integer quadrupolar nuclei, such as ^{2}H (*41*) and ^{14}N (*42*). Larsen *et al.* coupled QCPMG and MAS on ^{17}O spectra (*43*). Moreover, (Q)CPMG can be combined to other construction blocks in pulses sequences, to further enhance signal. Thereby, Multiple-Quantum MAS (MQMAS)-QCPMG allowed a fivefold enhancement on ^{23}Na (*44*). CP-QCPMG was applied on ^{25}Mg, or ^{67}Zn spectra (*45*). Schurko *et al.* added Double-Frequency Sweeps (DFS) and Rotor-Assisted Population Transfer (RAPT) to QCPMG for studying ^{39}K and $^{85/87}$Rb (*46*). Furthermore, continuous sensititvity improvement was obtained with adiabatic pulses such as Wideband Uniform Rate Smooth Truncation (WURST) (*47*) and BRoadband Adiabatic Inversion CP (Brain-CP) (*48*), resulting to an ultra wideline methodology (*49*). It should be noted that WURST pulses give raise to second order phase distortion (*23*). Finally, Bonhomme, Laurencin *et al.* added Variable Offset Cumulative Spectrum (VOCS) to maximise ^{87}Sr (*50*) and ^{127}I (*51*) with DFS VOCS WURST QCPMG.

Similarly, CPMG has been implemented many times on spin ½ nuclei to improve sensitivity, starting by Barrett *et al.* in 1990 to study ^{89}Y nucleus in a superconducting sample, with a gain of 2.5 (*52*). Siegel *et al.* focussed on ^{195}Pt, ^{199}Hg, and ^{207}Pb heavy nuclei with strong Chemical Shift Anisotropy (CSA) on static powders (*29*). Simultaneously, Hung *et al.* compared CP-CPMG and CP-MAS on ^{15}N, ^{109}Ag, ^{113}Cd, ^{199}Hg, and ^{207}Pb nuclei (*32*). WURST was added to CPMG to examine CSA of ^{119}Sn, ^{195}Pt, ^{199}Hg, and ^{207}Pb nuclei (*53*). ^{19}F→^{119}Sn BRAIN CP WURST CPMG spectra were obtained to study tin oxo-clusters (*51*). Solid 1,2,3-trimethoxybenzene was analysed by ^{13}C CP multi echoes PHase-cORrected Magic Angle Turning (PHORMAT) with very slow $MAS = 1.3$ Hz, in order to combine CSA patterns and sensitivity increase (*54*). CPMG and $MAS = 22$ kHz were combined to increase ^{77}Se sensitivity and to differentiate species based on their T_2 transversal relaxation (*55*).

Concerning ^{29}Si, in which we are especially interested, CPMG sensitivity enhancement methods were implemented as well. Ladd *et al.* highlighted that ^{29}Si CPMG echoes duration in a silicon single-crystal depended on ^{29}Si abundance and that in static condition with homonuclear decoupling, CPMG echoes can be measured during as long as one minute (*56*). CPMG sensitivity enhancement, again as a construction block in pulses programs, was combined with slow spinning ($MAS < 1$ kHz) techniques such as MAT (*57*), and STRAy-Field Imaging of rotating samples with Magic-Angle Spinning (STRAFI-MAS) (*58*). CPMG with higher spinning speeds reduced acquisition time by five on α-quartz (SiO_2), α-cristobalite (SiO_2), and Zircon ($ZrSiO_4$) (*43*). CPMG was also hyphenated with {^{27}Al}-^{29}Si HETeteronuclear CORrelation (HETCOR) and {^{1}H}-^{29}Si-^{29}Si refocused Incredible Natural Abundance DoublE QUAntum Transfer Experiment (INADEQUATE) (*59*), with Magic-Angle Flipping (MAF) (*60*) or with Dynamic Nuclear Polarisation (DNP) (*61*). In a remarkable paper, Malfait and Halter observed stretched exponential decays, depending on ^{29}Si chemical shift. They compared Regular echoes sum, weighted sum and T_2 reconstructed spectra (*62*).

B.4. Processing

B.4.a. Apodisation

CPMG echoes can be multiplied by weighting functions (Figure III.3), namely global apodisation (top), individual echo apodisation (middle top) or square function (middle bottom), either one or all of them. The first one avoids truncation effects on spikelets, the second one reduces noise introduced in between echoes, and the third one limits pulse ringing artefacts (*63*). To cite few examples, spikelet method with global apodisation was used in (*29*), global and individual apodisation in (*22*, *45*), double apodisation plus square multiplication in (*63*), and superposition method in (*43*, *52*, *64*). As this processing concerns only one-dimensional (1D) experiments, it has to be reproduced on each 1D slice of multi-dimensional spectra (Section B.1 of Chapter IV).

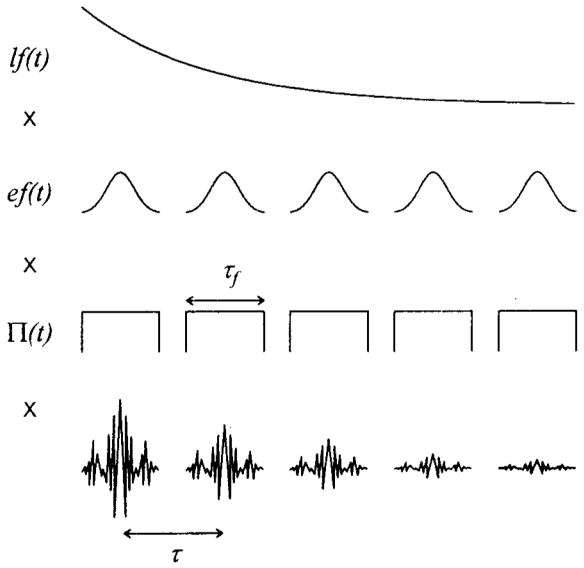


Figure III.3: Weighting functions applied to 1D CPMG FID. Top: global apodisation, middle top: individual echoes apodisation, middle bottom: pulse ringing removal, bottom: acquired echoes (*63*).

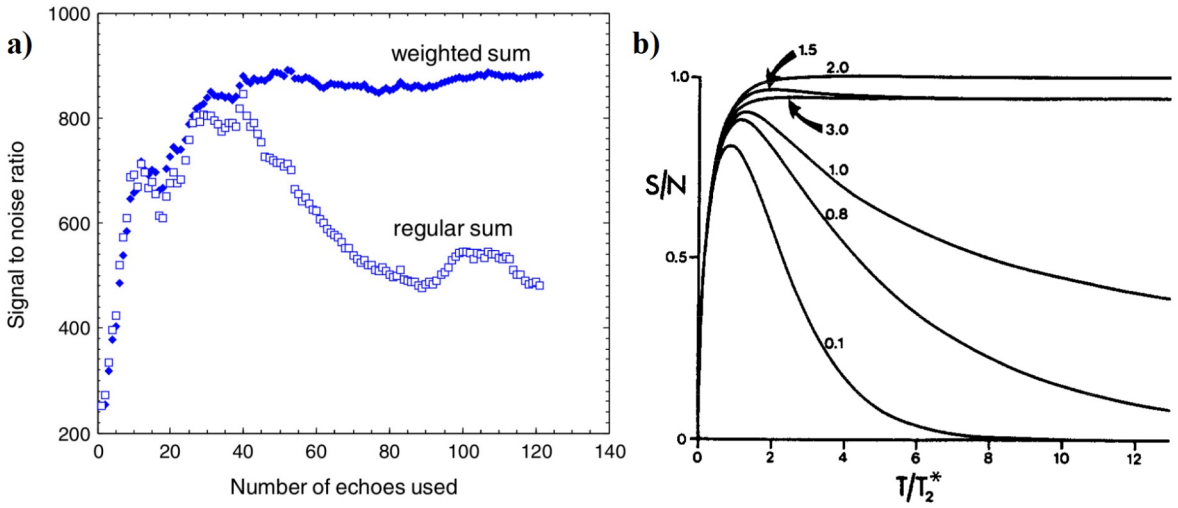


Figure III.4: a) Regular and weighted sum of CPMG echoes (*62*), b) Influence of multiplication of FID by an exponential whose constant is indicated (*65*).

Although $PSNR_{max}$ is increasing for first echoes, it decreases after a maximum (open squares on Figure III.4a) (*62*), as noise is proportional to the square root of the number of half-echoes (*64*). The situation is similar to a standard 1D acquisition of an exponentially decreasing signal with a time constant T_2^*, the apparent individual echo decrease, where $PSNR_{max}$ is maximised at 1.26 T_2^* (curve 1.0 on Figure III.4b) (*65*). In CPMG, echoes maxima decreases according to an averaged true T_2 value, the global CPMG echo train decrease. Apparent transversal relaxation T_2^* takes into account dispersion of chemical shifts, which are

refocused at each π, giving the true transversal relaxation T_2. By applying a matched filter on a standard FID, *i.e.*, an exponential apodisation (Section C.2 of Chapter I) with the same apparent constant T_2^*, signal decreases as $e^{-t/2T_2^*}$ and $PSNR_{max}$ raises up to a constant (curve 2.0 on Figure III.4b). An approximation of this matched filter on true T_2 can be obtained for CPMG by multiplying each echo by its own intensity value, so-called a weighted sum (filled diamonds on Figure III.4a) (*24*, *62*).

B.4.b. Spikelets and superposition methods

Two different approaches coexist to process (Q)CPMG spectra. The first one consists to directly FT the echoes, raising to a comb of narrow lines, so-called *spikelets method* (Figure III.5a) (*21*, *54*). The alternative *superposition method* involves a more complex processing (Figure III.5b), which folds the raising part of each echo into its decreasing part (b2), before summing all echoes together (b2') and performing FT (b3) (*63*, *64*). An acronym was even assigned to this technique and forgotten: Spin-Echo Fourier-Transform (SEFT) NMR (*66*). A less common procedure was to average points at top of echoes (*37*), which has the disadvantage of removing all chemical shift information, by centring peaks at middle of spectral window. CPMG Two-dimensional One Pulse (TOP) is also a fruitful processing possibility (*67*) and is available into Dmfit program (*68*).

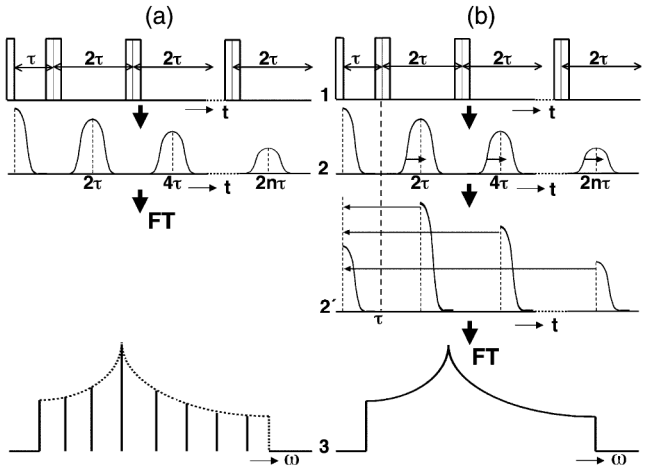


Figure III.5: Processing of 1D CPMG FID: a) spikelets and b) superposition methods. Each line corresponds to a processing step: 1) pulse sequence, 2) obtained echoes and individual half-echoes folding, 2') global folding into one half-echo, 3) final spectrum after Fourier transform (*54*).

B.4.c. Enhancement factor

A remaining question is the enhancement factor G that one can obtain with CPMG. A first answer was provided by Barrett *et al.*, who used the number of echoes N and the ratio r between amplitude of consecutive echoes (Equation III.3) (*52*).

$$G_{superposition} = \frac{1}{\sqrt{N}} \left[\frac{1 - r^N}{1 - r} \right] \qquad \text{III.3}$$

Lefort *et al.* proposed a more general formula based on relaxation rates, both apparent (T_{off}, the time needed for the signal to be confused with noise) and refocused (true T_2), and the echoes rate ν_{CPMG} (Equation III.4) (*63*).

$$\begin{aligned} G_{spikelets} &= 2\sqrt{T_2 T_{off}}\,\nu_{CPMG} \\ G_{spikelets}^{limit} &= 2\sqrt{\frac{T_2}{T_{off}}} \end{aligned} \qquad \text{III.4}$$

They also described the maximum sensitivity enhancement without truncation (G^{limit}). One should note that T_{off} has a rather vague definition, which depends on noise. A more precise definition could be $T_{off} = 5\,T_2^*$ (*59*). In such a case, 99 % of the transversal magnetisation would have disappeared. Hu and Wind compared the enhancement obtained from spikelets and superposition method (Equation III.5) (*54*).

$$\begin{aligned} G_{spikelets} &= \frac{1}{\sqrt{1+2N}} \left[1 + 2\sum_{n=1}^{N} \cos\left(2\omega n \tau_{echo}\right) \mathrm{e}^{-2n\frac{\tau_{echo}}{T_2}} \right] \\ G_{superposition} &= \frac{1}{\sqrt{1+2N}} \left[1 + 2\sum_{n=1}^{N} \mathrm{e}^{-2n\frac{\tau_{echo}}{T_2}} \right] \end{aligned} \qquad \text{III.5}$$

They proved that a cosine factor reduces spikelets intensities for all but the highest spikelet. τ_{echo} corresponds to the half-echo and the pre-factor accounts for noise increase with number of half-echoes. Trébosc *et al.* reformulated this expression using a change of variable (Equation III.6) (*69*). This more simple formula is directly usable to predict sensitivity gain from the number of echoes.

$$\begin{aligned} G_{superposition} &= \frac{1}{\sqrt{2N+1}} \left[1 + 2\sum_{n=1}^{N} x^n \right] \text{ with } x = \mathrm{e}^{-2\frac{\tau_{echo}}{T_2}} \\ G_{superposition} &= \frac{1}{\sqrt{2N+1}} \left[2x^n + \ldots + 2x^2 + 2x + 1 \right] \frac{x-1}{x-1} \\ G_{superposition} &= \frac{1}{\sqrt{2N+1}} \left[\frac{2x^{n+1} - x - 1}{x-1} \right] \end{aligned} \qquad \text{III.6}$$

B.5. Quantification

In addition to its processing complexity, the main drawback of CPMG is its need for long T_2 relaxation constants. Wiench *et al.* demonstrated that sensitivity on ^{29}Si species can be considerably improved by avoiding static condition (open triangles and bottom solid line on Figure III.6a) (*59*) and increasing MAS rate as much as possible. This is especially noticeable on T^3 species that have nearby ^{1}H nuclei helping to relax ^{29}Si (Figure III.6b). With 12 kHz low-power ^{1}H decoupling and $MAS = 40$ kHz, acquisition times of up to one second were obtained (open circles). As a consequence of the different true T_2 relaxation times between Q^4 and T^3 species, the areas measured on the spectrum were strongly varying (Figure III.6c). In such a case, it was especially difficult to obtain quantitative measurements. Despite these spectra were obtained with CP ($t_{cp} = 8$ ms), which is non-quantitative by nature and favours T species, the same discussion applies to evaluate the magnetisation just after CP, represented as $G = 1$ on curves. On this sample, T species were deconvoluted to 11 ± 2 % on quantitative spectra (*59*).

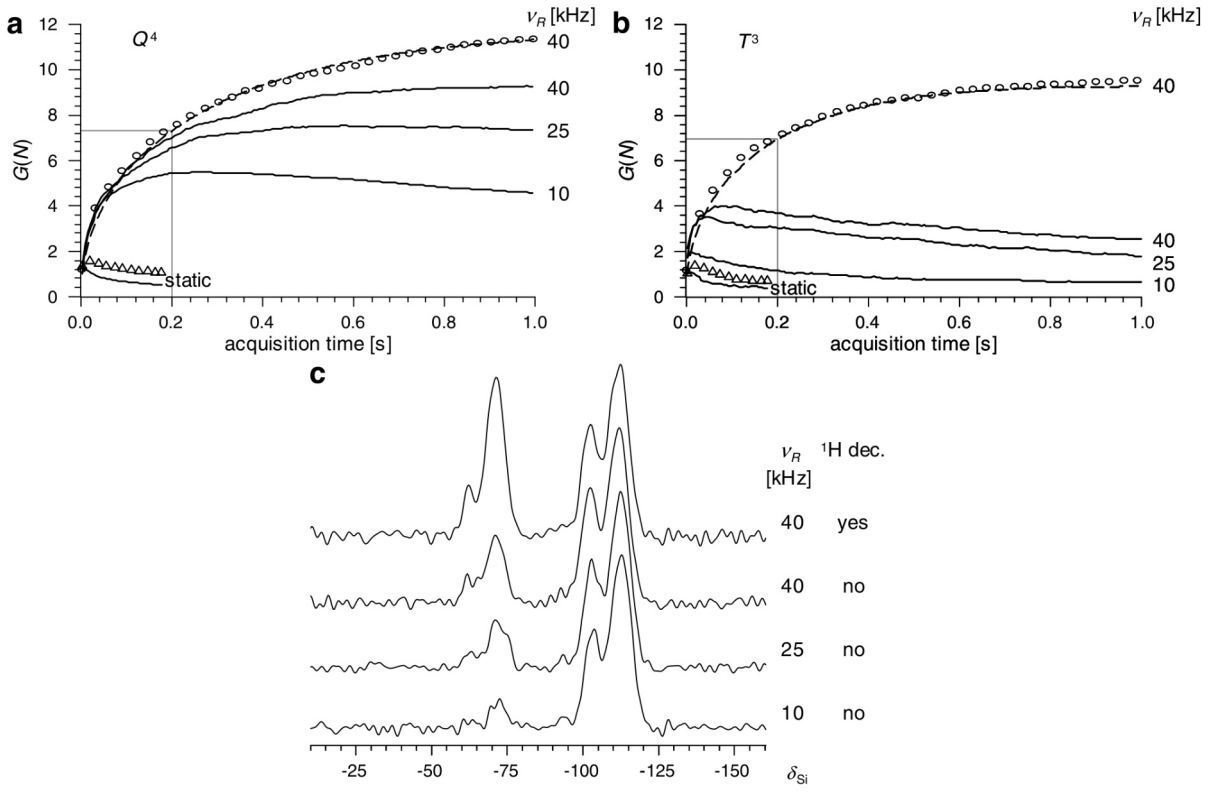


Figure III.6: Sensitivity gain *vs.* number of echoes for a) Q^4 sites and b) T^3 sites in mesoporous silica nanoparticle, obtained in static condition (bottom solid line) or with MAS (other solid lines). $\nu_{decoupling}(^1H) = 25$ kHz in static (open triangles) and 12 kHz at $MAS = 40$ kHz (open circles); dashed line: model fit. c) Reconstructed spectra with 33 echoes and various spinning speed and decoupling conditions. Adapted from (*59*).

In order to overcome this quantification limitation, Malfait and Halter measured a true T_2 relaxation with a stretched exponential, on the entire ^{29}Si CPMG train in frequency domain, and evaluated the intensities uncertainty with error propagation (*62*). Jaroszewicz *et al.* preferred to use Non-Negative Tikhonov Fitting (NNTF) routines for true T_2 separation in ^{35}Cl, ^{87}Rb, ^{71}Ga, and ^{119}Sn spectra (*70*). An innovative approach was proposed by Mason *et al.* combining CP contact time (T_{cp}), true T_2 from CPMG echoes and corresponding FID into a 3D data matrix (*71*). With tensor decomposition similar to Higher Order Singular Value Decomposition (HOSVD) (*72*), they could obtain a full various contact time curve for each specie with high $PSNR_{max}$ in 17 hours instead of 52.6 days. By fitting these curves, an estimation of quantification could be obtained, at least for nearby protonated species. Although these three techniques were interesting, they did not fully meet our need: a high initial $PSNR_{max}$ was necessary for the first approach, the second one was found too lately, and the third one needed a third dimension.

In Subchapter B, we highlighted that (Q)CPMG can be used to study dynamics, either in solution-state or in solid-state NMR. Precise timings are necessary to maximise signal and avoid artefacts. Additionally, (Q)CPMG provides a way to enhance sensitivity by discretising spectra. This is especially useful for highly anisotropic or low naturally abundant nuclei. These datasets required specific processing steps, whose enhancement factor depends. In case of multiple true T_2 relaxation constants, quantification of species is impacted.

C. Simulations

In this subchapter, our goal was to process (Q)CPMG datasets in a quantitative way, *i.e.*, that measured intensities reflect the magnetisation state after excitation, without being disturbed by true T_2 relaxation. In continuity of the program developed in Subsection B.3.d.ii of Chapter VI, we provided a Python program, interfaced with Bruker TopSpin software, and freely available on GitHub under GPL-3.0 licence (*3*). More than 2300 code lines were written, allowing to test installation, to measure $PSNR_{max}$, to denoise FID, and to simulate and process (Q)CPMG data.

Simulation of (Q)CPMG data will be presented in Section C.1. The (Q)CPMG processing workflow is shown on Figure III.7 and will be detailed in the forthcoming sections. Starting from the initial dataset (step a), echoes were aligned and individually apodised (step b, Section C.2). A global apodisation was then applied (step c, Section C.3). By

direct FT, the result of spikelets method was obtained (step d). On the contrary, summing (step e) before FT (step f) lead to superposition method (Section C.4). Our approach was to apply a partial sum before SVD denoising (step g), truncation of FID (step h) and FT (step i, Section C.5).

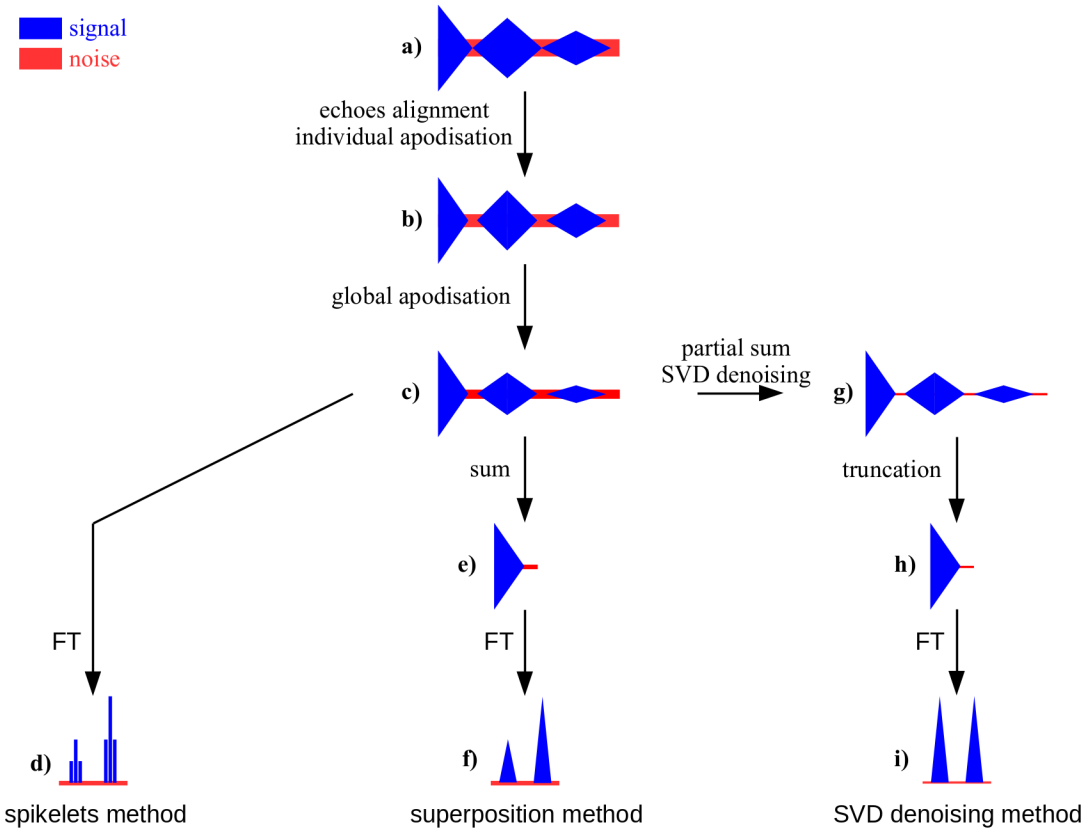


Figure III.7: (Q)CPMG workflow: a) initial dataset, b) after echoes alignment and individual apodisation, c) after global apodisation, d) using spikelets method, e) after sum of echoes, f) using superposition method, g) after partial sum and SVD denoising, h) after truncation at the end of the first echo, i) using SVD denoising method.

C.1. Generation of a train of echoes

From our point of view, the best way to ensure a viable processing algorithm is to start from a perfectly known dataset. That is why we designed a program simulating a (Q)CPMG echoes train in time domain (FID), cpmg_gen.py file of (*3*). Each parameter of Table III.1 can be adjusted. FID imperfections can be willingly introduced and checked, such as mismatch between sampling rate and echo delay, or missing points. Moreover, a noise-free dataset allows to check apodisation or folding effects with pinpoint accuracy.

Table III.1: Parameters in cpmg_gen.py program.

Parameter	Description	Parameter	Description
fullEcho	Full echo delay (s)	*nbEcho*	Number of echoes
firstDec	Presence of first decreasing half-echo (True / False)	*td*	Time domain points (real and imaginary points alternated)
dw	Dwell time between two successive points (s)	*de*	Dead time before acquisition (s)
mean	Mean value of noise	*std*	Standard deviation of noise
ph_0	Global zero order phase (deg)	amp_1	Amplitude of first peak
nu_1	Frequency of first peak (Hz)	T_{2_1}	True T_2 relaxation of first peak with Lorentzian shape (s)
$T_{2\_1}^*$	Apparent T_2^* relaxation of first peak with gl_1 shape (s)	gl_1	Gaussian / Lorentzian ratio for apparent relaxation ($0 \leq gl1 \leq 1$)

Such a noise-free FID is presented on Figure III.8a, with two frequencies at 2000 and -3000 Hz. The two signals had same amplitude and apparent relaxation time $T_2^* = 1$ ms, but different true relaxation times $T_2 = 30$ and 100 ms, respectively. When truncating signal after the first decrease and ignoring all (Q)CPMG echoes, FT gave the orange *reference* spectrum (SPC) on Figure III.8c, with identical height, width, shape and area for both peaks, mainly depending on the apparent T_2^*. Peaks proportionality changed in presence of (Q)CPMG echoes, where spikelets width depended on the true T_2, giving narrower spikelets for the peak at -3000 Hz. As peak area was constant, this increased peak height much more than for the peak at 2000 Hz (blue spectrum on Figure III.8c). As a consequence, a dissymmetry appears between both peaks clump, and the left one was hardly visible in presence of noise, whereas the right one was enhanced.

When acquiring an experimental FID, probe ringing occurs, which means that residual excitation radio-frequency corrupts the observed signal. As a consequence, the first points are dismissed, what is called a dead time. It was taken into account in our simulations (red arrow on Figure III.8b) and induced a first order dephasing on spectrum (in blue on Figure III.8c). Homoscedastic white Gaussian noise was subsequently added on FID. A noise standard deviation value of $std = 0.3$ corresponded to peak-peak noise ranging from -1 to 1. Final synthesised FID corresponded to step a of Figure III.7.

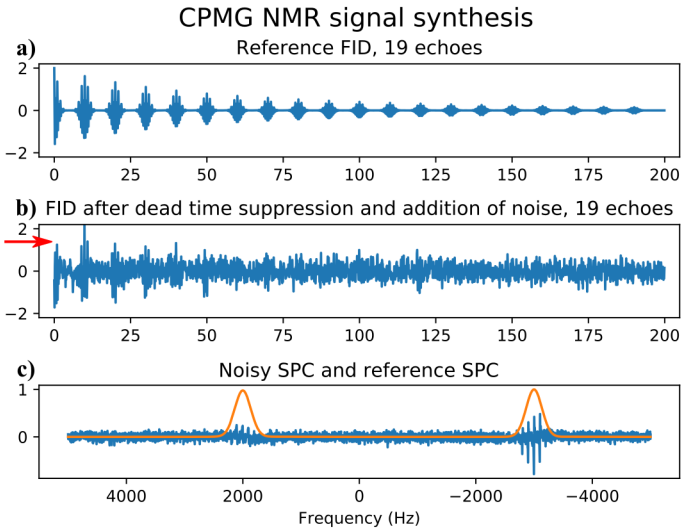


Figure III.8: (Q)CPMG signal synthesis: a) noise-free FID, b) noisy FID with dead time suppression (red arrow), c) noisy SPC (blue) and reference SPC (orange). Parameters of cpmg_gen.py program: *fullEcho* = 10 ms, *nbEcho* = 19, *firstDec* = True, *td* = 4000, *dw* = 50 µs, *de* = 100 µs, *mean* = 0, *std* = 0.3, ph_0 = 0, amp_1 = amp_2 = 1, nu_1 = 2000 Hz, nu_2 = -3000 Hz, T_{2_1} = 30 ms and T_{2_2} = 100 ms with Lorentzian shape, $T_{2\_1}^*$ = $T_{2\_2}^*$ = 1 ms with Gaussian shape.

C.2. Echoes alignment

After importing data with nmrglue (*73*), the first (Q)CPMG processing step in cpmg_prog.py program of (*3*) was to align echoes, *i.e.*, to ensure that top of echo was always at the middle of the echo (dotted vertical black line on Figure III.9a). To the best of our knowledge, this procedure was not detailed in the literature, but was used by authors without describing it. On simulated data, echo delay and dead-time were already exactly known. However, on experimental data, there could be some slight variations between the chosen values and the effective ones, for instance due to hidden electronic delays. Hence, we provide another program, cpmg_cal.py, to calibrate shift parameters without doing full processing. Similar curves were obtained with both programs.

If a dead time was present, FID had to be right shifted, to compensate for non-existent points. On the contrary, if acquisition started too early due to a misconfiguration of delays in pulse program, FID had to be left shifted. FID number of points was kept constant and missing points were replaced by zeroes. Correcting the echo position had the additional advantage to compensate first order phase. Maximum intensity of each echo was measured on its entire range and plotted against acquisition time (Figure III.9b). Individual apodisation (see next section) was applied before plotting to facilitate echo top visual detection. In case of

mismatch between sampling rate and echo delay, falls were observed (not shown), which were useful to carefully adjust echo delay. When the effective echo delay was not a multiple of 2 *dw* (one real and one imaginary point), partial points were in excess on each echo. The solution was to approximate echo top by removing two points at the beginning of some echoes when, and only when, two full points were in excess. This way, the echo top was oscillating around the middle of echoes. Data length and number of echoes were adjusted consequently. In addition, a true T_2 relaxation measurement was performed to verify signal decrease, and to update simulation or acquisition parameters if needed. The measured value was an average of the individual signals within noise error.

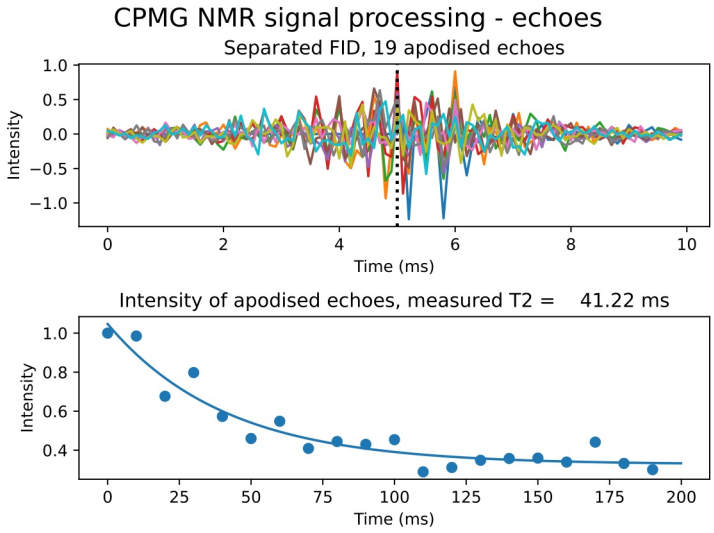


Figure III.9: Echoes alignment: a) superposition of echoes, b) measurement of true T_2 relaxation. Dotted vertical black line: centre of echoes. cpmg_proc.py program applied on noisy FID synthesised in Figure III.8. Parameters of cpmg_gen.py program: *fullEcho* = 10 ms, *nbEcho* = 19, *firstDec* = True, *td* = 4000, *dw* = 50 µs, *de* = 100 µs, *mean* = 0, *std* = 0.3, $ph_0 = 0$, $amp_1 = amp_2 = 1$, $nu_1 = 2000$ Hz, $nu_2 = -3000$ Hz, $T_{2\_1} = 30$ ms and $T_{2\_2} = 100$ ms with Lorentzian shape, $T_{2\_1}^* = T_{2\_2}^* = 1$ ms with Gaussian shape.

C.3. Apodisation

Two subsequent apodisation steps were applied: an individual echo one and a global one. This is similar to previous studies (*63*). Without adjusting top of echoes, it was difficult to distinguish them from noise (Figure III.10a). However, once top of echoes were carefully calibrated, it was possible to apply apodisation on each of them (Figure III.10b, corresponding to step b of Figure III.7). Two kind of echoes apodisation were tested: cosine attaining zero between two echoes, or exponential. The difference was that apodisation was more convex on

top of echo with cosine than with exponential function (Section C.2 of Chapter I). During this individual apodisation step, any point acquired after the end of last echo was discarded.

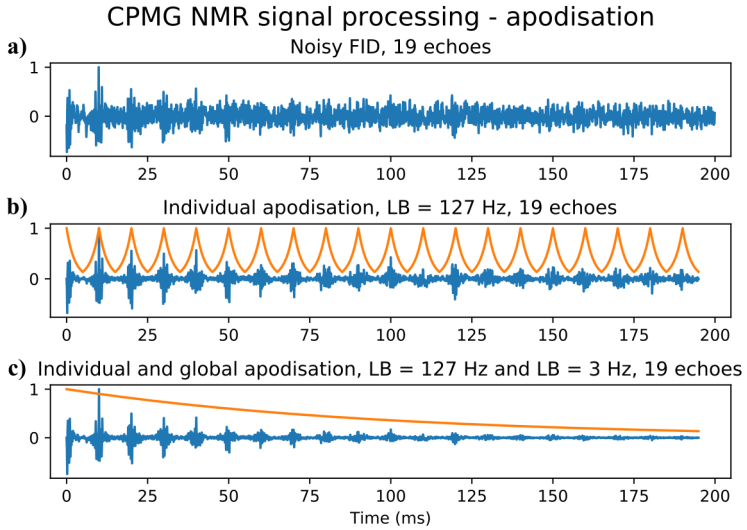


Figure III.10: Apodisation step: a) noisy FID, b) after individual apodisation, c) after global apodisation. Apodisation functions are presented in orange. cpmg_proc.py program applied on noisy FID synthesised in Figure III.8. Parameters of cpmg_gen.py program: *fullEcho* = 10 ms, *nbEcho* = 19, *firstDec* = True, *td* = 4000, *dw* = 50 µs, *de* = 100 µs, *mean* = 0, *std* = 0.3, ph_0 = 0, amp_1 = amp_2 = 1, nu_1 = 2000 Hz, nu_2 = -3000 Hz, T_{2_1} = 30 ms and T_{2_2} = 100 ms with Lorentzian shape, $T_{2\_1}^*$ = $T_{2\_2}^*$ = 1 ms with Gaussian shape.

Global apodisation was then applied, again either as cosine or as exponential (Figures III.10c, corresponding to step c of Figure III.7). Best results were obtained with exponential weightings with constants equal to 2 / (π × *halfEcho*) and to 2 / (π × *acquisition*) for individual echoes and global apodisation, respectively. *halfEcho* and *acquisition* stand for the half-echo delay and the total FID acquisition time, respectively. Further increase of exponential constant led to additional artefacts during SVD denoising (Section C.5). At this point, a common ancestor was reached between spikelets, superposition and SVD denoising methods, which will be explained and compared hereafter.

C.4. Spikelets and superposition methods

Spikelets SPC was obtained by direct FT of the apodised FID (Figure III.11a, corresponding to step d of Figure III.7) (*54*). Gratefully to the apodisation, $PSNR_{max}$ of this spectrum was considerably enhanced as compared to Figure III.8c. Moreover, first-order phase was corrected by echoes alignment. However, small peaks at the edge of clumps were still just above noise level. Noticeably, height of peaks at 2000 Hz were lower than the ones at

-3000 Hz, despite an identical simulated amplitude, highlighting non-quantitative results. This was explained by the different true T_2 relaxation.

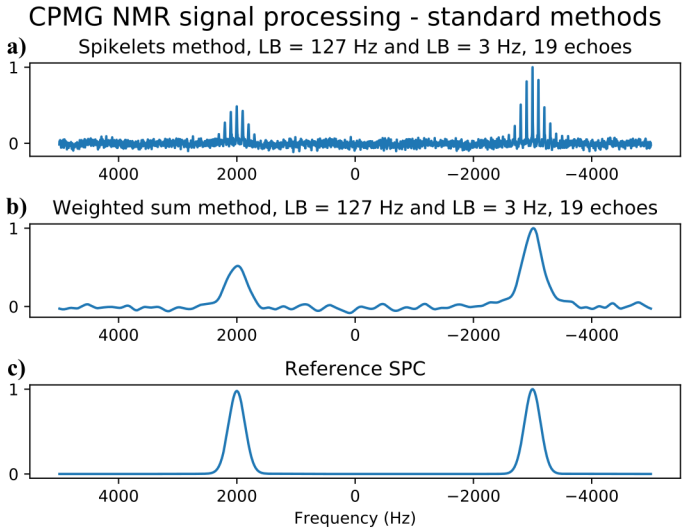


Figure III.11: Comparison of a) spikelets method, b) superposition method, c) reference spectrum. Apodisation is indicated on spectra. cpmg_proc.py program applied on noisy FID synthesised in Figure III.8. Parameters of cpmg_gen.py program: *fullEcho* = 10 ms, *nbEcho* = 19, *firstDec* = True, *td* = 4000, *dw* = 50 µs, *de* = 100 µs, *mean* = 0, *std* = 0.3, $ph_0 = 0$, $amp_1 = amp_2 = 1$, $nu_1 = 2000$ Hz, $nu_2 = -3000$ Hz, $T_{2\_1} = 30$ ms and $T_{2\_2} = 100$ ms with Lorentzian shape, $T_{2\_1}^* = T_{2\_2}^* = 1$ ms with Gaussian shape.

In order to obtain a reconstructed spectrum with the superposition method, apodised echoes were summed (step e of Figure III.7) (*54*). Before summing all echoes together, raising and decreasing parts of each echo were added. The subtlety was that a different treatment had to be applied on real and imaginary part of the data. Hence, real part was symmetric to top of echo (even), whereas imaginary part was asymmetric (odd). In other words, complex conjugate of raising part was folded to decreasing part (*43*). Unfortunately, this was only valid if the FID was zero-order phased. For an ordinary experimental FID, real and imaginary parts had no purely cosine and sine shapes. To correct this, the zero-order phase was calculated according to Equation III.7, where top_{real} and top_{imag} are the FID values at first full echo top.

$$phase_0 = \frac{-360}{2\pi} \arctan\left(\frac{top_{imag}}{top_{real}}\right) \qquad \text{III.7}$$
$$phase_0 = phase_0 - 180 \text{ if } top_{real} < 0$$

Moreover, there was only one point at top of echo, so this point was added to himself, but the first point of the raising part was discarded, in order to manipulate the same number of points from the raising and the decreasing parts. The obtained half echoes were then merged

into a single one, similar to a single FID without CPMG, but with enhanced $PSNR_{max}$. After FT, the result of superposition method was obtained (Figure III.11b, corresponding to step f of Figure III.7). Peak shape was similar to the overall envelope obtained with spikelets method. However, noise was very different, with high and low frequency for the spikelets and superposition methods, respectively. In agreement with Equation III.5, only a small $PSNR_{max}$ enhancement was obtained from spikelets to superposition methods. Once again, when comparing to reference SPC (Figure III.11c), superposition spectra exhibited a decreased intensity on peak at 2000 Hz, due to its faster T_2 relaxation rate. This clearly confirmed the default of quantification of these methods.

C.5. SVD denoising method

In order to combine (Q)CPMG enhancement and quantification, we proposed to use Singular Value Decomposition (SVD, Chapter V). Different approaches were tested. Firstly, the echo train after apodisation was converted to a 2D matrix, with a row per echo and SVD was applied on it. This was similar to a 2D version of (*71*), but results were not convincing. Two explanations were advanced: either matrix elongation due to the difference between number of points per echo and number of echoes (Subsection B.4.b of Chapter V); or an additional selectivity of SVD on 3D matrix.

Secondly, SVD was applied on the full echo train after individual and global apodisation (step c of Figure III.7). This was interesting as spikelets are Lorentzian peaks, for which quantitative denoising is expected, on the contrary to Gaussian peaks (Subsection B.4.f.iii of Chapter V). We are referring here to the individual spikelet quantification rather than to the overall envelope one. Computing SVD on the entire echo train had the drawback to be rather long (up to 600 s with the help of a Graphics Processing Unit (GPU), Subsection B.3.d.ii of Chapter VI). Furthermore, it was unnecessary to use the full echo train as only a few echoes were needed to obtain denoised spikelets, whereas the following ones were useless, except to avoid FID truncation. Once the denoised echo train was computed, its FT gave a denoised spikelets SPC (Figure III.12a). Only spikelets above automatic threshold, *i.e.*, above $PSNR_{max} \sim 2$ (Subsection B.4.e.ii of Chapter V) were detected. This was especially annoying for small peaks, which define the overall spikelets shape. For instance, extreme peaks around 2000 Hz disappeared. This spectrum was the low limit of $PSNR_{max}$ under which overall spikelets shape was degraded. For lower $PSNR_{max}$, the missing side spikelets narrowed reconstructed peak and decreased quantification. In extreme cases, no peak was reconstructed.

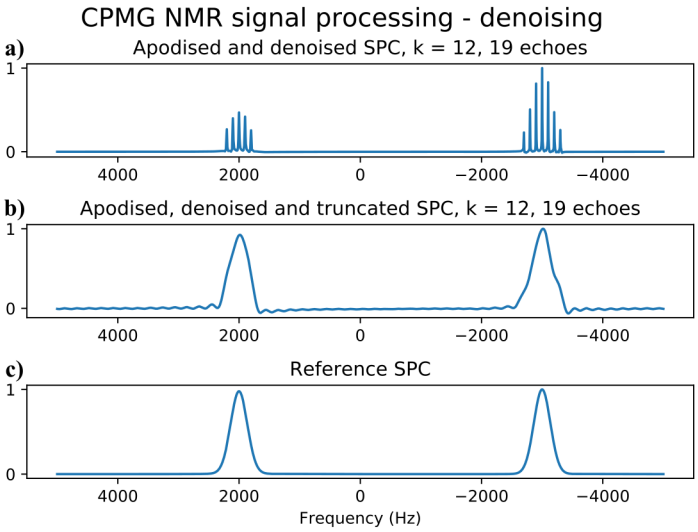


Figure III.12: Comparison of denoising method a) before truncation, and b) after truncation, c) reference spectrum. cpmg_proc.py program applied on noisy FID synthesised in Figure III.8. Parameters of cpmg_gen.py program: *fullEcho* = 10 ms, *nbEcho* = 19, *firstDec* = True, *td* = 4000, *dw* = 50 µs, *de* = 100 µs, *mean* = 0, *std* = 0.3, ph_0 = 0, $amp_1 = amp_2 = 1$, nu_1 = 2000 Hz, nu_2 = -3000 Hz, T_{2_1} = 30 ms and T_{2_2} = 100 ms with Lorentzian shape, $T_{2\_1}^* = T_{2\_2}^*$ = 1 ms with Gaussian shape.

As signal intensity is decreasing faster for species with short true T_2, quantification error is increasing with echo number. Only the first signal decrease is error-free. Hence, by truncation of the denoised FID after the first decrease (step h of Figure III.7), true T_2 relaxation effects were ignored, giving a more quantitative SPC after FT (Figure III.12b, corresponding to step i of Figure III.7). For comparison, reference SPC is presented on Figure III.12c. Quantitative deconvolution was not yet implemented in our Python program, but dmfit deconvolutions (*68*) will be tested on experimental spectra in next subchapter. Qualitatively, peaks at 2000 and -3000 Hz had the same amplitude, and a similar width, slightly enlarged as compared to reference SPC, due to apodisation. This was a major improvement over spikelets and superposition methods, for which different amplitudes were observed between both peaks. However, peaks were slightly distorted, more triangular than Gaussian.

SVD sensitivity strongly depended on FID $PSNR_{max}$. The question was thus how to further increase $PSNR_{max}$ before SVD. Our solution was to use a partial sum, by limiting echoes number and summing groups of echoes together, including first echo (step g of Figure III.7). Unfortunately, this reintroduced true T_2 discrepancies and lowered quantification. A maximum number of 25 echoes was chosen as a compromise to benefit from noise averaging

without too much impacting quantification. Nevertheless, a careful error measurement has still to be conducted.

In Subchapter C, a (Q)CPMG processing program was presented. After simulation of a perfectly known dataset, echoes alignment was performed. Global and individual apodisation enhanced signal-to-noise ratio. Finally, spikelets, superposition and SVD denoising methods were detailed. The two former suffered from true T_2 distortions, whereas the latter corrected them.

D. Application to materials

In order to further test our SVD denoising method on real (Q)CPMG experiments, we analysed two different samples by ^{29}Si solid-state NMR. Figure III.13 presents the spectra obtained with spikelets, superposition (sum), denoised, and no CPMG, from bottom to top, respectively. The later method was obtained by truncation of the apodised FID after the first decrease and was equivalent to the spectrum recorded without CPMG. These four spectra were processed starting from a unique CPMG dataset, which was necessary to precisely compare $PSNR_{max}$ enhancement with processing steps as close as possible from a method to another one.

The first sample was a sol-gel organic-inorganic material with MTEOS ($Si(OEt)_3CH_3$) and TEOS ($Si(OEt)_4$) in a 50:50 molar ratio. Synthesis was detailed in Section E.5 of Chapter I. It was studied by ^{29}Si Cross Polarisation (CP) CPMG MAS with 23 echoes (Figure III.13a). Overall shapes obtained by spikelets and superposition methods were very similar (blue and green spectra, respectively). By comparison with the spectrum obtained in absence of CPMG (grey spectrum), they clearly underestimated the region at -55 ppm (T^2 species, red arrows), while overestimating peaks at -100 and -110 ppm (Q^3 and Q^4 species, respectively), which was confirmed by modelling with dmfit software (*68*) (Table III.2). This was due to faster true T_2 relaxation for T^2 species, and slower one for Q^3 and Q^4 species. On the contrary, denoised spectrum (purple curve) was in much closer agreement with the spectrum in absence of CPMG. Artefacts were present in between spikelets, and were attributed to a mismatch between echo delay and MAS rate. $PSNR_{max}$ measured on Q^2 species with snr.py program of (*3*) increased from 37 without CPMG to 44 for spikelets, 62 for superposition and 190 for denoised methods. As expected from Equation III.5, superposition method gave better results than spikelets and no CPMG methods. With superposition method,

a time gain of $(62/37)^2 = 2.8$ was possible. Denoised method combined a good $PSNR_{max}$ enhancement with correct peak area quantification.

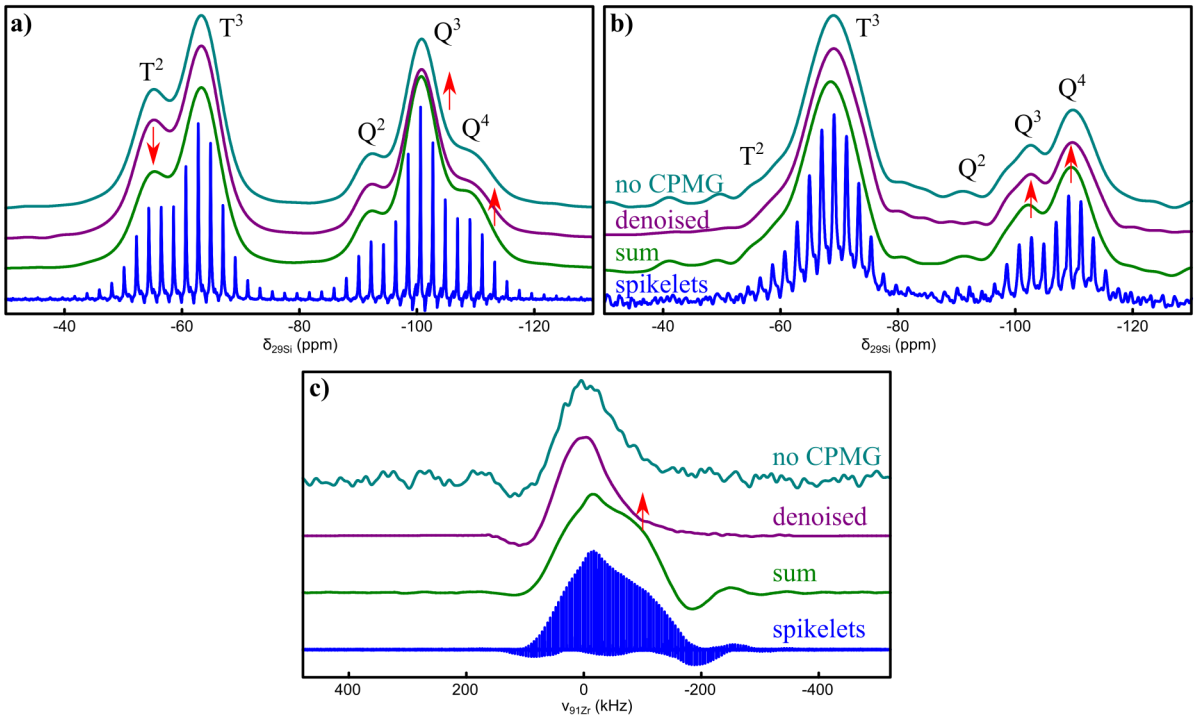


Figure III.13: Comparison of (Q)CPMG processing methods. From top to bottom on each subfigure: no CPMG, SVD denoising method, superposition, and spikelets methods. a) {^{1}H}-^{29}Si CP CPMG spectrum of 50:50 MTEOS:TEOS sample; b) {^{1}H}-^{29}Si HPDEC CPMG spectrum of CSPTC:TEOS:PVDF-HFP membrane; c) ^{91}Zr DFS-WURST-QCPMG spectrum of a zirconia ceramic. Parameters: a) B_0 = 7.0 T, *MAS* = 14 kHz, *RD* = 1 s, *NS* = 2048, t_{exp} = 0.57 h, t_{cp} = 5 ms, $\nu_{decoupling}(^1H)$ = 2.4 kHz SPINAL-64, *fullEcho* = 8 ms, *nbEcho* = 23, *firstDec* = True; b) B_0 = 7.0 T, *MAS* = 14 kHz, p_1 = 30°, *RD* = 30 s, *NS* = 3960, t_{exp} = 33 h, $\nu_{decoupling}(^1H)$ = 51 kHz SPINAL-64, *fullEcho* = 8 ms, *nbEcho* = 4, *firstDec* = True; c) B_0 = 16.4 T, *MAS* = 0 kHz, *RD* = 0.4 s, *NS* = 10000, t_{exp} = 1.1 h, no decoupling, *fullEcho* = 0.217 ms, *nbEcho* = 540, *firstDec* =False.

Similar results were obtained on a second sample consisting of an electrospinning film containing an organosilane (CSPTC), an orthosilicate (TEOS) and a fluorinated polymer (PVDF-HFP) (Figure III.13b). Synthesis was detailed in Section E.6 of Chapter I. As the aim was to quantify ^{29}Si species, High Power DECoupling (HPDEC) was preferred to CP, which is non-quantitative by nature. This spectrum was acquired in 33 h with *MAS* = 14 kHz. $PSNR_{max}$ was measured on Q^3 species (Table III.2). T^2 species were between detection and quantification level. Q^2 species were under detection limit. Only a slight time gain of $(6.5/4.7)^2 = 1.9$ was obtained with CPMG superposition method. Q^3 and Q^4 species increased, whereas T^3 species decreased. However, the T_2 discrepancies effect was limited as solely 4 echoes were acquired due to fast true T_2 relaxation, even with ^{1}H decoupling. Denoised

spectrum improved quantification, with values intermediate between superposition and no CPMG methods.

Table III.2: Deconvolution of ^{29}Si spectra presented on Figure III.13a-b.

		Species					
	Method	**T^2 (%)**	**T^3 (%)**	**Q^2 (%)**	**Q^3 (%)**	**Q^4 (%)**	***PSNR$_{max}$***
MTEOS TEOS	No CPMG	19.8	35.1	7.1	26.5	11.4	37
	Denoised	19.8	35.2	7.1	26.6	11.3	190
	Superposition	15.6	32.5	7.6	29.4	14.9	62
	Spikelets	-	-	-	-	-	44
CSPTC TEOS PVFD-HFP	No CPMG	~2.7	65.4	-	10.2	21.6	4.7
	Denoised	~3.0	64.2	-	11.3	21.5	18
	Superposition	~4.6	61.4	-	11.8	22.0	6.5
	Spikelets	-	-	-	-	-	5.7

On these two samples, the number of echoes was under 25 and no partial sum was applied before SVD denoising. Further tests were done with up to 540 echoes, for instance on ^{91}Zr QCPMG of zirconia ceramic (Figure III.13c and Section E.9 of Chapter I). A sensitivity gain of 12.6, corresponding to a time gain of 159, was observed between no CPMG and superposition methods. It is not clear yet if the different shape observed with denoising method accounts for relaxation of different species or for zero-order phase correction before folding of raising and decreasing parts of echoes. This folding could be avoided during (partial-)sum process, leading to a full echo rather than a half echo for superposition and denoising methods. An additional sensitivity gain of $\sqrt{2}$ could be obtained, but first-order phase correction will be needed.

In Subchapter D, our (Q)CPMG processing program was tested on experimental datasets. With a single command, no CPMG, spikelets, superposition and denoising methods were accessible under Bruker Topspin software. This is a major gain of processing time. The main difficulty was to carefully adjust echoes position for the best results. On ^{29}Si spectra, true T_2 were corrected, giving correct quantification in presence of (Q)CPMG.

E. Chapter conclusion

In this chapter, we focussed on (Q)CPMG experiments that can either be used to probe dynamics or to increase sensitivity of 1D solid-state NMR experiments or direct dimension of

nD experiments (Section B). For the latter objective, specific processing steps were underlined, either using spikelets or superposition methods. Different equations from the literature were presented to estimate the enhancement factor. One major drawback of (Q)CPMG is to loose quantification in presence of multiple species with different relaxation constants. Malfait and Halter proposed a relaxation compensated method, but to the detriment of signal enhancement (*62*).

In order to circumvent relaxation downside, we developed a processing program written in Python (Section C) (*3*). After synthesising a perfectly known (Q)CPMG train of echoes, we applied echoes alignment. This evidenced the need to have a perfect synchronisation between inter points delay, half-echo delay, and MAS period. Following step was to apodise FID, both individually on each echo and globally. The spikelets and superposition results were then obtained by direct FT or folding followed by sum and FT, respectively. Denoising approach was rather to apply SVD on a limited number of echoes. This way, SVD was applied on multiple narrow Lorentzian lines, enhancing peak detection (Subsection B.4.e.ii of Chapter V) and avoiding the extracted error obtained for a single Gaussian shape (Subsection B.4.f.iii of Chapter V). Additional echoes increased computational cost without further signal improvement. Truncating denoised FID after the first decrease and applying FT resulted to a denoised quantitative spectrum. However, for correct reconstructed peak shape, small spikelets should have a $PSNR_{max} \sim 2$ after apodisation. To improve their detection, we applied a partial sum before SVD in case of a high number of echoes. Unfortunately, this decreased quantification in case of multiple true T_2 relaxation constants.

Results on experimental spectra were provided in Section D on two different samples studied by (Q)CPMG. A time gain of 2.8 was achieved on ^{29}Si with 23 echoes, and of 86 on ^{91}Zr with 600 echoes. Despite spikelets and superposition methods gave non-quantitative results, quantification was improved with SVD denoising method. One advantage our program is that it is fast to process and directly callable from TopSpin software. However, a more precise error measurement would be useful, and some adjustment of number of echoes and of echoes folding process would improve overall shape reconstruction. Alternatively, true T_2 resolved 2D will provide an insight to characterise multiple species (*55*, *59*). Some investigations on PIETA technique (*31*), which is very close from (Q)CPMG but less sensitive to pathway errors, may probably be fruitful.

F. Chapter bibliography

1. H. Y. Carr, E. M. Purcell, *Phys. Rev.* **94**, 630–638 (1954).
2. S. Meiboom, D. Gill, *Rev. Sci. Instrum.* **29**, 688–691 (1958).
3. G. Laurent, P.-A. Gilles (2017), available at https://github.com/gul916/NMR_post_proc.
4. E. L. Hahn, *Phys. Rev.* **80**, 580–594 (1950).
5. P. Mansfield, D. Ware, *Phys. Lett.* **22**, 133–135 (1966).
6. E. D. Ostroff, J. S. Waugh, *Phys. Rev. Lett.* **16**, 1097–1098 (1966).
7. H. S. Gutowsky, R. L. Vold, E. J. Wells, *J. Chem. Phys.* **43**, 4107–4125 (1965).
8. J. P. Carver, R. E. Richards, *J. Magn. Reson. 1969*. **6**, 89–105 (1972).
9. A. Allerhand, *J. Chem. Phys.* **44**, 1–9 (1966).
10. Z. Tošner, A. Škoch, J. Kowalewski, *ChemPhysChem*. **11**, 638–645 (2010).
11. J. A. Aguilar, M. Nilsson, G. Bodenhausen, G. A. Morris, *Chem. Commun.* **48**, 811–813 (2012).
12. J. P. Loria, M. Rance, A. G. Palmer, *J. Am. Chem. Soc.* **121**, 2331–2332 (1999).
13. R. Ishima, P. T. Wingfield, S. J. Stahl, J. D. Kaufman, D. A. Torchia, *J. Am. Chem. Soc.* **120**, 10534–10542 (1998).
14. A. L. Hansen, P. Lundström, A. Velyvis, L. E. Kay, *J. Am. Chem. Soc.* **134**, 3178–3189 (2012).
15. A. G. Palmer, C. D. Kroenke, J. Patrick Loria, in *Methods in Enzymology*, T. L. James, V. Dötsch, U. Schmitz, Eds. (Academic Press, 2001), vol. 339 of *Nuclear Magnetic Resonance of Biological Macromolecules - Part B*, pp. 204–238.
16. J. Greener *et al.*, *J. Am. Ceram. Soc.* **83**, 623–627 (2000).
17. L. B. Casabianca, D. Mohr, S. Mandal, Y.-Q. Song, L. Frydman, *J. Magn. Reson.* **242**, 197–202 (2014).
18. J. Hennig, A. Nauerth, H. Friedburg, *Magn. Reson. Med.* **3**, 823–833 (1986).
19. R. V. Mulkern, R. G. S. Spencer, *Magn. Reson. Imaging*. **6**, 623–631 (1988).
20. R. T. Constable, A. W. Anderson, J. Zhong, J. C. Gore, *Magn. Reson. Imaging*. **10**, 497–511 (1992).
21. A. N. Garroway, *J. Magn. Reson. 1969*. **28**, 365–371 (1977).
22. P. M. Henrichs, V. A. Nicely, *Macromolecules*. **24**, 2506–2513 (1991).
23. I. Hung, Z. Gan, *J. Magn. Reson.* **204**, 256–265 (2010).
24. F. A. Perras, J. Viger-Gravel, K. M. N. Burgess, D. L. Bryce, *Solid State Nucl. Magn. Reson.* **51–52**, 1–15 (2013).
25. F. H. Larsen, H. J. Jakobsen, P. D. Ellis, N. Chr. Nielsen, *J. Phys. Chem. A*. **101**, 8597–8606 (1997).
26. B. A. Cowans, J. B. Grutzner, *J. Magn. Reson. A*. **105**, 10–18 (1993).
27. T. K. Pratum, *J. Magn. Reson. 1969*. **78**, 123–126 (1988).
28. T. Gullion, D. B. Baker, M. S. Conradi, *J. Magn. Reson. 1969*. **89**, 479–484 (1990).
29. R. Siegel, T. T. Nakashima, R. E. Wasylishen, *J. Phys. Chem. B*. **108**, 2218–2226 (2004).
30. R. Siegel, T. T. Nakashima, R. E. Wasylishen, *Concepts Magn. Reson. Part A*. **26A**, 62–77 (2005).
31. J. H. Baltisberger *et al.*, *J. Chem. Phys.* **136**, 211104, 1–4 (2012).
32. I. Hung, A. J. Rossini, R. W. Schurko, *J. Phys. Chem. A*. **108**, 7112–7120 (2004).
33. M. M. Maricq, J. S. Waugh, *J. Chem. Phys.* **70**, 3300–3316 (1979).
34. S. Swanson, S. Ganapathy, S. Kennedy, P. M. Henrichs, R. G. Bryant, *J. Magn. Reson. 1969*. **69**, 531–534 (1986).
35. S. Bank, J. F. Bank, P. D. Ellis, *J. Phys. Chem.* **93**, 4847–4855 (1989).

36. M. Engelsberg, C. S. Yannoni, *J. Magn. Reson. 1969*. **88**, 393–400 (1990).
37. S. L. Grage, A. S. Ulrich, *J. Magn. Reson.* **138**, 98–106 (1999).
38. J. T. Cheng, P. D. Ellis, *J. Phys. Chem.* **93**, 2549–2555 (1989).
39. S. E. Shore, J.-P. Ansermet, C. P. Slichter, J. H. Sinfelt, *Phys. Rev. Lett.* **58**, 953–956 (1987).
40. D. L. Bryce, M. Gee, R. E. Wasylishen, *J. Phys. Chem. A*. **105**, 10413–10421 (2001).
41. F. H. Larsen, H. J. Jakobsen, P. D. Ellis, N. Chr. Nielsen, *Chem. Phys. Lett.* **292**, 467–473 (1998).
42. L. A. O'Dell, R. W. Schurko, *J. Am. Chem. Soc.* **131**, 6658–6659 (2009).
43. F. H. Larsen, I. Farnan, *Chem. Phys. Lett.* **357**, 403–408 (2002).
44. T. Vosegaard, F. H. Larsen, H. J. Jakobsen, P. D. Ellis, N. C. Nielsen, *J. Am. Chem. Soc.* **119**, 9055–9056 (1997).
45. A. S. Lipton, J. A. Sears, P. D. Ellis, *J. Magn. Reson.* **151**, 48–59 (2001).
46. R. W. Schurko, I. Hung, C. M. Widdifield, *Chem. Phys. Lett.* **379**, 1–10 (2003).
47. L. A. O'Dell, R. W. Schurko, *Chem. Phys. Lett.* **464**, 97–102 (2008).
48. K. J. Harris, A. Lupulescu, B. E. G. Lucier, L. Frydman, R. W. Schurko, *J. Magn. Reson.* **224**, 38–47 (2012).
49. R. W. Schurko, *Acc. Chem. Res.* **46**, 1985–1995 (2013).
50. C. Bonhomme *et al.*, *J. Am. Chem. Soc.* **134**, 12611–12628 (2012).
51. D. Laurencin *et al.*, *ChemistrySelect*. **1**, 4509–4519 (2016).
52. S. E. Barrett *et al.*, *Phys. Rev. B*. **41**, 6283–6296 (1990).
53. A. W. MacGregor, L. A. O'Dell, R. W. Schurko, *J. Magn. Reson.* **208**, 103–113 (2011).
54. J. Z. Hu, R. A. Wind, *J. Magn. Reson.* **163**, 149–162 (2003).
55. M. Deschamps *et al.*, *Solid State Nucl. Magn. Reson.* **40**, 72–77 (2011).
56. T. D. Ladd, D. Maryenko, Y. Yamamoto, E. Abe, K. M. Itoh, *Phys. Rev. B*. **71**, 014401, 1–12 (2005).
57. D. Sakellariou, J.-F. Jacquinot, T. Charpentier, *Chem. Phys. Lett.* **411**, 171–174 (2005).
58. A. Wong, D. Sakellariou, *ChemPhysChem*. **12**, 3529–3532 (2011).
59. J. W. Wiench, V. S.-Y. Lin, M. Pruski, *J. Magn. Reson.* **193**, 233–242 (2008).
60. J. H. Baltisberger *et al.*, *J. Magn. Reson.* **268**, 95–106 (2016).
61. A. J. Rossini *et al.*, *Chem. Sci.* **3**, 108–115 (2012).
62. W. J. Malfait, W. E. Halter, *J. Non-Cryst. Solids*. **354**, 4107–4114 (2008).
63. R. Lefort, J. W. Wiench, M. Pruski, J.-P. Amoureux, *J. Chem. Phys.* **116**, 2493–2501 (2002).
64. F. H. Larsen, J. Skibsted, H. J. Jakobsen, N. Chr. Nielsen, *J. Am. Chem. Soc.* **122**, 7080–7086 (2000).
65. J. C. Lindon, A. G. Ferrige, *Prog. Nucl. Magn. Reson. Spectrosc.* **14**, 27–66 (1980).
66. A. Allerhand, D. W. Cochran, *J. Am. Chem. Soc.* **92**, 4482–4484 (1970).
67. K. K. Dey, J. T. Ash, N. M. Trease, P. J. Grandinetti, *J. Chem. Phys.* **133**, 054501, 1–10 (2010).
68. D. Massiot *et al.*, *Magn. Reson. Chem.* **40**, 70–76 (2002).
69. J. Trébosc, J. W. Wiench, S. Huh, V. S.-Y. Lin, M. Pruski, *J. Am. Chem. Soc.* **127**, 7587–7593 (2005).
70. M. J. Jaroszewicz, L. Frydman, R. W. Schurko, *J. Phys. Chem. A*. **121**, 51–65 (2017).
71. H. E. Mason, E. C. Uribe, J. A. Shusterman, *Phys. Chem. Chem. Phys.* **20**, 18082–18088 (2018).
72. T. Kolda, B. Bader, *SIAM Rev.* **51**, 455–500 (2009).
73. J. J. Helmus, C. P. Jaroniec, *J. Biomol. NMR*. **55**, 355–367 (2013).

Chapter IV. Non-uniform sampling

'Everything should be made as simple as possible, but not simpler.'

Albert Einstein (1879-1955), physicist

A. Chapter introduction

After increasing sensitivity of one-dimensional (1D) spectra through CPMG echoes, we focussed on multi-dimensional (nD) experiments with Non-Uniform Sampling (NUS), which allows considerable acquisition time decrease. NMR acquisition has been evolving for over six decades, with an increasing speed of nD signal acquisition (Figure IV.1) (*1*). A brief history of NMR can be found in (*2*). Some of these innovations are listed hereafter: Fourier Transform (FT) in 1966 (*3*), two-dimensional (2D) NMR in 1971 (*4*), structure of a protein with 57 amino-acids in 1985 (*5*), first 4D experiment in 1992 (*6*), and single-scan 2D NMR in 2002 (*7*). A new comprehension of signals sparsity was provided by Compressed Sensing (CS) in 2011 (*8*).

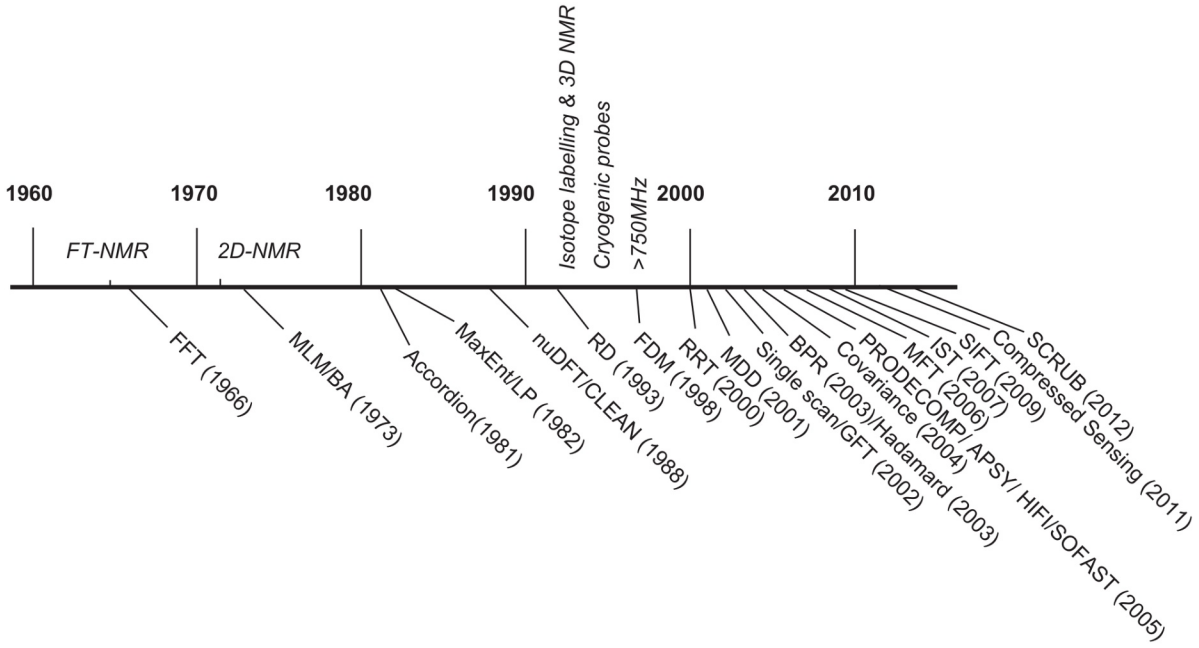


Figure IV.1: Chronology of NMR developments (*1*).

A few techniques are known to enhance acquisition speed of nD experiments. Among them, single-scan NMR, also called UltraFast (UF), consists to selectively excite a spatially encoded sample and to acquire successive corresponding echoes (*7*). Band-Selective Optimized Flip-Angle Short-Transient (SOFAST) (*9*), later renamed as Band-selective Excitation Short-Transient (BEST), which uses a band-selective excitation to enhance relaxation through the non-excited neighbours. Recently, NMR by Ordered Acquisition using ^{1}H-detection (NOAH) combined up to 5 successive 2D experiments into a single pulse

program (*10*). Non-Uniform Sampling (NUS) allowed to acquire only a part of the usual indirect points (*1*). SOFAST has already been combined with Non-Uniform Sampling (NUS) (*11*), whereas UF experiments are not compatible, due to the single-scan coding process. Most of these techniques are gaining interest for solution-state NMR, but here we focussed solely on NUS for solid-state NMR.

This work started from the availability of NUS in Bruker TopSpin software and strongly benefited from an international school in Sweden (*12*). A global historical perspective on NUS can be found in reference (*13*) and will be detailed hereafter for each NUS approach. We will first explain multi-dimensional experiments and their limitations (Subchapter B), before optimising data acquisition and reconstruction (Subchapters C.1 and D, respectively). These procedures will be applied to solid-state NMR in Subchapter E. Finally, practical details will be provided in Subchapter F.

B. Multi-dimensional experiments

In this Subchapter, after describing two- and higher dimensional experiments, we will focus on resolution and sensitivity before examining the impact of FID truncation, whose principle is similar for all dimensions.

B.1. Two-dimensional experiments

2D spectra are based on an idea of Jeener (*4*) and were developed in the 1970s (*14*, *15*). As shown on Figure IV.2a, the 2D principle is to acquire a succession of 1D FID, so-called *slices*, with a step by step delay increase along t_1 (*indirect* dimension). After a mixing step, the t_2 dimension is digitalised in one-shot (*direct* dimension). Moreover, frequency sign detection requires *phase quadrature*, which double the number of FID. To obtain a 2D spectrum, two complex FT are necessary, one for each time evolution.

B.2. Higher-dimensional experiments

The same principle can be extended to higher dimensions, that have nothing in common with our 3D visual space (Figure IV.2b). A 3D will thus have two indirect dimensions and be a succession of 2D planes. Figure IV.2c presents a 4D hCOCANH experiment (*16*), useful to study proteins by NMR. Amide ^{1}H magnetisation is first transferred to carbonyl ^{13}C, then to alpha ^{13}C, to amide ^{15}N and finally to amide ^{1}H. In a single

experiment, one can correlate two successive amino acids of the protein and obtain spectra of all their connecting nuclei.

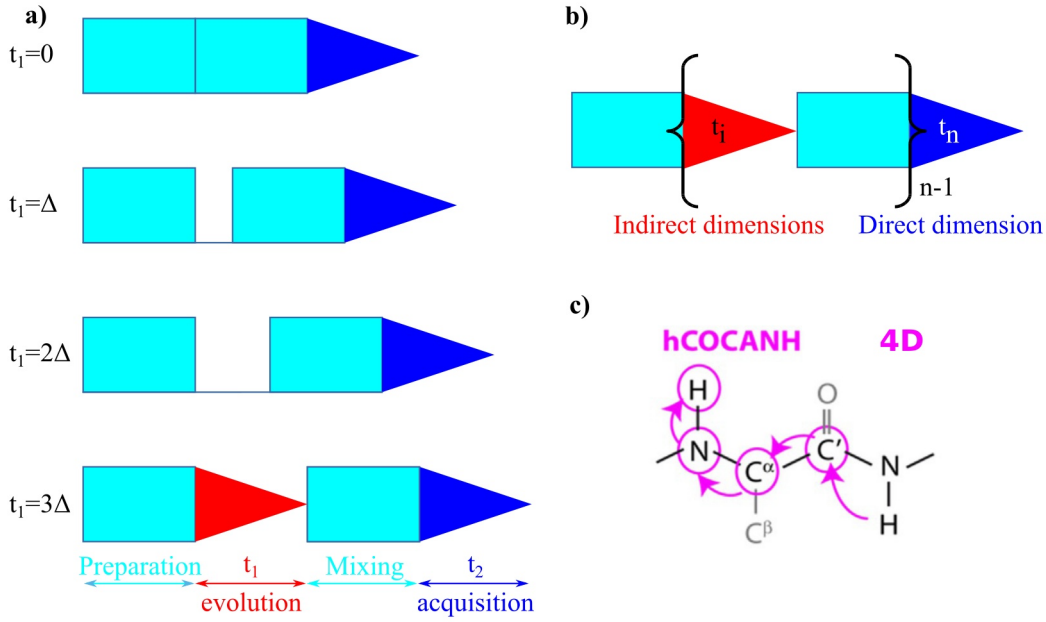


Figure IV.2: Principle of a) 2D and b) nD experiments, adapted from (*17*). Indirect and direct dimensions are point by point and one-shot acquisition, respectively.
c) 4D experiment hCOCANH (*16*).

By increasing dimensions, experimental time T_{exp} is growing according to Equation IV.1, where N is the number of points per n indirect dimension, and T_{1D} is 1D acquisition time (*18*).

$$T_{\exp} = N^{n-1} 2^{n-1} T_{1D} \qquad \text{IV.1}$$

The factor 2^{n-1} reflects phase quadrature for each indirect dimension. Corresponding T_{exp} are presented by dashed lines on Figure IV.3 from 2D to 5D experiments. Plain lines refers to sparse sampling, which will be explained in Subchapter C.1. Table IV.1 details acquisition parameters for top right point of each dashed line. For these calculations, T_{1D} is fixed to only 4 seconds, which reflects a very favourable case where sufficient Peak Signal-to-Noise ratio (*PSNR*) is achieved in only 2 scans, each one acquired in 2 seconds, with magnetisation selection by pulsed field gradients instead of phase cycling (*18*). A higher number of scans will consequently increase T_{exp}. Experimental time has been limited to 500 hours, *i.e.*, 3 weeks, which is the maximum reasonable for an NMR experiment. In such conditions, the NMR spectrometer has to be perfectly stable concerning magnetic field homogeneity and radio-frequency pulses imperfections. As a consequence, resolution has to decrease from high

(2D) to low (5D), and is much poorer than achievable on modern NMR spectrometers, which are providing incomparable narrow lines, especially on high field magnets. Another drawback of nD experiments is disk space, that increases with number of FID.

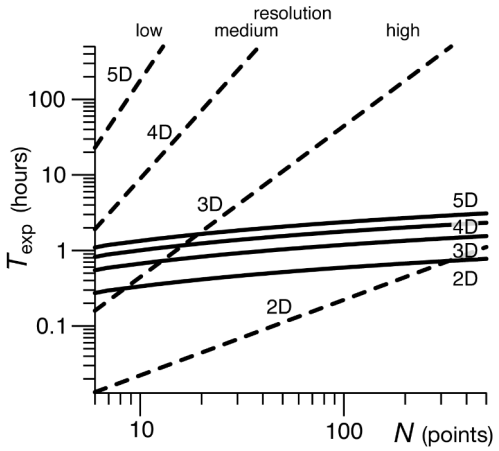


Figure IV.3: Comparison of full sampling (dashed lines) and sparse sampling (plain lines) acquisition times for nD NMR experiments, including direct dimension (*18*).

Table IV.1: Comparison of 1D to 5D experiments, including direct dimension. Dims: dimensions; phase quad: phase quadrature; sim: simultaneous.

Dims	**Direct points**	**Indirect increments**				**Phase quad**	**Total FID**	**Disk space**	**Reso-lution**	**Time**
		#1	**#2**	**#3**	**#4**					
1D	512	-	-	-	-	2 (sim)	2 (sim)	4.0 ko	high	4 seconds
2D	512	512	-	-	-	4	2.0e4	4.2 Mo	high	1.1 hour
3D	512	334	334	-	-	8	8.9e5	1.8 Go	high	20.7 days
4D	512	38	38	38	-	16	8.8e5	1.8 Go	medium	20.3 days
5D	512	13	13	13	13	32	9.1e5	1.9 Go	low	21.1 days

In the rest of this chapter, direct dimension will be neglected due to its one-shot character, unless otherwise stated. We will investigate how to decrease the acquisition time of indirect dimensions. Next section will describe the impact of truncation on resolution and sensitivity.

B.3. Resolution and sensitivity

RESolution of a spectrum (*RES*) can be defined as peak separation capability. It is driven by the corresponding FID duration (*AQ*, AcQuisition time), which itself is determined by Spectral Width (*SW*) and number of points (*TD*, Time Domain), as shown by Equation IV.2. Longer the FID, smaller and better the resolution.

$$RES(Hz) = \frac{1}{AQ(s)} = \frac{SW(Hz)}{TD/2} \qquad \text{IV.2}$$

As demonstrated by Rovnyak *et al.* (*19*), discrimination of two resonances separated by 1.2 times their line width requires to acquire signal up to 3 T_2^*, the apparent transversal relaxation time (green circle in Figure IV.4a), whatever its longitudinal relaxation time T_1. On the contrary, *PSNR* and sensitivity, are maximised at 1.26 T_2^*, value above which more noise than signal is introduced into the FID (Equation IV.3 and green arrow in Figure IV.4b).

$$PSNR(t_{max}) \propto \frac{T_2^*\left(1 - e^{\frac{-t_{max}}{T_2^*}}\right)}{\sqrt{t_{max}}},\ t_{max} \sim 1.26\,T_2^* \qquad \text{IV.3}$$

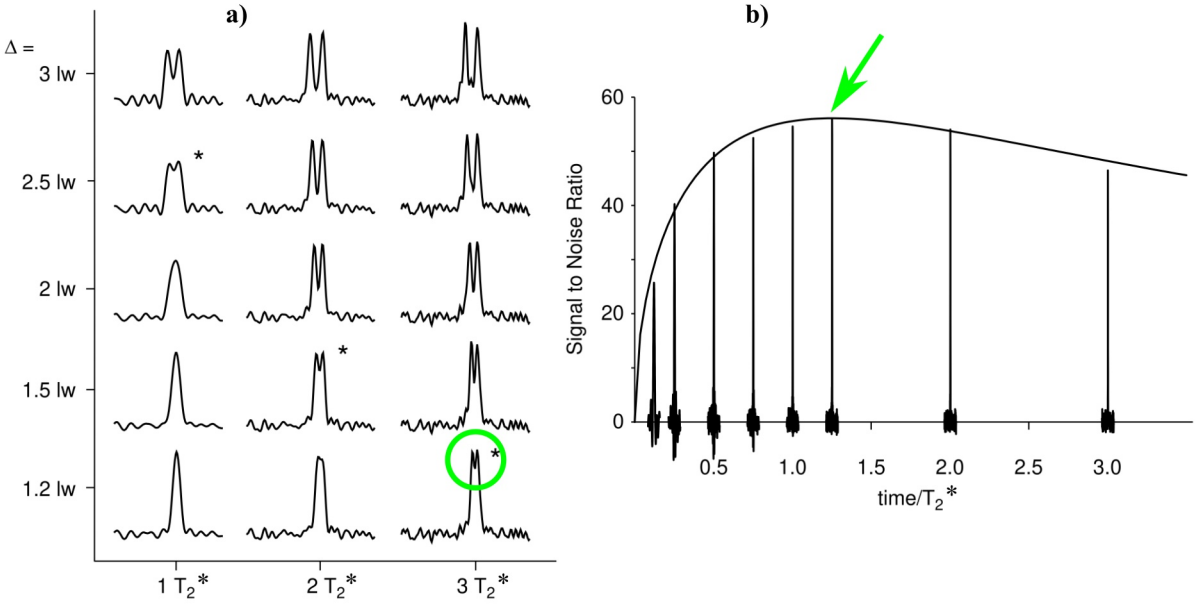


Figure IV.4: Influence of acquisition time a) on resolution and b) on sensitivity. Optimal values are highlighted by green circle and green arrow. Δ: peak separation, *lw*: line width, T_2^*: apparent transversal relaxation time. Adapted from (*19*).

A compromise has thus to be made between resolution and sensitivity. In order to maximise sensitivity and decrease experimental time of nD experiments, an easy solution is to truncate the FID, at the detriment of resolution. What is the impact of this choice?

B.4. FID truncation

FID truncation not only leads to resolution decrease (Figure IV.5a-b), but it also induces artefacts on spectrum after FT (red circle in Figure IV.5b). By multiplying the FID with an exponential apodisation, artefacts are reduced at the cost of peak broadening (red

circle in Figure IV.5c). Further details on FT artefacts will be provided in Section C.4. A common workaround is to use forward linear prediction to extrapolate the FID to a longer acquisition time (*20*, *21*). However, despite strongly improving spectra, this creates distortion on both FID and spectrum (red circles in Figure IV.5d), as missing data cannot be guessed unambiguously. In particular, the number of predicted components has a strong impact on final result. In addition, backward linear prediction can be used to circumvent missing points at the beginning of the FID, due to electronic delays.

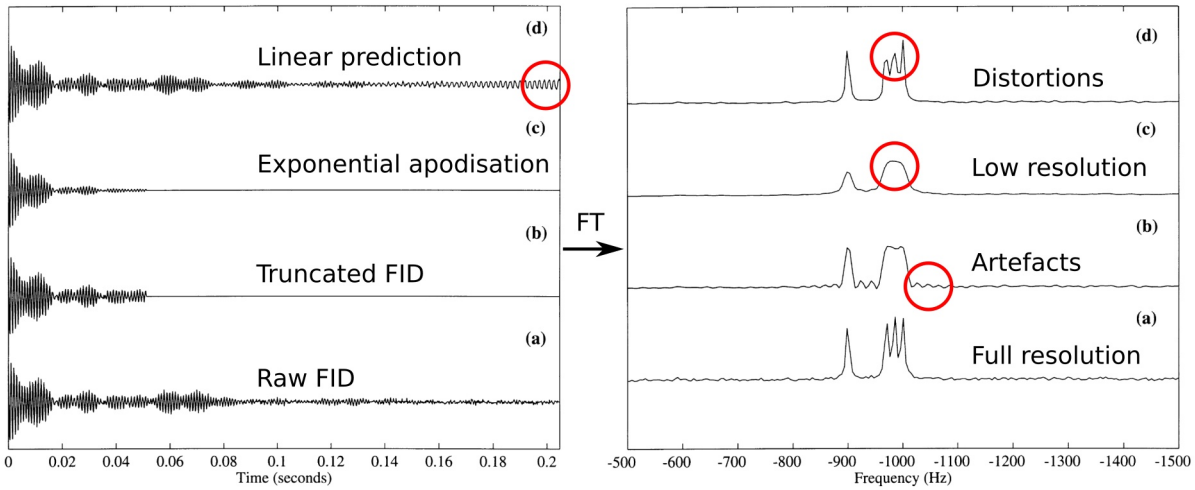
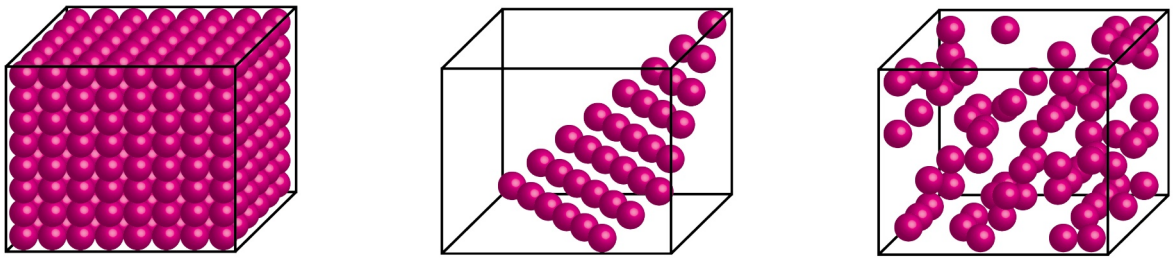


Figure IV.5: left) Time domain and right) their respective spectra after FT of a) Full FID, b) truncated FID, c) truncated FID with exponential apodisation, d) truncated FID with linear prediction. Red circles highlight undesired features. Adapted from (*22*).

As demonstrated in Subchapter B, full sampling of a nD experiment is a balance between resolution (*sampling limited regime*) and sensitivity (*sensitivity limited regime*), in a given experimental time (*23*). In many cases, only a few indirect increments are kept, giving a poor resolution if $AQ < 3\ T_2^*$, and even a low sensitivity if $AQ < 1.26\ T_2^*$ (Figure IV.4).

C. Data acquisition

Radial sampling (Figure IV.6b) and non-uniform sampling (Figure IV.6c) schemes have been developed as alternatives to full sampling (Figure IV.6a). We remind here that direct dimension is not represented, as it is fully sampled at no additional cost. The 3D cube corresponds thus to a 4D NMR experiment. Sparse sampling acquisition times are much smaller than uniform sampling ones, without losing resolution (plain lines in Figure IV.3). Radial sampling and non-uniform sampling are detailed hereafter, as well as NUS sampling strategies, point spread functions and sampling quality.



a) full sampling **b)** radial sampling **c)** non-uniform sampling

Figure IV.6: a) Full sampling, b) radial sampling and c) non-uniform sampling of a 4D experiment. The direct dimension is fully sampled and is neglected. Adapted from (*12*).

C.1. Radial sampling

The basic idea of radial sampling is to simultaneously and regularly increase multiple time dimensions, with a constant ratio between them, giving an angle α (Figure IV.7a). Selected axes or planes are acquired but other points are ignored. The resulting FID can be processed with a standard FT, which gives a projection of the nD spectrum perpendicularly to α (Figure IV.7b). By combining multiple angles, true and false cross-peaks can be discriminated. A true cross-peak correlates with all projections (red disk) while a false one involves only some of them (yellow disk). Radial sampling is useful for high dimensionality ($\geq$ 3D), without needing to reconstruct all points in nD space. The main drawback of this technique is that peaks have strong ridges on projection directions (Figure IV.7c), which may mask small peaks. Moreover, enough projections are mandatory to discriminate all peaks. The correct number of projections is difficult to estimate. A solution is to count the number of peaks on each projection (this is equivalent to compressed sensing with l_0-norm) and to add projections angles if the number of peaks is not constant against angle (*24*). Between 40 to 100 2D projections were used for 5D to 7D experiments (*25*).

The first NMR usage of this concept was Nagayama *et al.* experiment, with processing of a standard 2D to highlight specific cross-sections (*26*). The “accordion” experiment, proposed later by Bodenhausen and Ernst, implemented radial sampling to faster probe a chemical exchange 3D space (*27*). In the latter measurement, the evolution period and the mixing time were concomitantly stretched, hence the name. The 3D space is projected on a 2D plane, thus giving a (3,2)D spectrum. This was generalised as the so-called *reduced dimensionality* experiments (*28*).

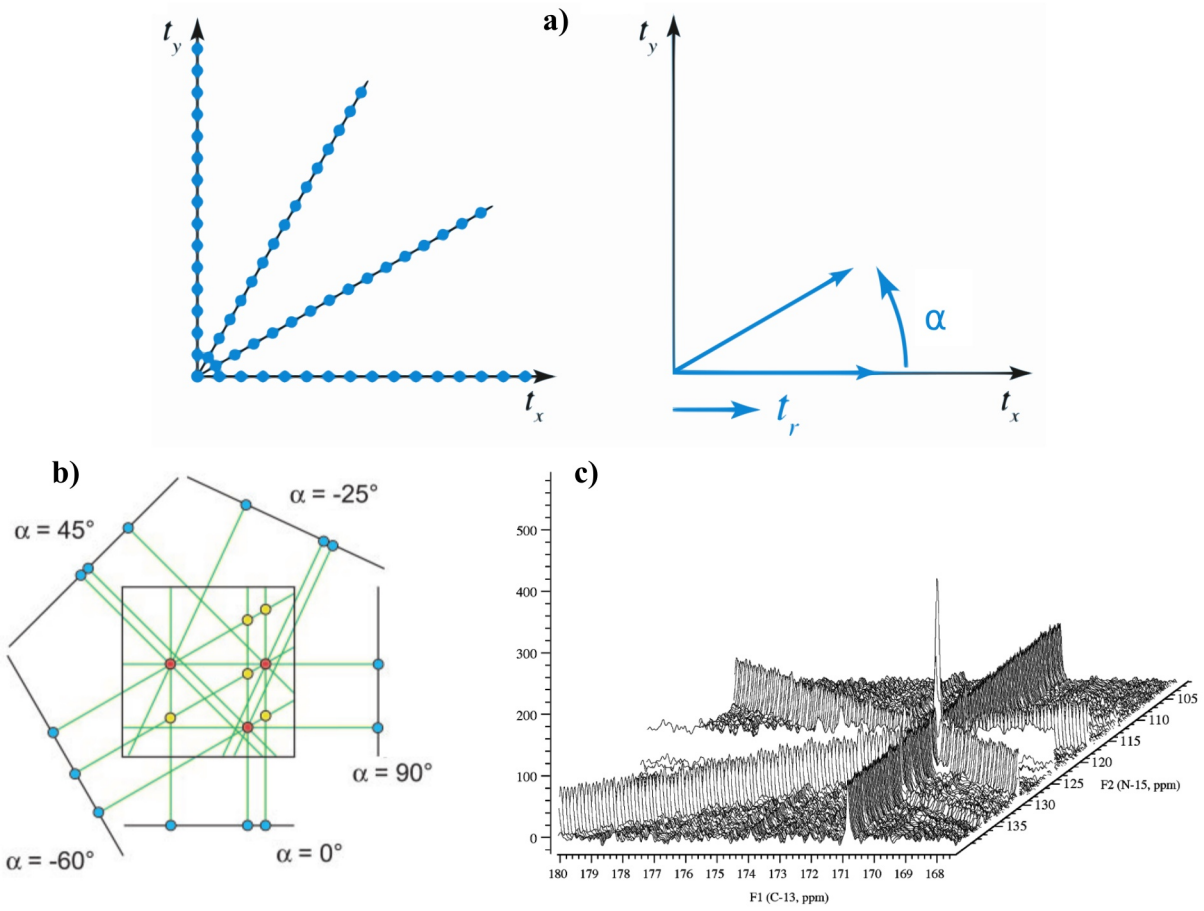


Figure IV.7: Principle of radial sampling: a) constant-time increments with multiple angles (*29*), b) reconstruction by multiple projections, red and yellow disks highlight true and false peaks, respectively, adapted from (*30*), and c) strong artefacts on projections directions (*31*).

In order to reconstruct the nD spectrum obtained with radial sampling, 1D spectra can be combined in a system of linear equations, so-called a G-matrix Fourier Transform (GFT) (*32*). Another solution is to use the projection-reconstruction approach (*33*) based on inverse Radon transform (*34*). A similar idea was applied earlier for astronomy (*35*) and X-ray tomography (*36*). Reconstruction algorithms were detailed in reference (*37*). An additional possibility is to keep only a peak list with N coordinates by use of Automated Projection Spectroscopy (APSY) (*38*). It has already been successfully performed on (7,2)D experiments (*25*). A more detailed historical review of radial sampling can be found in reference (*29*).

C.2. Non-uniform sampling

Barna *et al.* proposed a totally different approach than radial sampling (*39*): instead of acquiring points regularly on specific axes or planes (Uniform Sampling, US), points were

sampled irregularly in all indirect dimensions (Non-Uniform Sampling, NUS). An exponential weighting was chosen to maximise sampling at the beginning of the FID, where *PSNR* was higher (Section B.3), and to decrease acquisition time. Resulting FID and their respective spectra are presented on Figure IV.8a. By simply replacing the missing points by zero and applying FT, a noisy spectrum was obtained. Indeed, for each factor 2 of subsampling, *PSNR* was divided by a factor $\sqrt{2}$, that is to say -3 dB (Figure IV.8b) (*40*). The reason will be detailed in Section C.4. However, Barna *et al.* reconstructed a correct spectrum using maximum entropy. Maximum entropy method will be detailed in Section D.3.

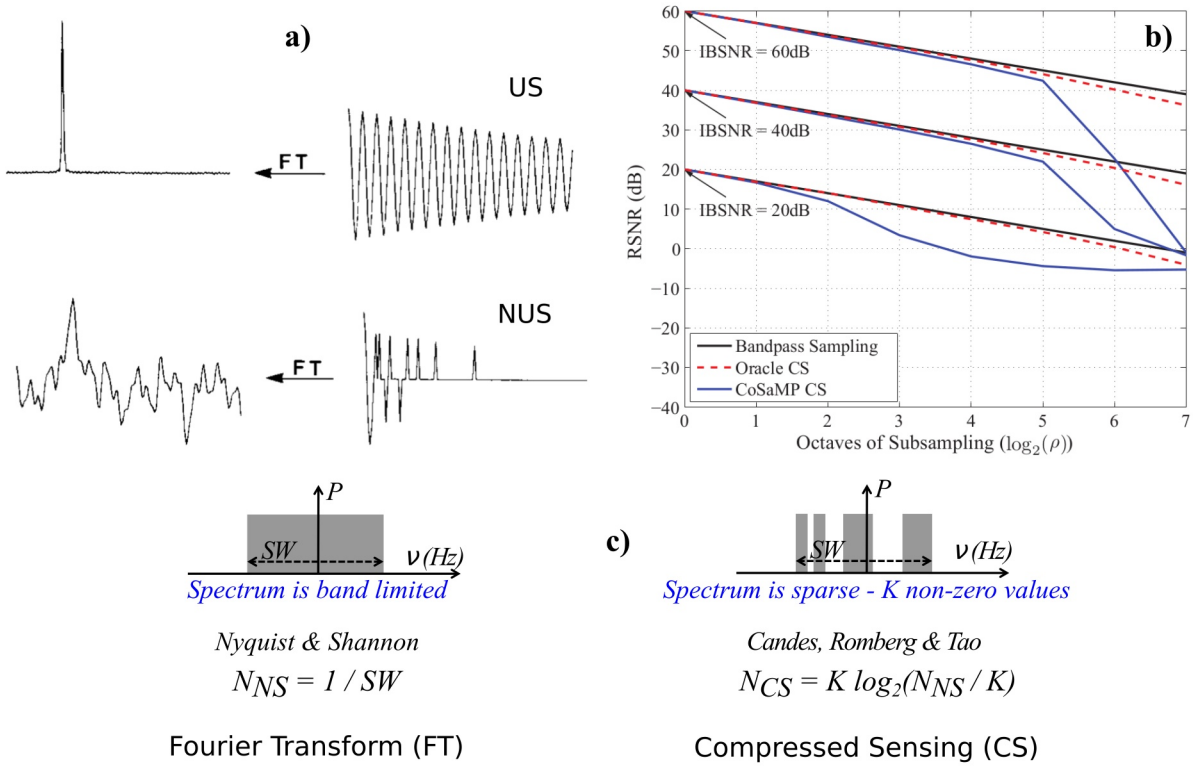


Figure IV.8: a) Comparison of US and NUS FID and their respective spectra (*39*).
b) Impact of subsampling on *PSNR* (*40*). Bandpass sampling: regular undersampling; Oracle CS: algorithm with known zero elements; CoSaMP CS: l_1-norm algorithm.
c) NS *vs.* CS theorems, adapted from (*41*).

In order to overcome subsampling limitation, the key point is to replace the hypothesis from Nyquist and Shannon theorem (NS) (*42*, *43*) to Compressed Sensing (CS) one, which was demonstrated by Candès *et al.* (*44*). While the former states that all frequencies are relevant on Spectral Width (*SW*), the latter postulates that spectrum is sparse, either with *K* non-zero values (Figure IV.8c) (*41*) or with a low Total Variation (TV) as in magnetic resonance imaging. As a consequence, much less points (N_{CS}) are needed to fully reconstruct the spectrum with CS than with NS (N_{NS} points) (*45*). CS reconstruction will be explained in

Section C.3. Probability of success (in red on Figure IV.9) and of failure (in blue) was called the Donoho-Tanner phase transition (*46*), with an intermediate uncertain region (in yellow). It depends on sparsity ratio and sampling coverage. Low sampling coverage is eligible only for low sparsity ratio, *i.e.*, for highly sparse spectra. Gratefully to the resolution increase at higher magnetic field, sparser peaks are obtained improving success rate (black arrow). Moreover, by increasing spectrum dimensionality, the number of peaks is only slightly increasing, whereas sparsity is strongly increasing, enhancing spectrum reconstruction.

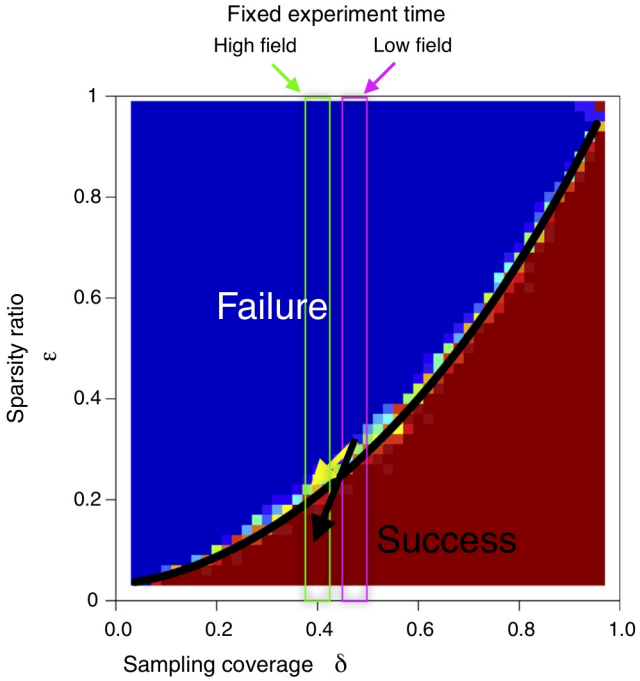


Figure IV.9: Probability of CS reconstruction (*47*). Blue and red regions depict failure and success, respectively, while intermediate region is unsure; black arrow highlights the success improvement due to resolution enhancement at high magnetic field.

C.3. Sampling strategies

In their experiment, Barna *et al.* suggested to sample following an exponential bias, with a higher amount of points at the beginning than at the end of the FID (*39*). Figure IV.10a presents this strategy graphically (*12*, *48*). This *envelope-matched sampling* is in agreement with a recent theoretical study highlighting that any NUS function with a decreasing probability, improves sensitivity compared to the US spectrum (*49*). Alternatively, sensitivity can be maximised for J-coupling experiments using *beat-matched sampling* (Figure IV.10b) (*18*), where more points are acquired on top of oscillations (*50*). However, in the latter case,

additional artefacts may occur (*51*). A further sensitivity gain was proposed by Kumar *et al.* with exponential averaging, by increasing the number of scans for first points (*52*).

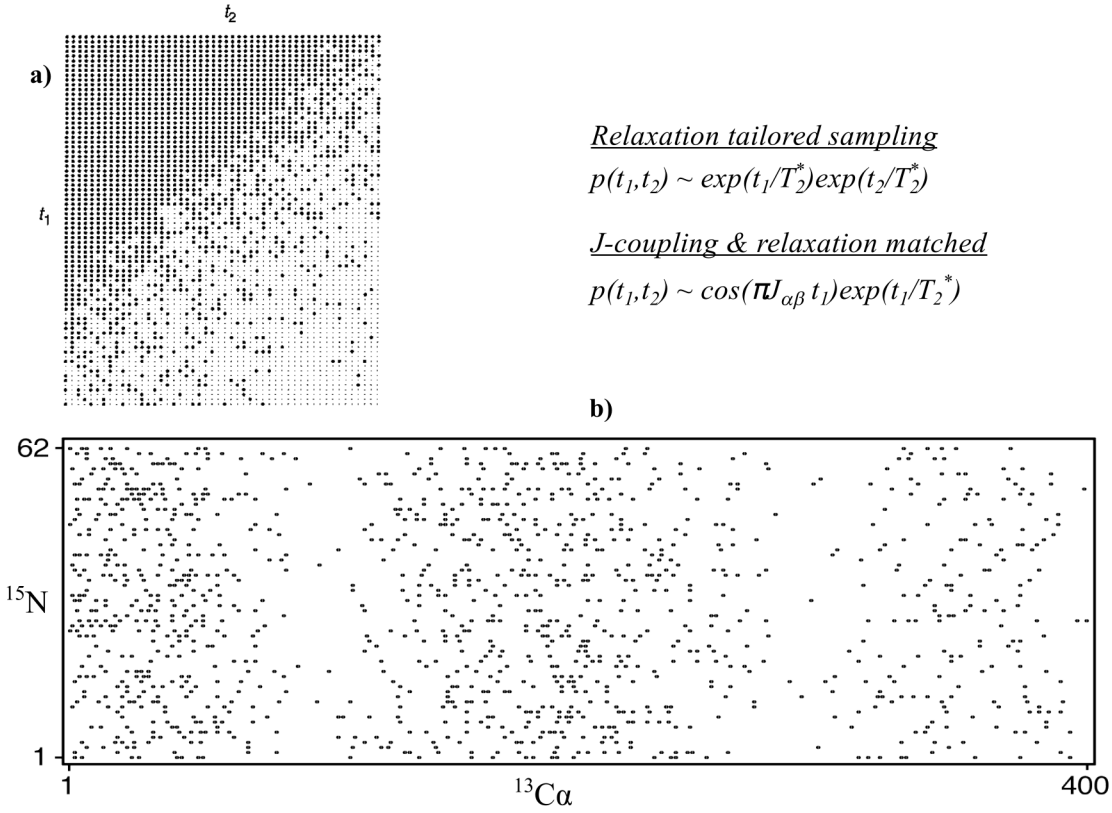


Figure IV.10: 2D sampling schemes with a) relaxation tailored (*envelope-matched sampling*) (*48*) and b) J-coupling & relaxation matched (*beat-matched sampling*) (*18*), adapted from (*12*).

The optimal number of sampling points in an experiment can be checked with targeted acquisition (*53*). Its principle is to randomly acquire points, to reconstruct spectrum in real-time, and to stop when the number of found peaks is constant. A similar idea can be used to monitor chemical reactions with Time-Resolved NUS (TR-NUS): the FID is randomly sampled and for each time-sliding constant number of points, a 2D NMR spectrum is obtained within a few minutes (*54*, *55*). New developments on NUS have also been tested on phase quadrature, considering it as an additional dimension (*56*).

C.4. Point spread function

In this section we will investigate why some sampling schemes give better results than others. In each FID, points are aligned on a Nyquist grid (*42*) and discrete Fourier transformed (Section C.3 of Chapter I). If points are missing, they are replaced by zeroes. This *zero-*

augmented data is equivalent to multiply the full signal by a comb function, whose FT is a complex oscillation, called Point Spread Function (PSF) (*57*, *58*). The FT convolution theorem is presented in Equation IV.4, where ◦ is the element-wise multiplication, and * is the convolution (*59*).

$$FT(a \circ b) = FT(a) * FT(b) \qquad \text{IV.4}$$

In words, FT of a product of two vectors *a* and *b*, is the convolution of the FT of each vector. Applied to sampling schemes, this means that NUS spectrum is the convolution of US spectrum and PSF (Equation IV.5) (*59*).

$$NUS_{spectrum} = US_{spectrum} * PSF \qquad \text{IV.5}$$

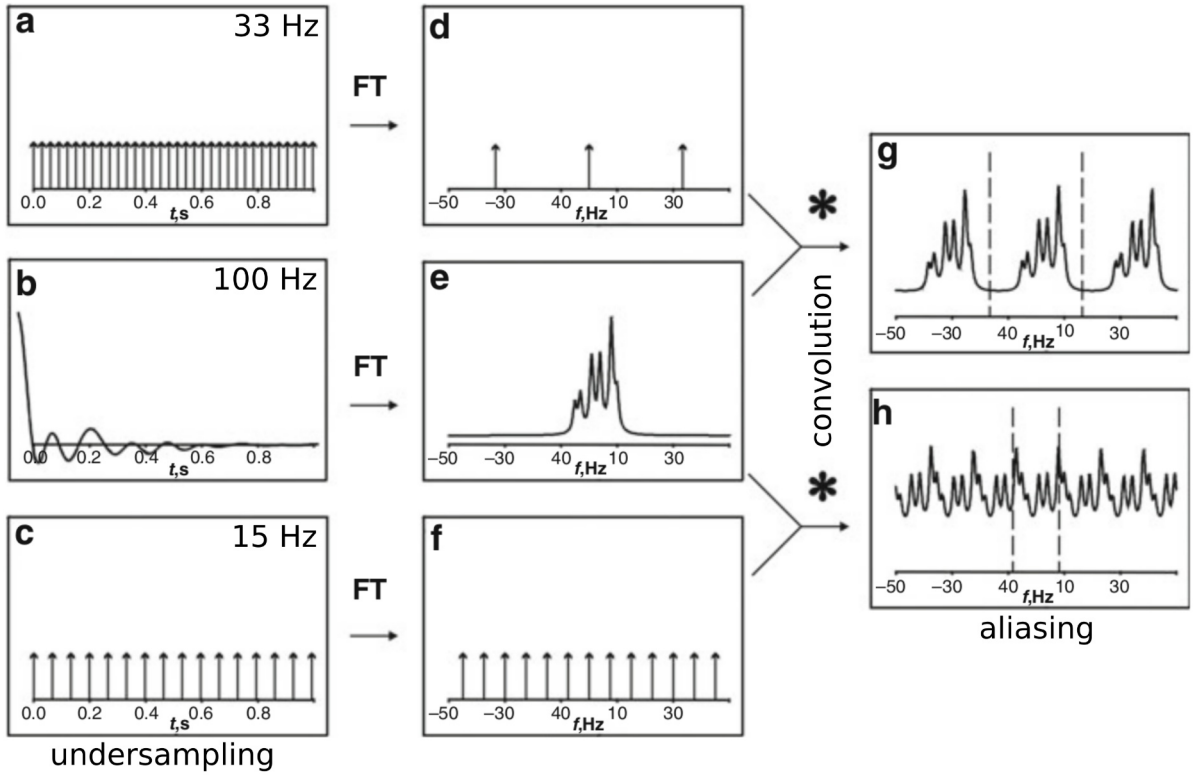


Figure IV.11: Effect of undersampling on spectrum: b) time function sampled at 100 Hz and e) its corresponding spectrum; a) and c) regular undersampling function at 33 and 15 Hz, respectively, and d) and f) their corresponding PSF; g) and h) convolution of e) with d) and f), respectively. Adapted from (*60*).

PSF is thus of crucial importance for NUS. The more points are missing, the more the noise coming from the PSF is intense. On the contrary, artefacts are minimised with full sampling. The simplest comb function is regular undersampling (Figure IV.11). When digitalising a signal at 33 Hz instead of 100 Hz (a and b, respectively), the resulting convolution is a three-fold aliased spectrum (g), which is not so annoying in this example.

However, a worst case is obtained at a digitalisation rate of 15 Hz (c) with an unreadable six-fold aliased spectrum (h).

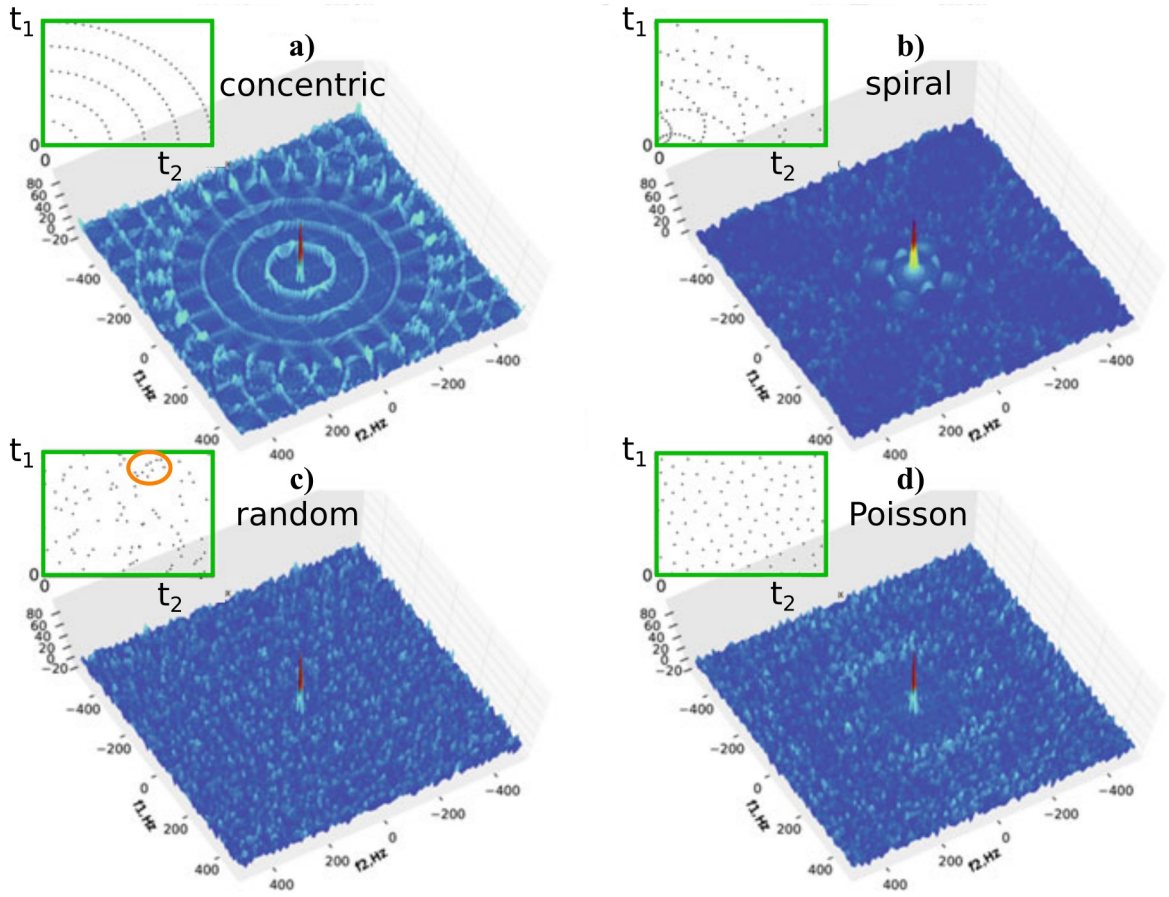


Figure IV.12: 2D sampling scheme (insert) and their corresponding PSF (main figure) for a) concentric, b) spiral, c) random and d) Poisson sampling. Adapted from (*60*).

Similar peak images are obtained with more complex sampling functions, such as triangular pattern (*61*). Many of them were analysed by Maciejewski *et al.* (*62*). Figure IV.12 presents such 2D sampling functions (green inserts) and their corresponding PSF (blue planes) (*60*). Full sampling would have filled the green insert. While concentric (a) and spiral schemes (b) exhibited strong artefacts, random sampling (c) minimised them. Small artefacts were present, similar to noise. It should be noted that this is purely sampling noise arising during processing. In presence of additional thermal noise during acquisition, *PSNR* degrades consequently. However, clusters of points are present with random sampling (orange ellipse in (c) insert), which are overcome with Poisson distribution (d), where the distribution of contiguous skipped measures follows a Poisson law, instead of Laplace law for simple random sampling (*63*). As a consequence, clusters of points are avoided and noise is pushed

away from peak to higher frequencies, improving peak discrimination. Poisson scheme has thus to be preferred (*64*).

C.5. Sampling quality

As computers are not using random points generation but rather pseudo-random series, it is possible to have a reproducible list of sampling points, by selecting a specific series, so-called a *random seed*. The impact of random seed on spectra has been studied multiple times (*65*, *66*). Craft *et al.* proposed to define sampling schedules according to quantiles, and to randomly jitter 25-50% of points around the desired value to remove undersampling artefacts (*67*). While Poisson sampling induces *blue noise*, namely noise at high frequency, bursts of points leads to *red noise* (*68*), *i.e.*, increased noise at the neighbouring of intense peaks (Section B.1 of Chapter I). This latter solution may be useful for Nuclear Overhauser Effect SpectroscopY (NOESY) experiments, where artefacts arising from intense diagonal peaks have to be placed close to the diagonal to avoid masking small peaks away from diagonal. Once a satisfying sampling scheme has been found for an application, it can easily be applied to another sample.

Even for Poisson distribution, and on simulated data in the absence of noise, different random seeds will give different results (*69*). In this example, the worst sampling schedule will give rise to strong vertical artefacts (red arrows on Figure IV.13a) while the best one will not (Figure IV.13b). Following this observation, one need a scoring function to discriminate sampling quality. The PSF ratio (*58*) or its inverse the Peak-to-Sidelobe Ratio (PSR) (*59*) can be such useful tools. PSF ratio compares the sidelobes of PSF to its zero frequency peak. The lower PSF ratio is and the lower FT artefacts are. For instance, PSF ratio equals to 1 for regular undersampling and tends to zero for random sampling (*56*). Unfortunately, PSF ratio does not correlate well to Root Mean Square (RMS), which takes into account any spectrum distortions: no clear correlation is obtained and high PSF can even correlates to low RMS (Figure IV.13c).

Another possible indicator is l_2-norm that measures the total deviation from US spectrum (*70*). A better agreement is obtained between l_2-norm and RMS, with a more linear curve and the appearance of clusters in the graph (Figure IV.13d). However, comparing multiple schedule with l_2-norm or RMS needs a full spectrum reconstruction for each of them, which is difficult and time-consuming. To overpass this limitation, Aoto *et al.* developed a

program called NUSscore (*69*), that combines a weighting score for sensitivity and a randomness score for robustness. It generates 10 000 schedules within a few seconds and rank them by quality. It is therefore very easy to use the best one for a given sampling ratio and a maximum number of indirect increments.

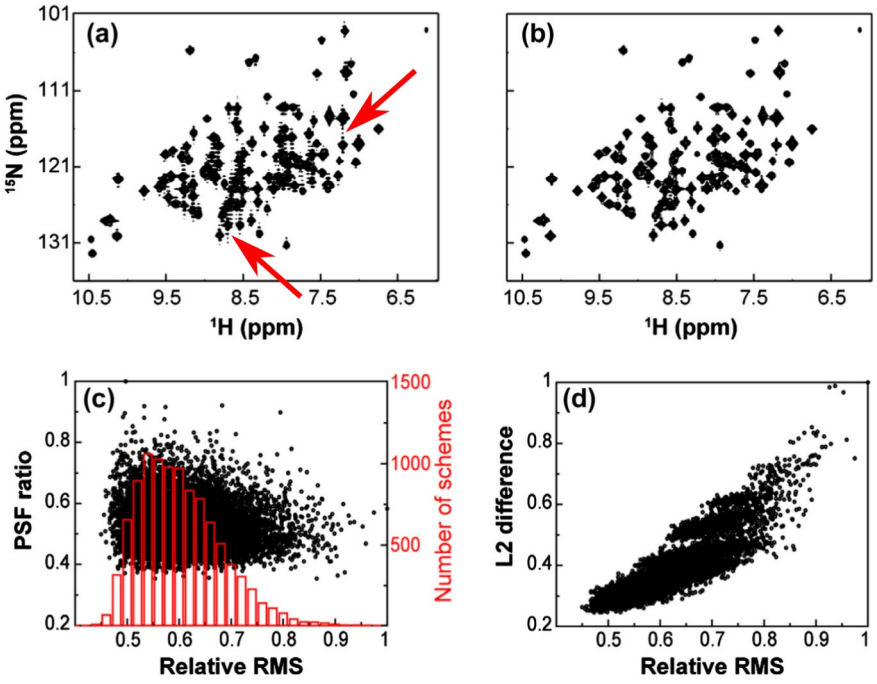


Figure IV.13: Reconstructed spectra for identically processed synthetic noiseless data with resonances approximating an 18 kDa protein and sampled with a) the worst and b) the best 60 % Poisson schedule, based on RMS measurement. c) Correlation of PSF ratio to RMS (black) and number of sampling scheme for each RMS value (red histogram). d) Correlation of l_2-norm to RMS. Adapted from (*69*).

As we have demonstrated in Subchapter C.1, non-uniform sampling can be very efficient to reconcile resolution and sensitivity. The former can be improved with nD acquisition times as long as 3 T_2^*, whereas the latter benefits of a higher number of scans in a given time and of reduced noise introduction, due to lower density of points above 1.26 T_2^*. Nevertheless, sampling scheme has to be defined carefully in order to avoid FT artefacts.

D. Data reconstruction

After optimising sampling scheme, data has to be reconstructed. Theory will first be introduced using a system of linear equations. We will then describe the available reconstruction algorithms and focus on maximum entropy and compressed sensing before characterising reconstruction quality.

D.1. System of linear equations

Using FT to obtain a complex spectrum *s* from a complex FID *f* is similar to solve an inverse system of linear equations (Figure IV.14a). This problem is inverse as one observe the influence of parameters on data (oscillation on FID *f*) without knowing these parameters (frequencies). In case of US acquisition, the inverse FT (iFT) matrix Φ is a square, with $N \times N$ rows and columns (Figure IV.14b). The spectrum *s* is the unique solution to this system of linear equations.

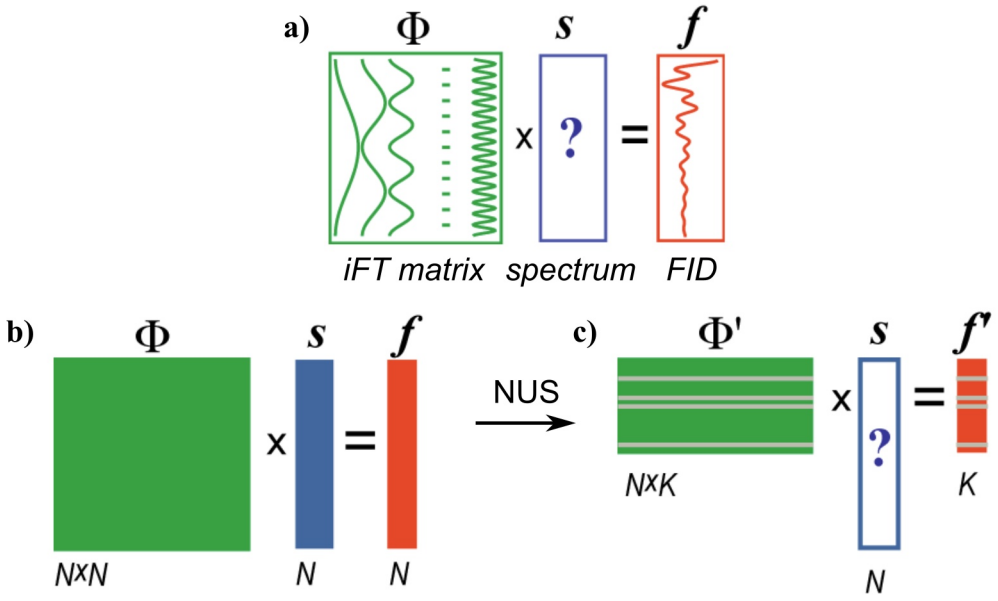


Figure IV.14: System of linear equations for a) and b) US and c) NUS. Φ, Φ', *s*, *f* and *f'* are inverse FT US matrix, inverse FT NUS matrix, spectrum, US FID and NUS FID, respectively. Adapted from (*12*).

In case of NUS, only *K* points are acquired on FID *f'* (Figure IV.14c), with missing regions reduced to grey lines. As a consequence, the iFT matrix Φ' is not a square any more and this system is under-determined, with less equations than variables. This implies that its solution is not unique and that additional assumptions are needed (*71*). However, any presupposition can have strong consequences, as was demonstrated by a deliberately wrong model, where a photograph of Einstein arose from 1000 pure white noise images (Figure IV.15) (*72*).

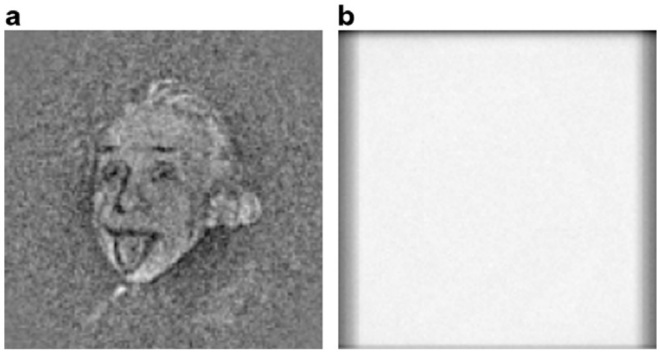


Figure IV.15: A photograph of Einstein is used as a model on 1000 pure white noise images (*72*). a) Wrong reconstruction using alignment on model. b) Close to random reconstruction based on mutual correlation of points.

The key point of Compressed Sensing (CS) is to assume that *s* dataset is sparse in frequency domain, *i.e.*, it has only a few non-zero elements, and that it is incoherent, *i.e.*, it spreads out in the complementary time domain (73). Sparsity consists in maximizing the number of null values in s (l_0-norm) which was shown to be generalizable to all l_p-norms, where p ≤ 1. The l_p-norm of a vector *f* with *N* values f_k is defined by Equation IV.6 (*74*):

$$\|f\|_{lp} = \left(\sum_{k=1}^{N} |f_k|^p \right)^{\frac{1}{p}} \qquad \text{IV.6}$$

l_p-norm functions are presented on Figure IV.16 for a 2 points vector *f* with $p = 2$, 1 and 0, from left to right, respectively (*75*). For longer vectors, l_p-balls will correspond to mathematical objects with *N* axes. l_2-norm (left) and l_1-norm (middle) are convex, whereas l_0-norm (right) is non-convex (*76*). As a consequence, the latter is much sensitive to sparse than to non-sparse solutions (black and white disks, respectively). Here, sparse solutions have values on only one axis.

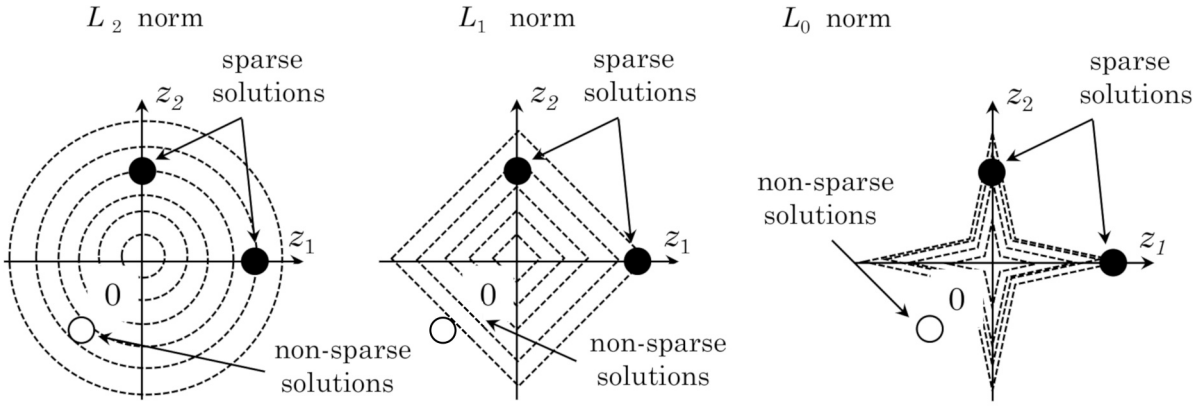


Figure IV.16: l_p-balls for $N = 2$ and $p = 2$, 1 and 0, adapted from (*75*).

CS way to solve NUS problem is to minimise l_p-norm (Equation IV.7),

$$min\|s\|_{lp} \text{ subject to } \Phi' s = f' \quad \text{IV.7}$$

with at least *2K* sampling points for l_0-norm or a slightly higher number $K\ log_2(N_{NS} / K)$ sampling points (*77*) for l_1-norm. However, the l_0-norm is non-continuous and cannot be minimised, involving more demanding computations. The l_p-norm with $p < 1$ is non-convex, which can result in multiple solutions. l_1-norm is often used instead of l_0-norm, with a high success rate. When data is corrupted by noise, a more general formula can be used (Equation IV.8),

$$min\left[\|\Phi' s - f'\|_{l2}^2 + \lambda C(s)\right] \quad \text{IV.8}$$

where λ is a Lagrange multiplier, that keeps the balance between the measured data (first term) and the sparsity (second term) (*78*). *C(s)* is a cost function, either l_p-norm, total variation, maximum entropy or matrix rank. In NMR, this can be rewritten as Equation IV.9, where η is a noise estimation.

$$min\, C(s) \text{ subject to } \|\Phi' s - f'\|_{l2}^2 \leq \eta \quad \text{IV.9}$$

D.2. Reconstruction algorithms

Table IV.2: Available reconstruction algorithms and their limitations (*79*).
Except FT, all methods are non-linear.

Method	Compatible with		Limitations
	Sampling scheme	Experiments/signals	
LP	US	Lorentzian signals	False positives and frequency shifts for low S/N or severely truncated data
GFT	Radial	nD with n ≥3	No reconstruction of the nD spectrum (only subspectra) Not suitable for low S/N and/or crowded spectra Sampling artifacts
Linear least square	All	All	Previous knowledge of all line shapes Choice of a noise threshold Same limitation as MDD
MDD	All	All	Artifacts for low S/N, overlapping signals or highly sparse sampling scheme
Covariance	All	Homonuclear correlation 2D	Only applicable to nD spectra with specific symmetries
SIFT	All	All	Previous knowledge of spectral regions without signals Artifacts produced by the peak tails
MaxEnt	All	All	Non linear
MINT	All	All	Artifacts for severely truncated sampling
CS	All	All	Artifacts or signal suppression for severely truncated sampling or broad peaks

Many reconstruction algorithms are available to process nD sparse spectra, which were described numerous times (*1*, *80*) and summarised in Table IV.2 by Lesot *et al.* (*79*). All acronyms in the method column and some others are explained hereafter. Optimisation algorithms (gradients, proximal function, iterative thresholding, or iteratively reweighted least

squares) have not be confused with regulation parameters (l_0-norm, l_1-norm, maximum entropy, or matrix rank).

A first case of signal reconstruction is truncated FID. It can be overcome with Linear Prediction (LP) (*20*), potentially coupled to Singular Value Decomposition (LPSVD) (*81*). However, it assumes that all peaks are Lorentzian (*82*), which may not be the case, especially in solid-state NMR, where Gaussian or more complex lineshapes are predominant. Moreover, if FID truncation is too strong, it is impossible to separate peaks at close frequencies, simply because not enough discriminating information has been acquired.

Specific algorithms are needed for radial sampling, such as G-matrix FT (GFT) (*32*) and projection-reconstruction (*33*), that were presented in Section C.1. More detailed informations are available in reference (*37*). Their principal limitation is the need to have at least three NMR dimensions. Covariance is another special case, that is only applicable to homonuclear correlations (*83*).

For NUS sampling, alternative methods have been developed to circumvent missing values. The simplest reconstruction algorithm consist to simply ignore missing values, without adding any zeroes, and to apply discrete FT on the resulting data set. As a consequence, frequency oscillations are not regularly digitalised in time domain and distortions are observed on spectrum, but no sampling noise is induced. Another possibility is to replace missing points on FID by zeroes before FT. This solution is called *zero-augmented FT* or *minimal power* (*76*). Indeed, as stated by Parseval's theorem, integrated power is equal in frequency and time domains (Equation IV.10) (*84*).

$$\int_{-\infty}^{+\infty} [S(\nu)]^2 \, d\nu = \int_{-\infty}^{+\infty} [s(t)]^2 \, dt \qquad \text{IV.10}$$

According to this theorem, by minimising time signal amplitude, power of frequency signal is also minimised. This is equivalent to l_2-norm optimisation and many non-sparse values are obtained (white disk on Figure IV.16), which correspond to artefacts as highlighted in Section C.4. Nevertheless, this latter method is useful for fast spectrum computation. A suitable alternative is Non-Uniform FT (NU-FT). When points are non-regularly sampled, it is necessary to recast them on-grid and to weight them accordingly to the actual delay between points (*85*). While a Lagrange interpolation can be implemented for 1D data sets, Voronoi cells are necessary for higher dimensions (*86*).

Another possibility is to use least-squares to fill in the gaps. It can benefit from previous knowledge of peaks through a 1D NMR experiment (*87*). The Technique for Importing Greater Evolution Resolution (TIGER) (*88*) and the Application of a Restricted Linear Least Squares Procedure (ANAFOR) (*89*) are such options. Multi-Dimensional Decomposition (MDD) is a generalisation of these methods by describing nD FID as the product of 1D vectors (*90*). A unique solution is obtained only for spectra higher or equal to 3D, with the need to have at least one point per plane (*91*).

Other solutions are available, such as Spectroscopy by Integration of Frequency and Time domain information (SIFT), which alternates FT and iFT to impose measured time intensities and known zero frequencies (*80*). Low-Rank (LR) is also useful to reconstruct NUS FID (*92*). This process is based on Hankel matrices and SVD, as explained in Subsection B.3.b. of Chapter V. However, an elongated matrix has been chosen in LR, imposing that matrix rank is less than 10 %. This is not the best choice as it decreases computation time at the expense of denoising quality. Moreover, SVD has a strong impact on Gaussian shapes. Finally Virtual Echo (VE, Section D.5), has been developed from causality principle (*93*). This preliminary step, before NUS spectra reconstruction, avoids additional errors introduced by the missing negative times, and improves overall quality. VE can be combined with many reconstruction procedures.

In addition to this menagerie of algorithms, the two most popular are maximum entropy and compressed sensing, which will be developed in the two following sections.

D.3. Maximum entropy

Maximum Entropy Method (MaxEnt or MEM) was developed in 1957 (*94*) and consequently applied to spectral analysis (*95*, *96*), astronomy (*97*, *98*), X-ray tomography (*99*), electron spin-echo spectroscopy (*100*), and NMR spectroscopy (*101*, *102*). As MaxEnt can theoretically improve both resolution and sensitivity, it has been used either to complete missing acquisitions or to denoise data.

The schematic diagram of MaxEnt is presented on Figure IV.17 (*103*), and could as well explain any gradient based optimisation. MaxEnt simulates a trial spectrum f, typically starting from a blank spectrum, and compares its iFT mock data m to the empirical FID d. f is then updated by applying gradients until m difference to d is minimised. Upon the infinity of

corresponding spectra, the one with the lowest amount of information, *i.e.*, the maximum entropy is chosen (*1*, *104*). This is a constrained optimisation, that maximises entropy instead of minimising l_1-norm.

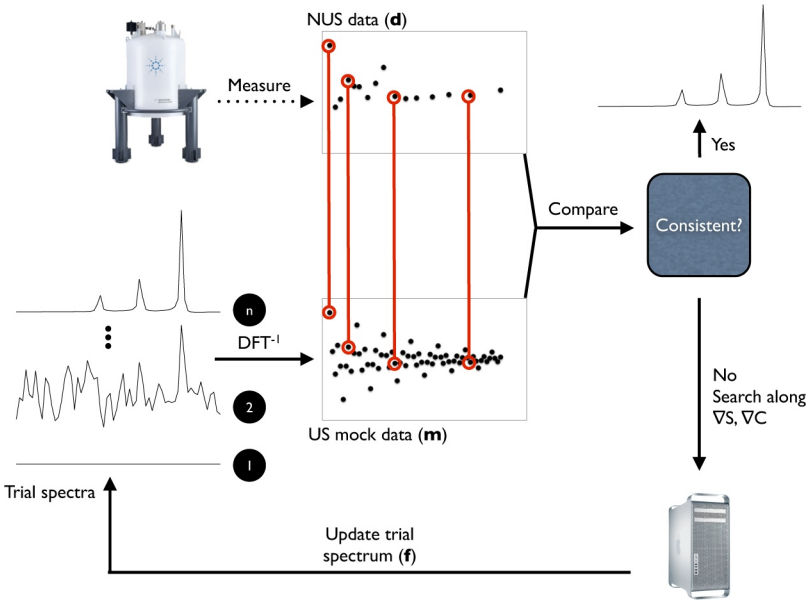


Figure IV.17: Schematic diagram for MaxEnt algorithm (*103*). See text for details.

However, all methods but FT are non-linear processes, whose sensitivity gain has to be taken with caution. It can improve apparent *PSNR* without enhancing actual detection of small peaks. Hence, Donoho *et al.* warned that 'the noise suppression offered by maximum entropy reconstruction could (in this special case) be equally well obtained by a "cosmetic" device: simply displaying the conventional Fourier transform reconstruction using a certain nonlinear plotting scale for the vertical (*y*) coordinate' (*105*).

To avoid propagation of nonlinearity, MaxEnt has to be applied simultaneously on all NUS dimensions. Alternatively, Maximum entropy INTerpolation (MINT) improves MaxEnt linearity by using a high Lagrange multiplier (*106*). Forward Maximum entropy (FM) is also a valuable tool that does not enhances experimental data points, at the detriment of noise reduction (*107*).

D.4. Compressed sensing

Candès *et al.* and Donoho separately developed CS concept in 2004 (*44*, *108*), that revolutionised Nyquist and Shannon sampling theory (*42*, *43*), by demonstrating that exact signal reconstruction can be obtained under sparse sampling in specific conditions and

without noise. They theorised some results already known in NMR and Magnetic Resonance Imaging (MRI) (*58*). Mathematical details on CS can be found in reference (*109*). CS algorithms and their pitfalls were presented in reference (*110*).

A first algorithm named CLEAN, which is not an acronym, was implemented for radio-astronomy in 1974 (*111*). It was then applied to 2D and 4D NMR (*112*, *113*). CLEAN process is presented on Figure IV.18 (*110*). After minimal power FT, the maximum point is added to the sought spectrum, whose iFT is subtracted from non-zero experimental data. After a few iterations, noise level is reached and the sought spectrum becomes the final spectrum. CLEAN belongs to a family known as matching pursuit (*114*).

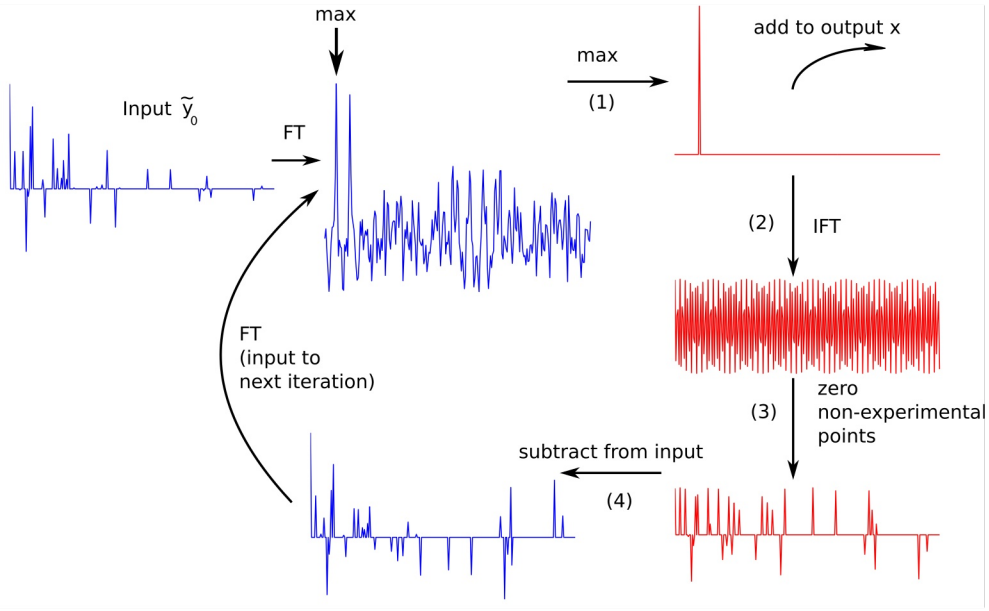


Figure IV.18: Schematic diagram for CLEAN algorithm (*110*). See text for details.

Another family of CS algorithms is Iterative Soft Thresholding (IST) (*115*). Two variants are available: IST-S by Stern *et al.* (*116*) and IST-D by Drori *et al.* (*117*). IST-S replaces only zero-points in experimental data, tending to US FID, and was used by early versions of MDDnmr, the software used to process data with Multi-Dimensional Decomposition (MDD) reconstruction (*8*, *90*), freely available at (*118*). IST-D is similar to CLEAN, which tends to noise level, but with multiple points at a time, and only above the desired threshold. IST-D was used in later versions of MDDnmr and in Harvard Medical School IST (hmsIST) (*119*). IST-S and IST-D were compared to NESTerov's Algorithm (NESTA), which is an acceleration of IST method (*120*, *121*). In addition, low-rank plus

sparse (L+S) matrix decomposition, combining Singular Value Thresholding (SVT) and IST proved to be efficient in MRI, both for denoising and for recovering sparsity (*122*).

Finally, Iteratively Reweighted Least Squares (IRLS) uses a norm factor tending to zero $l_{p\to 0}$ (*8*), thus being close from a hard thresholding method, as opposed to soft thresholding (*123*). Soft thresholding (l_1-norm) is a continuous function with a translation of data (Figure IV.19a) and convexity, whereas hard thresholding (l_0-norm) exhibits a discontinuity for values under the threshold (Figure IV.19b) and non-convexity. Other hard thresholding methods are available (*124*). They are convenient to decrease the number of NUS points needed (Section D.1). However, this minimisation increases sparsity, with the risk to have multiple narrow lines for a single broad peak (*116*), and is computationally expensive (*125*).

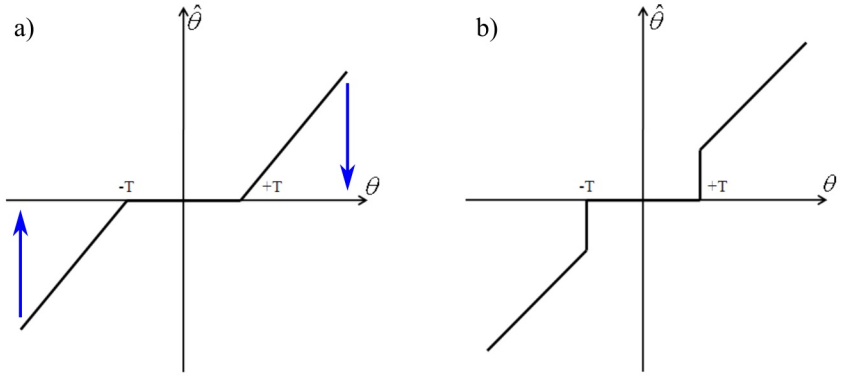


Figure IV.19: a) Soft thresholding and b) hard thresholding, adapted from (*123*).
Blue arrows highlight the vertical shift for values above the soft threshold.

D.5. Virtual echo

Mayzel *et al.* improved CS using Virtual Echo (VE) (*93*). Without NUS, the FID started with a maximum and decreased to noise (Figure IV.20b). This was due to the causality principle, telling that the excitation corresponds to time zero and that no negative time can be observed before. In other words, there is 'no output before the input' (*126*). As a consequence, imaginary part in dispersion is present on the spectrum (in red on Figure IV.20a). However, using NUS and CS reconstruction, the algorithm tends to minimize the imaginary part on the spectrum (Figure IV.20c-d). According to the Kramers–Kronig relations (*127*, *128*), also involved in Hilbert transform (*129*), this is equivalent to add a negative time scale. However, as the number of points is constant and as FT is a cyclic function, an echo appears at end of FID. Total acquisition being twice shorter, resolution is decreased. Moreover, imaginary part

suppression on spectrum is not perfect (red circle) and this induces quantification errors. Result can be improved using Zero-Filling (ZF) before CS (Figure IV.20g-h). Mayzel's approach was to reflect FID from left to right instead of adding zeroes. In such a case, imaginary artefacts are minimised and quantification is optimised (Figure IV.20e-f).

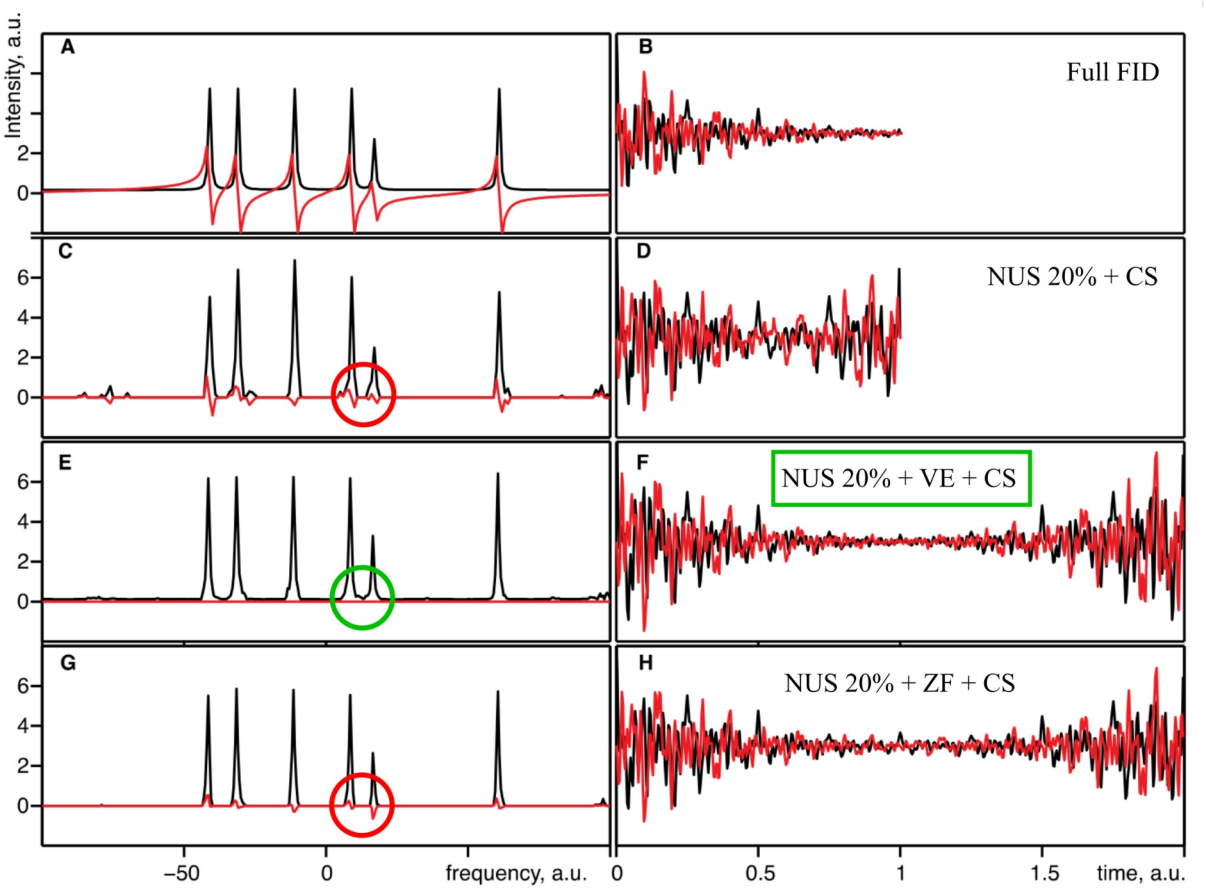


Figure IV.20: Virtual Echo (VE), adapted from (*93*). a) and b) US acquisition; c) and d) NUS 20% and CS; e) and f) NUS 20% with VE and CS; g) and h) NUS 20% with ZF and CS; left and right columns: spectral and time domains, respectively.

D.6. Reconstruction quality

Similarly to sampling quality, one needs an adequate indicator to select the best reconstruction algorithm. Up to now, no consensus has been found and it may depend on applications and on chosen parameters. All in all, the best choice will maximise true positive signals, while minimising false positive detection, when artefacts or noise are wrongly attributed to signals.

Wu *et al.* proposed a pair of noise-normalised measurements to detect the optimum number of reconstruction iterations, and noise increase by additional iterations (overfitting)

(*130*). Their cross-validation technique necessitated to exclude 5-10 % of the acquired data. Noise on processed data (R^{noise}_{work}) was then compared to noise from excluded data (R^{noise}_{free}). However, dismissing data may be counter-productive when sampling is already at a very low level, risking to have too few points to accurately recover the spectrum. They compared convex l_1-norm minimisation, MaxEnt and IST-S and obtained similar reconstructed spectra for the optimum number of iterations, but divergence for additional ones.

In situ Receiver Operating Characteristic analysis (IROC) is another relevant tool for algorithm selection (*131*). Synthetic data are added at blank experimental frequencies and their presence and distortions are checked in processed data. Figure IV.21 exhibits such an IROC curve and is characterised by its Maximum Recovery rate at Minimum False discovery rate (MRMF, the leftmost point), its Distance to the Perfect Classifier (DPC, closest point from top left corner) and by its Area Under the Curve (AUC). MRMF and AUC have to be maximised, whereas DPC has to be minimised.

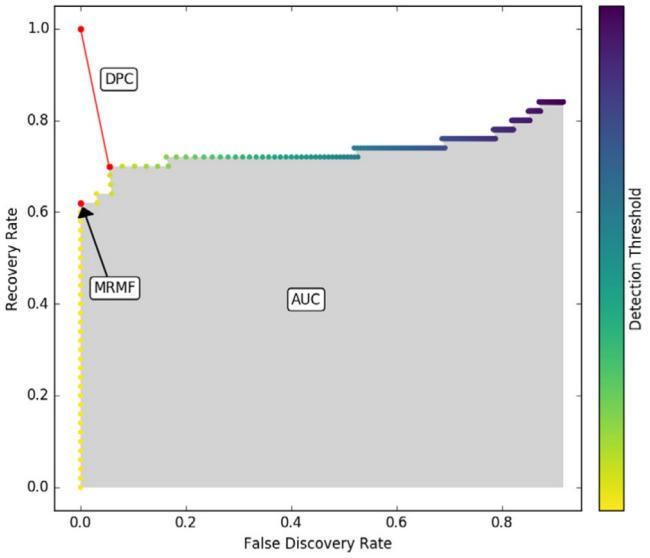


Figure IV.21: *In situ* Receiver Operating Characteristic analysis (IROC) (*131*).
MRMF: Maximum Recovery rate at Minimum False discovery rate, DPC: Distance to the Perfect Classifier, AUC: Area Under the Curve.

In Subchapter D, we have highlighted that multiple reconstruction algorithms are available, each one with its own advantages and drawbacks. MaxEnt and CS are the most used ones. The chosen parameters set strongly influences the IROC curve.

E. NUS in solid-state NMR

In this subchapter NUS will be applied to solid-state NMR, starting with the available literature. After comparing reconstruction algorithms, some sampling schemes will be checked (TopSpin, random, Poisson and hybrid). However, for a more complete study, statistics on many random seed would be needed, which is out of present work focus.

E.1. Literature

Despite abundant literature on NUS in liquid-state NMR, the number of studies in solid-state NMR is much more reduced. Lesot *et al.* published a detailed review on NUS in solids and mesophases (*79*).

Reduced dimensionality was applied to H-C-N triple resonance experiments, either for (3,2)D single quantum spectrum of N-Aceyl-Valine-Leucine tri-peptide (*132*), for (3,2)D double quanta acquisition of histidine amino acid (*133*), or for (3,2)D and (4,3)D correlation of GB1 protein (*134*). Zero-augmented FT was preferred to obtain ^{1}H-^{1}H distances in ubiquitin through 3D and 4D Dipolar Recoupling Enhanced by Amplitude Modulation (DREAM) (*135*). A 3D spectrum of a microcrystalline sample of GB1 protein domain was processed with SIFT (*136*).

Truncation in the indirect dimension was thwarted with TIGER for Magic Angle Turning (MAT) of 2,6-dimethoxynaphthalene (*88*), and for an isotropic-anisotropic separation experiment of small molecules (*137*). ANAFOR was implemented for Multiple Quantum Magic Angle Spinning (MQMAS) derived experiments of aluminophosphates (*89*, *138*). Covariance proved to be very efficient to simultaneously compensate FID truncation and decrease noise for homonuclear correlations (*139–142*). Peak Signal-to-Noise ratio (*PSNR*) dependence on sampling scheme (*143*) and on number of scans (*144*) was analysed with histidine.

MaxEnt method was first applied in NUS solid-state NMR to ^{17}O MQMAS of various samples, including hydroxyapatite (*145*). In a study on 2D Polarisation Inversion Spin-Exchange at the Magic Angle (PISEMA), it was highlighted that improved *PSNR* can be obtained by applying MaxEnt both in direct and indirect dimensions (*146*). An improved linearity was obtained with MINT that yields to 2-4 experimental time decrease on

thioredoxin (*106*). By combining MINT with Paramagnetic-relaxation-Assisted Condensed data Collection (PACC), it was even possible to attain 16-fold time saving on 3D spectra (*147*). In the context of high dynamic range, such as ^{13}C-^{13}C Dipolar Assisted Rotational Resonance (DARR), where intense peaks artefacts can mask small peaks, it was suggested to apply MINT with a conservative NUS sampling schedule keeping 50 % of the points (*148*).

MDD allowed semi-automatic backbone assignment of an insoluble, non-crystalline protein assembly (*149*). The influence of hybrid US/NUS schedule was studied with CS-IST-D algorithm on simulated data and on 2D and 3D Separated Local Field (SLF) spectra of a membrane protein in aligned bicelles (*142*, *150*). CS-IRLS was also performed on Natural Abundance Deuterium (NAD) 2D of enantiomers aligned in lyotropic Chiral Liquid Crystals (CLC) (*77*). An original study used CS-l_1-norm for NMR magnetometry of glycine under magic angle spinning (*151*). Finally, in a 4D ^{1}H-^{1}H correlation with diagonal suppression, Linser *et al.* sampled only 2 % of points and processed spectra with CS-hmsIST on proteins (*152*).

E.2. Reconstruction algorithms

In a first attempt to use NUS in solid-state MAS NMR, we applied it to a mobile sample, namely gelatin, which is the denatured form of collagen protein, itself being the main organic phase of bone (Section E.3 of Chapter I). A 2D $\{^{1}H\}$-^{13}C Insensitive Nuclei Enhanced by Polarization Transfer (INEPT) NMR spectrum of this sample was acquired with NUS (Figure IV.22 bottom left) and compared to the US spectrum obtained in twice the time (top left). Default Topspin random sample scheme was used here (seed 54321), with 256 slices and 25 % of points sampled.

This figure and the following ones are organised as follows. Two representative 2D spectra are presented on top left and bottom left, with dotted horizontal black lines showing *PSNR* measurement and dotted vertical black lines showing the extracted slices. Middle of figure displays the extracted vertical slices from the two left 2D, highlighted with arrows, and additional extracted slices from other 2D with different acquisition or processing parameters. Sampling scheme is presented at the right of the extracted slices (pink curve). The used pulse sequence is drawn on top right of the figure. Coloured ellipses evidence result quality based on peak shape and noise level, with bad, intermediate or good results in red, orange or green, respectively. $PSNR_{max}$ based on maximum of noise was measured according to:

$$PSNR_{max} = \frac{H_{signal}}{h_{noise\_peak\_peak}/2} \quad \text{IV.11}$$

Zero-augmented FT, MDD, IST, IST with Virtual Echo (IST_VE), IRLS_VE and full sampling ^{1}H slices were processed using MDDnmr. Unfortunately, it was not possible to compare reconstruction from MaxEnt algorithms, which required a totally different workflow based on Rowland NMR ToolKit (RNMRTK) (*153*, *154*). Unsurprisingly, FT results (brown curve) were very noisy, due to the absence of FID reconstruction. A slight improvement was obtained with MDD (purple curve). This poor result was explained by the need to have 3D or higher dimensions for a unique MDD solution. IST (red curve) gave much better results with decreased noise. IST_VE (green curve) marginally improved peak shape (orange circles). Finally, peaks were narrowed with IRLS_VE (orange curve), that gave the best results on this spectrum. It was thus possible to obtain a $PSNR_{max}$ multiplied by 1.5 in half of the time with NUS (one quarter of slices and twice scans), as compared to full sampling (blue curve), *i.e.*, a gain of 2.1 in $PSNR_{max}$ per unit of time.

Visually, noise in indirect dimension differed depending on reconstruction algorithm. To better characterise it, corresponding histograms of noise values obtained under Python software are presented on Figure IV.23. With zero-augmented FT (brown curve), noise distribution was irregular, with holes into histogram, due to artefacts of point-spread function. MDD noise (purple curve) was lowered but very similar. With IST (red curve) and IST_VE (green curve), many points contained low noise but some others were much higher. This was especially disturbing, as such high noise values could easily be confused with signals. A more regular noise distribution was obtained with IRLS_VE (orange curve), similar to full sampling noise (blue curve). These results illustrated that processing noise was not necessarily Gaussian, whereas US spectra are Gaussian. In that case, $PSNR_{rms}$ based on Root Mean Square noise (Equation IV.12, Section B.3 of Chapter I) was not a relevant measurement and $PSNR_{max}$ had to be preferred (*155*, *156*).

$$PSNR_{rms} = \frac{H_{signal}}{\sigma_{noise}} \quad \text{IV.12}$$

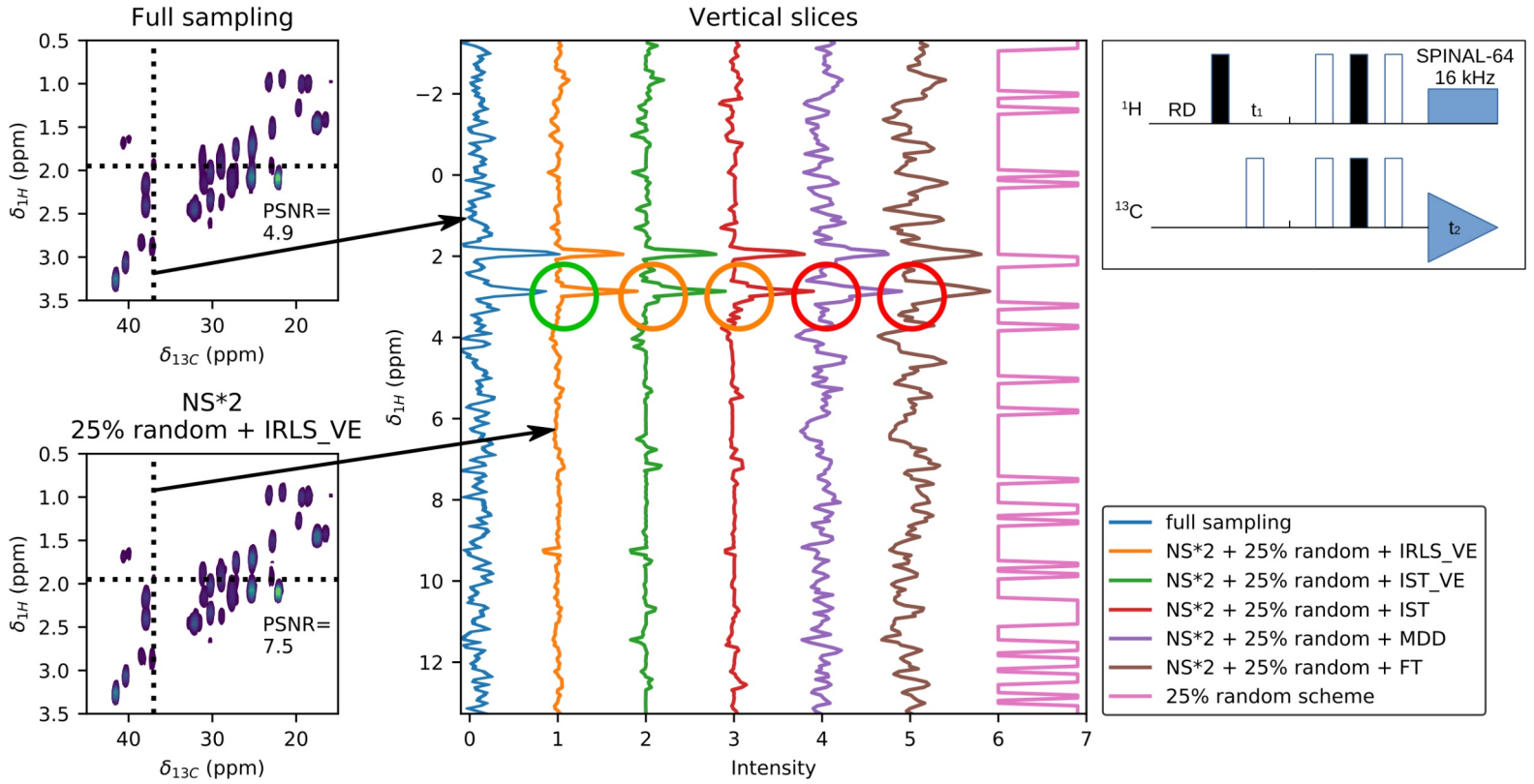


Figure IV.22: Influence of reconstruction algorithm on 2D {^{1}H}-^{13}C INEPT NMR spectra of gelatin with 0.5 M acetic acid. Top left: full sampling, $NS = 128$, $t_{exp} = 9.1$ h; bottom left: 25 % random NUS with IRLS_VE reconstruction, $NS = 256$, $t_{exp} = 4.6$ h; middle: 2D extracted ^{1}H slices on smallest peak, with 25 % random NUS and various reconstruction algorithms; top right: pulse sequence. Dotted horizontal black line: $PSNR_{max}$ measurement; dotted vertical black line: extracted ^{1}H slices; pink: sampling scheme; coloured circles: result quality. Parameters: $B_0 = 7.0$ T, 7 mm rotor, $MAS = 2.5$ kHz, $RD = 1$s, $TD(^{1}H) = 128$ complex slices, $SW(^{1}H) = 5$ kHz, $\nu_{decoupling}(^{1}H) = 16$ kHz SPINAL-64, cosine apodisation, processed with MDDnmr.

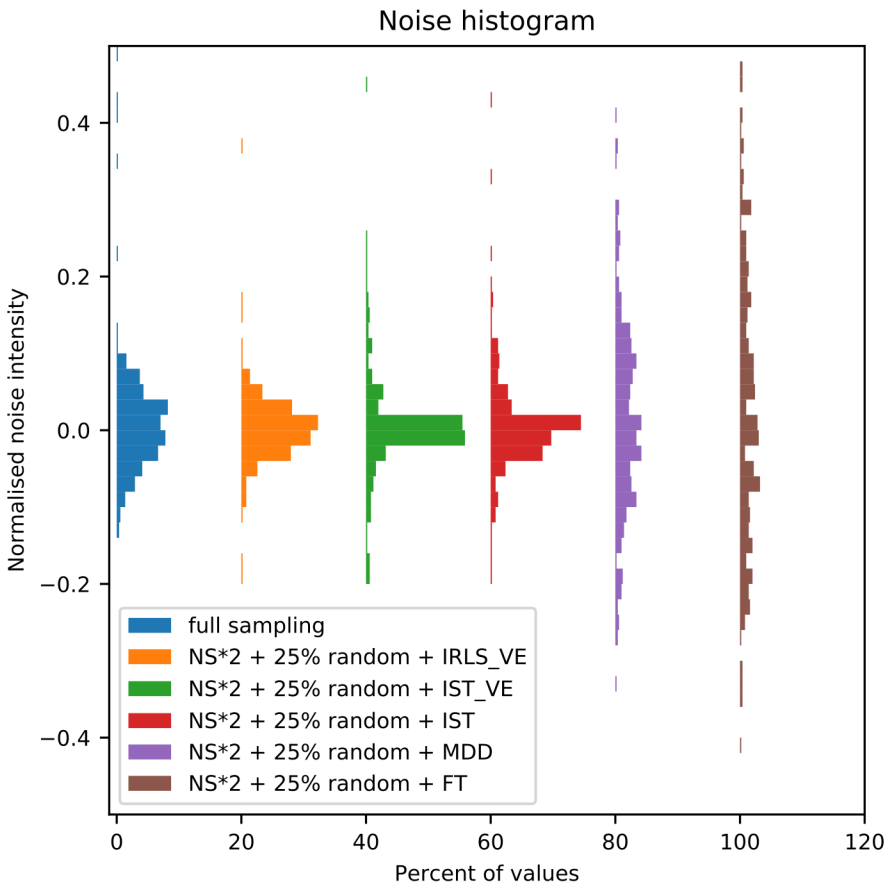


Figure IV.23: Noise histogram in ^{1}H slices extracted from Figure IV.22 right, depending on reconstruction algorithm. For each extracted slice, the maximum was normalised to 1 and 50 boxes were measured in the range [-0.5, 0.5].

E.3. TopSpin default sampling scheme

Figure IV.24 presents the application of NUS with Topspin default sampling scheme, on a rigid sample, namely a ^{13}C enriched carbonated HydroxyApatite (HAp), containing 4.8 wt % of carbonates (Section E.1 of Chapter I). HAp is the mineral phase of bone and can be substituted by many ions, such as carbonates. As hydrogens are isolated in HAp crystal structure, ^{1}H-^{1}H homonuclear dipolar interaction is small and narrow ^{1}H NMR lines are observed. 2D ^{1}H→X→^{1}H double cross-polarisation (CP) HETeronuclear CORrelation (HETCOR) (*157*). NUS spectra were acquired with X = {^{13}C, ^{31}P}. Unfortunately, corresponding US spectra were not acquired. After zero-augmented FT, strong artefacts were present (red circles on brown and green curves at right), which were only reduced with the best processing method chosen in previous section, namely IRLS_VE (orange circles on purple and orange curves). In that case, artefacts were not problematic with very high

$PSNR_{max}$ = 769 (bottom left), whereas the resulting spectrum was unusable with a lower $PSNR_{max}$ = 36.8 (top left), as vertical artefacts were as high as low intensity signals. Topspin random seed 54321, which is the default, has thus to be especially avoided. Similar warning has already been reported by Sidebottom (*158*).

E.4. Random *vs.* Poisson

In order to overcome default sampling scheme limitations, we compared random and Poisson sampling on a $\{^{1}H\}\rightarrow^{13}C\rightarrow^{31}P$ HETCOR of the aforementioned ^{13}C enriched carbonated HAp (Figure IV.25). US spectrum was resampled with MDDnmr (Section F.2) according to the chosen scheme. Random sampling (pink curve) and Poisson sampling (red curve), both with exponential weighting, looked very similar, but the latter was more regular, especially for the last points (top right). As a consequence, artefacts were visible on zero-augmented FT spectrum with random sampling (brown curve), whereas they were strongly reduced with Poisson sampling (green curve). This tendency was confirmed with IRLS_VE processing (purple and orange curves). Moreover, when focussing on the upper edge of the peak, no shoulder was visible with random sampling, whereas peak shape was almost correctly reconstructed with Poisson sampling and IRLS_VE (green circle). In such conditions, an improved $PSNR_{max}$ of 11.1 instead of 9.5 was obtained, *i.e.*, a gain of 1.2 in a quarter of the time, or 2.3 per unit of time. This also highlighted that noise can be reduced, simply by resampling the full spectrum, when decreasing sampling weight above 1.26 T_2^* (Section B.3). A 3D $^{1}H/^{13}C$ (labelled)/^{31}P NMR spectrum was attempted on this sample but was unusable due to ^{13}C probe detuning, which strongly decreased signal during acquisition.

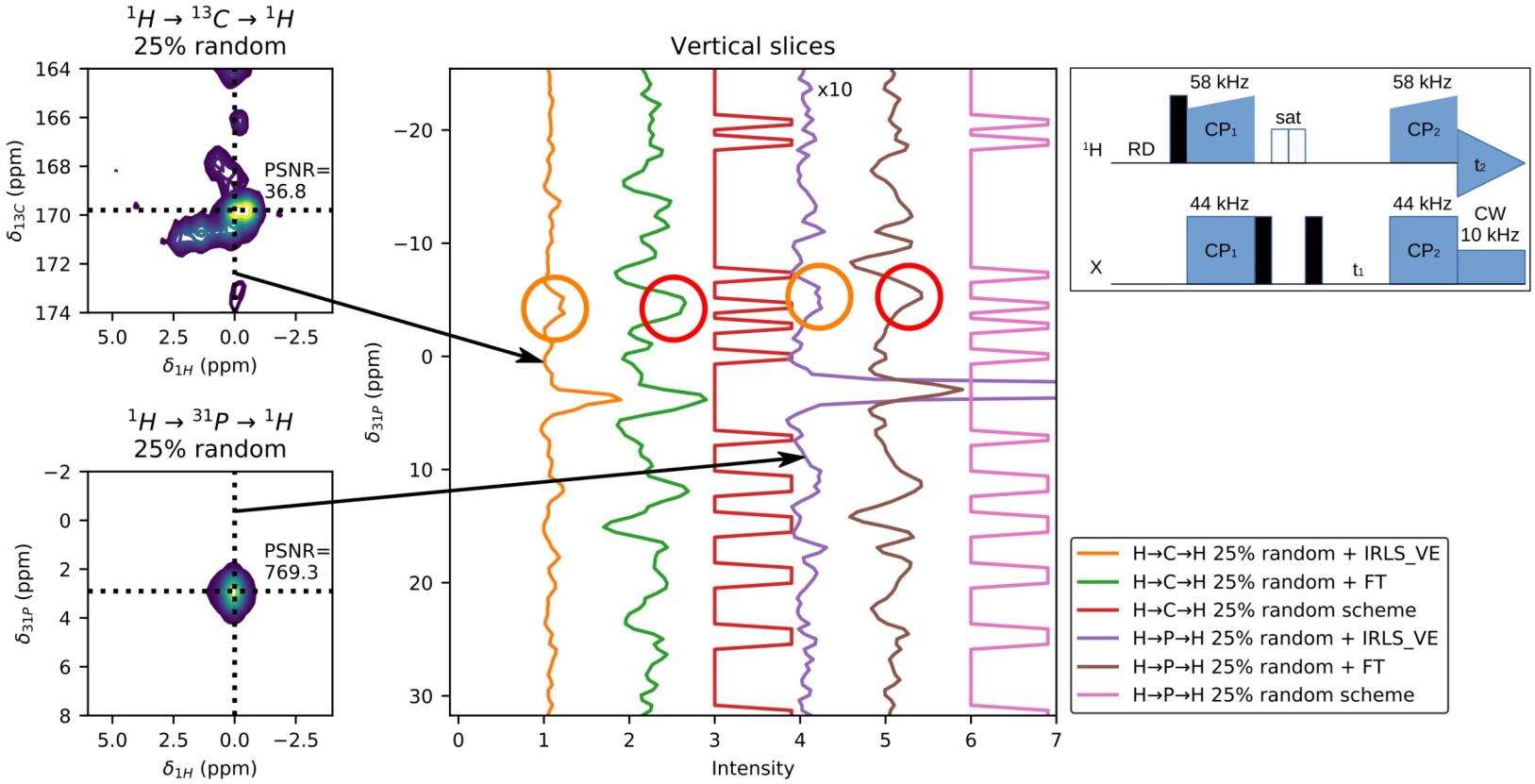


Figure IV.24: Default TopSpin sampling scheme on 2D ^{1}H→X→^{1}H HETCOR spectra of ^{13}C enriched carbonated HAp. Top left: X = ^{13}C, *NS* = 256, t_{exp} = 6.8 h, *SW(X)* = 3.5 kHz; bottom left: X = ^{31}P, *NS* = 128, t_{exp} = 3.4 h, *SW(X)* = 7 kHz; middle: 2D extracted X slices with zero-augmented FT and IRLS_VE processing; top right: pulse sequence. Dotted horizontal black line: $PSNR_{max}$ measurement; dotted vertical black line: extracted X slices; pink and red: sampling scheme; coloured circles: result quality. Parameters: 25 % random NUS, B_0 = 7.0 T, 4 mm rotor, *MAS* = 14 kHz, *RD* = 3s, *TD(X)* =64 complex slices, $t_{cp1} = t_{cp2}$ = 10 ms, $\nu_{CP}(^1H)$ = 58 kHz, $\nu_{CP}(X)$ = 44 kHz, $\nu_{decoupling}(X)$ = 10 kHz CW, cosine apodisation, processed with MDDnmr.

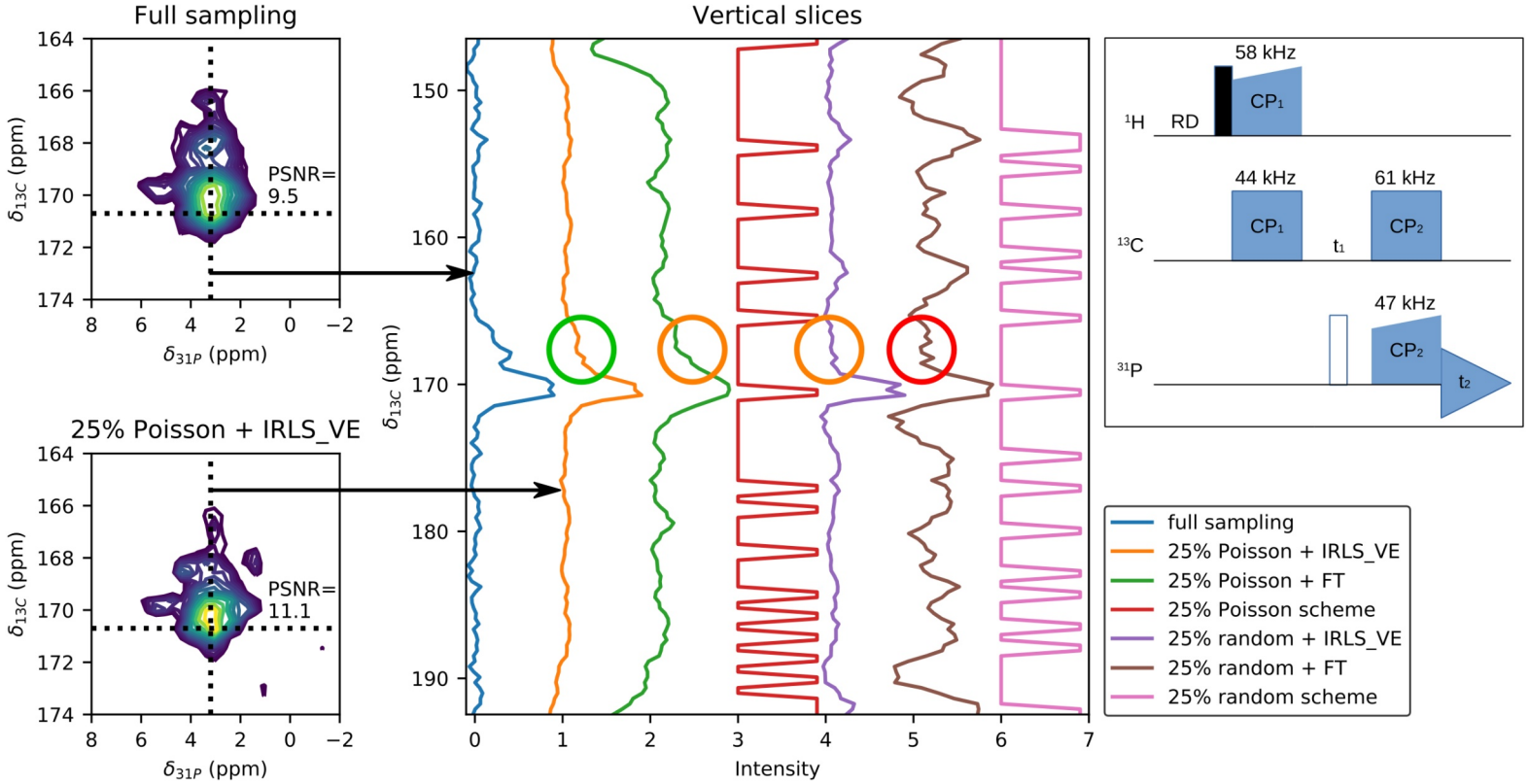


Figure IV.25: Random *vs.* Poisson sampling on 2D {^{1}H}→^{13}C→^{31}P HETCOR spectra of ^{13}C enriched carbonated HAp. Top left: full sampling, t_{exp} = 4.7 h; bottom left: resampled US spectrum with 25 % Poisson NUS, t_{exp} = 1.2 h, $TD(^{13}C)$ =16 complex slices; middle: 2D extracted ^{13}C slices with 25 % random or Poisson NUS; top right: pulse sequence. Dotted horizontal black line: $PSNR_{max}$ measurement; dotted vertical black line: extracted ^{13}C slices; pink and red: sampling scheme; coloured circles: result quality. Parameters: B_0 = 7.0 T, MAS = 14 kHz, RD = 3s, NS =176, $TD(^{13}C)$ =64 complex slices, $SW(^{13}C)$ =3.5 kHz, t_{cp1} = 10 ms, t_{cp2} = 20 ms, $\nu_{CP1}(^{1}H)$ = 58 kHz, $\nu_{CP1}(^{13}C)$ = 44 kHz, $\nu_{CP2}(^{13}C)$ = 61 kHz, $\nu_{CP2}(^{31}P)$ = 47 kHz, no decoupling, cosine apodisation, processed with MDDnmr.

E.5. Hybrid sampling

Up to now, we focussed on narrow peaks (45 Hz on gelatin for ^{1}H studied peak and 300 and 500 Hz on carbonated HAp for ^{13}C and ^{31}P, respectively) which are well suited for NUS. On the contrary, NUS cannot be applied to broad peaks, due to the lack of sparseness, which is a necessary condition. However, many spectra are a mixture of narrow and broad resonances, where hybrid sampling, combining NUS and US acquisition could be a solution to decrease acquisition time (*64*, *150*). We applied it to solid-state NMR of crystalline platelets of HAp surrounded by an amorphous calcium phosphate phase (Section E.2 of Chapter I). This structure is a model of the inorganic phase of bone (Figure I.12 of Chapter I). The same $\{^{1}H\}$-^{31}P HECTOR spectrum was reported on Figure IV.26 (top left), where the crystalline phase (HAp) was visible as a diamond at 0 ppm in ^{1}H dimension, whereas the amorphous phase (HPO_4^{2-} and adsorbed water) presented a broad peak from 3 to 17 ppm in ^{1}H dimension. As ^{31}P chemical shifts of both phases were almost identical near 3 ppm, the 2D spectrum was the only way to clearly discriminate amorphous from crystalline domains. However, a few indirect points were needed to sample broad part, whereas a much higher amount was necessary to digitalise the narrow peak, being time-consuming.

With 10 % Poisson sampling (pink line at right) and IRLS_VE processing, poor result was obtained (brown line), and the amorphous part was undistinguishable from noise. We gradually added complex uniform sampling on first indirect points. At 8 points US followed by 10 % NUS (purple line), noise was reduced but results were still unsatisfactory. At 16 points US followed by 10 % NUS (red line), the broad peak started to be distinguishable. At 32 points US followed by 10 % NUS (green and orange lines), the broad peak was correctly reconstructed. In particular, the shoulder at 13 ppm was clearly visible. The corresponding 2D spectrum is presented at bottom left. Very similar shapes were obtained between full and hybrid sampling. However, a slight decrease in $PSNR_{max}$ from 30.1 to 27.4 was measured, while this hybrid sampling corresponded to a quarter of the acquisition time. This was equivalent to a $PSNR_{max}$ gain per unit of time of 1.8. Hybrid NUS is thus a suitable option to decrease acquisition time of mixed narrow/broad spectra.

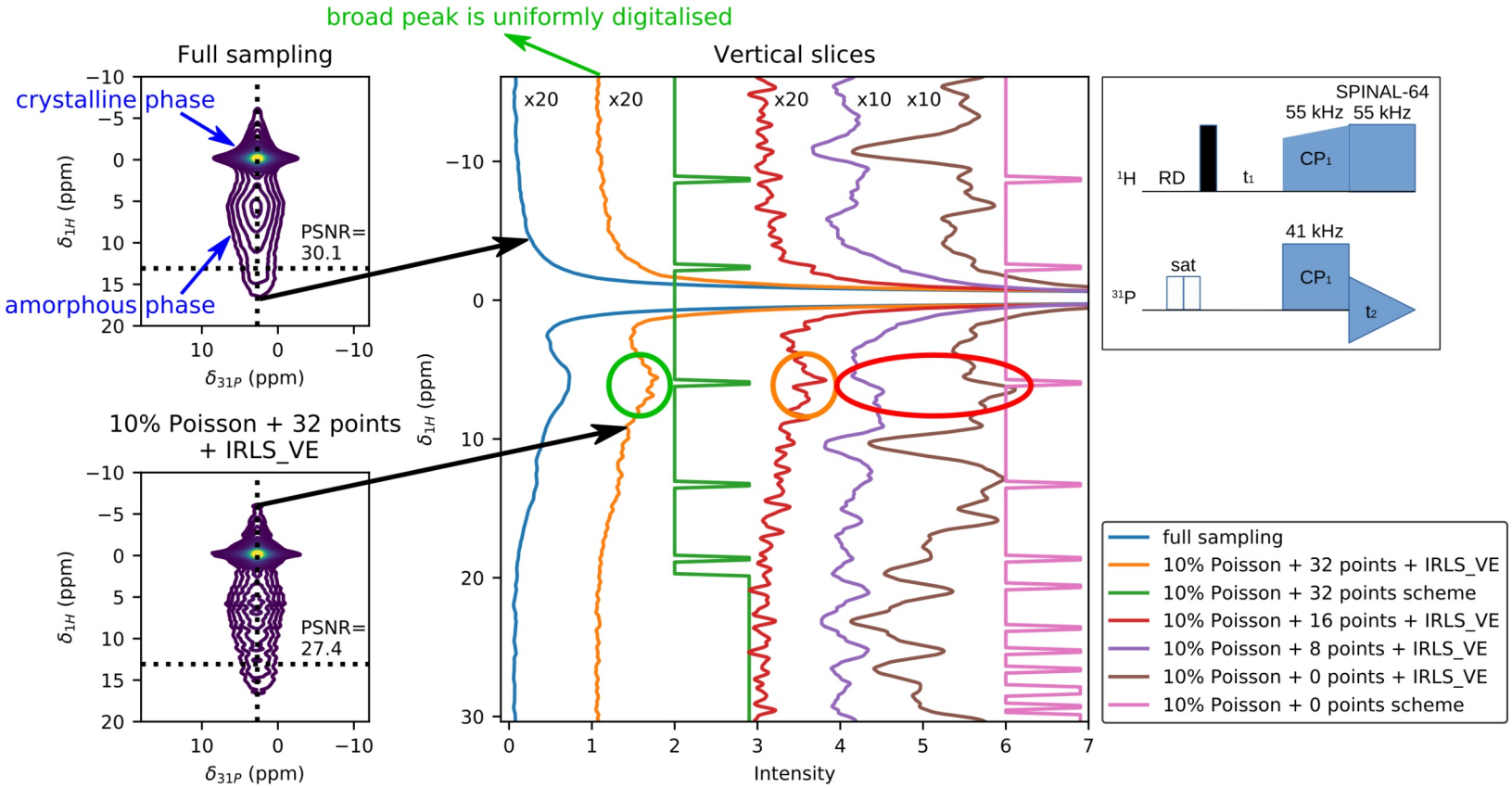


Figure IV.26: Hybrid NUS on 2D $\{^{1}H\}\rightarrow^{31}P$ HETCOR spectra of amorphous/crystalline HAp. Top left: full sampling, t_{exp} = 2.5 h; bottom left: resampled US spectrum with hybrid US/NUS, t_{exp} = 0.7 h; middle: 2D extracted ^{1}H slices with 10 % Poisson NUS and an increasing amount of US complex first points; top right: pulse sequence. Dotted horizontal black line: $PSNR_{max}$ measurement; dotted vertical black line: extracted ^{1}H slices; pink: sampling scheme; coloured ellipses: result quality. Parameters: B_0 = 7.0 T, MAS = 14 kHz, RD = 1s, NS =32, $TD(^{1}H)$ =140 complex slices, $SW(^{1}H)$ =14 kHz, t_{cp} = 1 ms, $\nu_{CP}(^{1}H)$ = 55 kHz, $\nu_{CP}(^{31}P)$ = 41 kHz, $\nu_{decoupling}(^{1}H)$ = 55 kHz SPINAL-64, 100 Hz exponential apodisation, processed with MDDnmr.

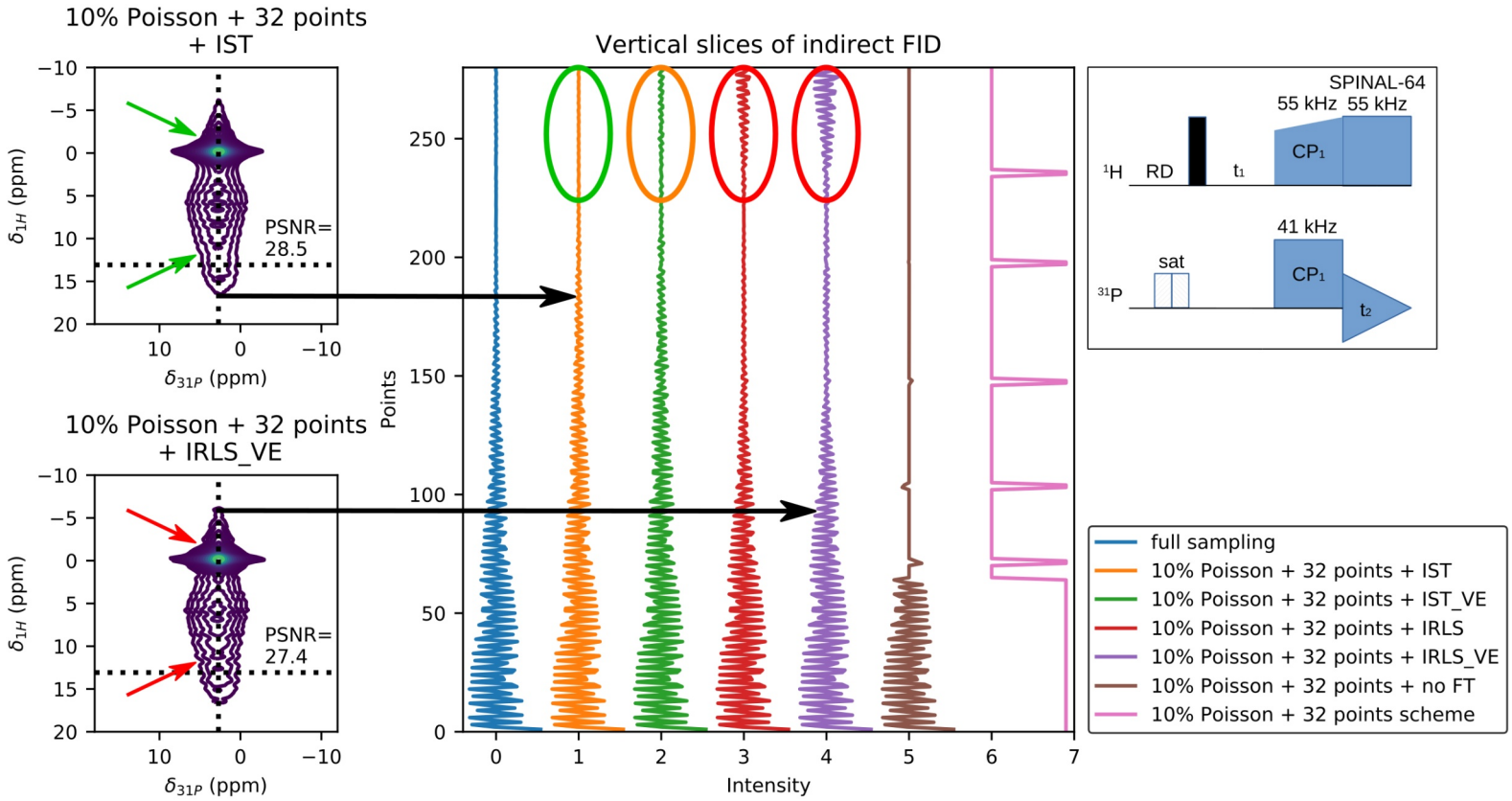


Figure IV.27: IST and IRLS reconstruction for hybrid NUS on 2D {^{1}H}→^{31}P HETCOR spectra of amorphous/crystalline HAp. Top left: IST of resampled US spectrum with 10 % Poisson NUS and 32 US complex first points; bottom left: IRLS_VE; middle: 2D extracted ^{1}H slices; top right: pulse sequence. Dotted horizontal black line: $PSNR_{max}$ measurement; dotted vertical black line: extracted ^{1}H slices; pink: sampling scheme; coloured ellipses: result quality. Parameters: B_0 = 7.0 T, MAS = 14 kHz, RD = 1s, NS =32, $TD(^{1}H)$ =140 complex slices, t_{exp} = 0.7 h, $SW(^{1}H)$ =14 kHz, t_{cp} = 1 ms, $\nu_{CP}(^{1}H)$ = 55 kHz, $\nu_{CP}(^{31}P)$ = 41 kHz, $\nu_{decoupling}(^{1}H)$ = 55 kHz SPINAL-64, 100 Hz exponential apodisation, processed with MDDnmr.

When looking in details to the resulting 2D spectrum (bottom left of Figure IV.26), the result looked ‘wavy’, as highlighted by red arrows on bottom left of Figure IV.27. Sampling scheme and raw FID are presented in pink and purple, respectively, at right of Figure IV.27. We noticed that an echo appeared on reconstructed FID with IRLS_VE (red ellipse on purple curve), whereas it was not present on full sampling FID (blue curve). This echo gradually diminished with IRLS, IST_VE and IST processing (red, orange and green ellipses, respectively). With IST without VE, a more regular 2D was obtained (top left and green arrows) and $PSNR_{max}$ sightly increased from 27.4 to 28.5. We can draw three observations: (i) the best processing for narrow lines, IRLS_VE, was not adequate for broad peaks and IST without VE had to be preferred; (ii) VE was detrimental for hybrid sampling reconstruction; (iii) IRLS induced a strong FID echo and some oscillations on spectra, similarly to forward linear prediction (Section B.4). The first observation highlighted the fact that there is no absolute rule for NUS reconstruction, and that the chosen algorithm has to be validated for the desired application. The second observation was explained by VE reflecting points at long time, which contained signal for narrow line but only noise for broad peak (Section D.5). The third observation was due to l_p-norm tending to zero for IRLS, which induced additional sparsity in reconstruction (Figure IV.16). As a consequence, IRLS minimised the number of peaks and discretised broad components.

In Subchapter E, we have highlighted that NUS has already been applied multiple times to solids and mesophases, but a detailed analysis on NUS influence on peak shape was missing. We first studied narrow lines on a mobile sample by solid-state MAS NMR. TopSpin default random sampling scheme was unusable, whereas Poisson scheme gave much better results. Finally, we implemented hybrid sampling to acquire a mixture of narrow and broad peaks. Surprisingly, best reconstruction algorithm differed for narrow lines (IRLS_VE) and for peaks widths mixture (IST without VE).

F. Practical aspects

In this subchapter, practical details will be provided for NUSscore, MDDnmr and TopSpin software, in order to optimise NUS acquisition, from sampling scheme to processing.

F.1. Sampling scheme optimisation with NUSscore

Efficient sampling scheme can be generated using NUSscore (Section C.5), available under Linux[1] and MacOS. Under Windows, a Linux Virtual Machine (VM) can be used, for instance with VirtualBox. Alternatively, NMRbox.org provides a suitable environment with online Linux VM (*159*). This shared computational platform for NMR at the Center for NMR Data Processing and Analysis (USA) provides multiple software (TopSpin, MDDnmr, NMRPipe, RNMRTK…), zero-configuration and version persistence. Community training such as workshops or tutorials are also available. NUSscore usage is presented on Figure IV.28, with the following parameters. For hybrid sampling (Section E.5), the resulting sampling file has to be edited with a text editor to complete first points.

- NUSscore parameters:
  - *TD*: number of slices in full sampling, including phase quadrature,
  - *Constant time*: to ignore *SW*, *Obs Freq* and $1/T_2^*$,
  - *NUS*: amount of points to sample,
  - *# of schemes*: number of schemes to generate and compare.

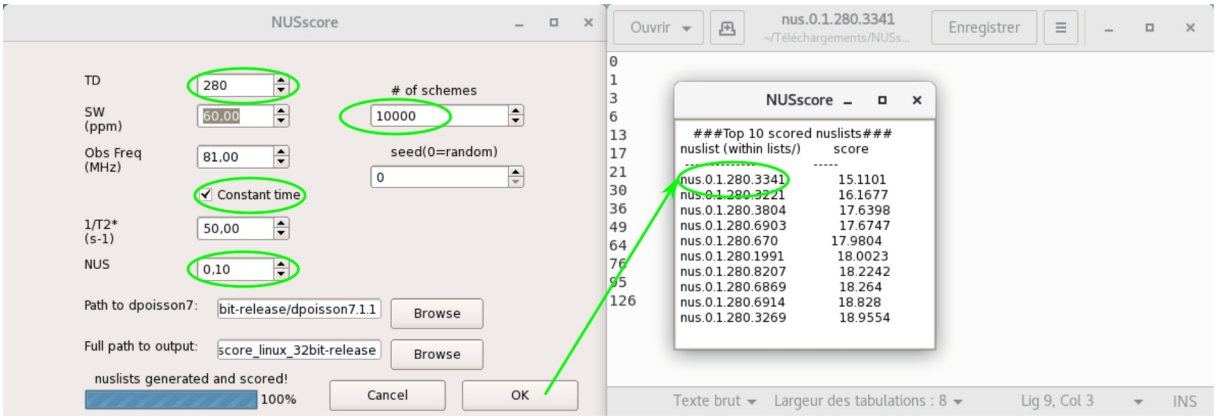


Figure IV.28: NUSscore usage (*69*). Left: parameters selection, right: scored results and best scheme at background.

F.2. Resampling of US spectrum with MDDnmr

Resampling means that some slices in indirect dimensions are discarded, while only a few of the original ones are kept, thus simulating NUS results from the US spectrum. This can be very useful to check the adequateness of a sampling scheme with given data. MDDnmr (*8*,

[1] Under Linux CentOS 7 64 bits, the following 32 bits libraries need to be installed, as root: 'yum install glib2.i686 fontconfig.i686 libXext.i686 libXrender.i686 libSM.i686 libjpeg-turbo.i686 libpng12.i686'

90) needs Linux environment with NMRPipe installed (*160*). Despite being a graphical software, some options are only accessible by text file edition and command line start[2]. This is the case for resampling. We suggest to graphically create MDDnmr files with a dummy processing, and to edit them afterwards. They are written in shell language, where an anti-slash '\' means that same line continues on next line, and pipe '|' means that output of first command is entered as input of second command. The following three files were used to resample data with hybrid sampling in Section E.5. Lines 6, 10, and 13 of proc.sh were especially useful for this purpose. Results can be viewed either with NMRDraw, which is part of NMRPipe, or with importation to Python thanks to nmrglue (*161*).

- MDDnmr files:
  - *proc.sh*: main script (Figure IV.29),
  - *fidSP.com*: processing of direct dimension (Figure IV.30),
  - *recFT.com*: processing of indirect dimensions (Figure IV.31).

```
#!/bin/tcsh
setenv FID ../33
setenv fidSP fidSP.com
setenv REC2FT recFT.com
setenv in_file nls.in
setenv selection_file nuslist_pois_sin_140_10_32
setenv FST_PNT_PPM 43
setenv ROISW 80
setenv proc_out FID_IST_VE_pois_sin_140_10_32.dat
setenv SPARSE              n
setenv NUS_TABLE_ORDER     '1'
setenv NUS_TABLE_OFFSET    0
setenv NUS_POINTS          37
setenv NI                  '140 1 '
setenv NIMAX               '140 1 '
setenv NDIM                2
setenv MDDTHREADS          10
setenv METHOD              CS
setenv CS_alg              IST
setenv CS_norm             1
setenv CS_lambda           1.0
setenv CS_niter            200
setenv CS_VE               n
setenv phase               '0 0'
setenv SRSIZE              0.5

mddnmr4pipeN.sh  1 2 3 4 5
set ecode=$? ;if ( $ecode ) exit($ecode)
```

Figure IV.29: *proc.sh* file in MDDnmr.

[2] The following commands are needed to start MDDnmr in command line, taking care to spaces: 'tcsh; set path = (. $path); proc.sh'

```
bruk2pipe -in ./ser -bad 0.0 -noswap -DMX -decim 960 -dspfvs 20 -grpdly 67.9839935302734 \
-xN              1024  -yN              280  \
-xT               512  -yT              140  \
-xMODE        Complex  -yMODE    States-TPPI  \
-xSW        20833.333  -ySW       14005.602  \
-xOBS         121.560  -yOBS        300.290  \
-xCAR           3.000  -yCAR          7.000  \
-xP0             33.6  -yP0             0.0  \
-xP1              0.0  -yP1             0.0  \
-xLAB             31P  -yLAB            1H  \
-ndim 2 -aq2D States                          \
| nmrPipe -fn EM -lb 100 -c 0.5                \
| nmrPipe -fn ZF -auto                         \
| nmrPipe -fn FT -auto                         \
| nmrPipe -fn PS -hdr                          \
# \
# for NUS table and PSF                        \
#| nmrPipe -fn SET -r 1.0 -i 0.0               \
#| nmrPipe -fn PS -p0 0 -p1 0 -di              \
# \
| nmrPipe -fn PS -p0 -21 -p1 0 -di            \
| nmrPipe -fn POLY -auto -ord 5                \
| nmrPipe -fn EXT -sw                          \
| pipe2xyz -z -out ft/data%03d.DAT -ov -nofs -verb
```

Figure IV.30: *fidSP.com* file in MDDnmr.

```
#!/bin/tcsh -f
echo '| ' in $0 $1
if( $#argv < 1 ) then
echo "Use: $0 <input pipe> <template for output spectrum>"
echo "nmrPipe processing of YZ dimensions after MDD reconstruction"
exit 1
endif
set ft4trec=$1
if( $#argv > 1 ) set proc_out=$2
if( ! -f $ft4trec ) then
ls $ft4trec
echo $0 failed
exit 2
endif
echo '|   Processing time domain MDD reconstruction '
echo
echo Processing Y dimensions
showhdr $ft4trec
cat $ft4trec                                  \
| nmrPipe  -fn TP -auto                       \
| nmrPipe  -fn EM -lb 100 -c 0.5              \
| nmrPipe  -fn ZF -auto                       \
| nmrPipe  -fn FT -auto                       \
| nmrPipe  -fn PS -hdr                        \
| nmrPipe  -fn PS -p0 0.0 -p1 -0.0 -di        \
| nmrPipe  -fn TP  -auto                      \
-ov -out $proc_out
echo $proc_out ready
exit
```

Figure IV.31: *recFT.com* file in MDDnmr.

F.3. Acquisition and processing with TopSpin

TopSpin is the software provided by Bruker, an NMR spectrometer manufacturer. NUS implementation depends on TopSpin version. In this chapter, we used TopSpin $\geq$ 3.5. Acquisition and processing relevant parameters are detailed hereafter.

- TopSpin version:
  - TopSpin < 3.0: NUS is not natively supported, manual modification of pulse programs and usage of external software for processing are required (*162*),
  - TopSpin ≥ 3.0: MDD, CS-IST and CS-IRLS algorithms are implemented,
  - TopSpin ≥ 3.5: improved CS algorithms with Virtual Echo (VE, Section D.5), algorithm defaulted to CS for 2D and to MDD for higher dimensions experiments,
  - TopSpin ≥ 3.5.5: 2D processing does not require a NUS licence, but only with CS-IST algorithm and without VE, all other options still need a licence,
  - TopSpin ≥ 3.5.6: CS-IST-VE is available for 4D experiments, to enable a faster computation than MDD.

- TopSpin acquisition parameters (Figure IV.32):
  - *FnTYPE*: select non-uniform sampling,
  - *FnMODE*: phase quadrature detection,
  - *TD*: number of slices in full sampling, including phase quadrature,
  - *NusPOINTS*: number of lines in *NUSLIST*,
  - *NUSLIST*: text file containing the NUS scheme,
  - Calculate button will overwrite *NUSLIST* with incorrect parameters,
  - Show button will display sampling scheme progress during acquisition.

- TopSpin processing parameters (Figure IV.33):
  - *Mdd_mod*: CS or MDD method,
  - *MddSRSIZE*: subregion size for NUS reconstruction, to slowly increase above 0.15 in case of error when processing without NUS licence.
- TopSpin hidden processing parameters:
  - *Mdd_CsALG*: IST or IRLS algorithm, defaulted to IST without NUS licence,
  - *Mdd_CsVE*: True or False, defaulted to False without NUS licence.

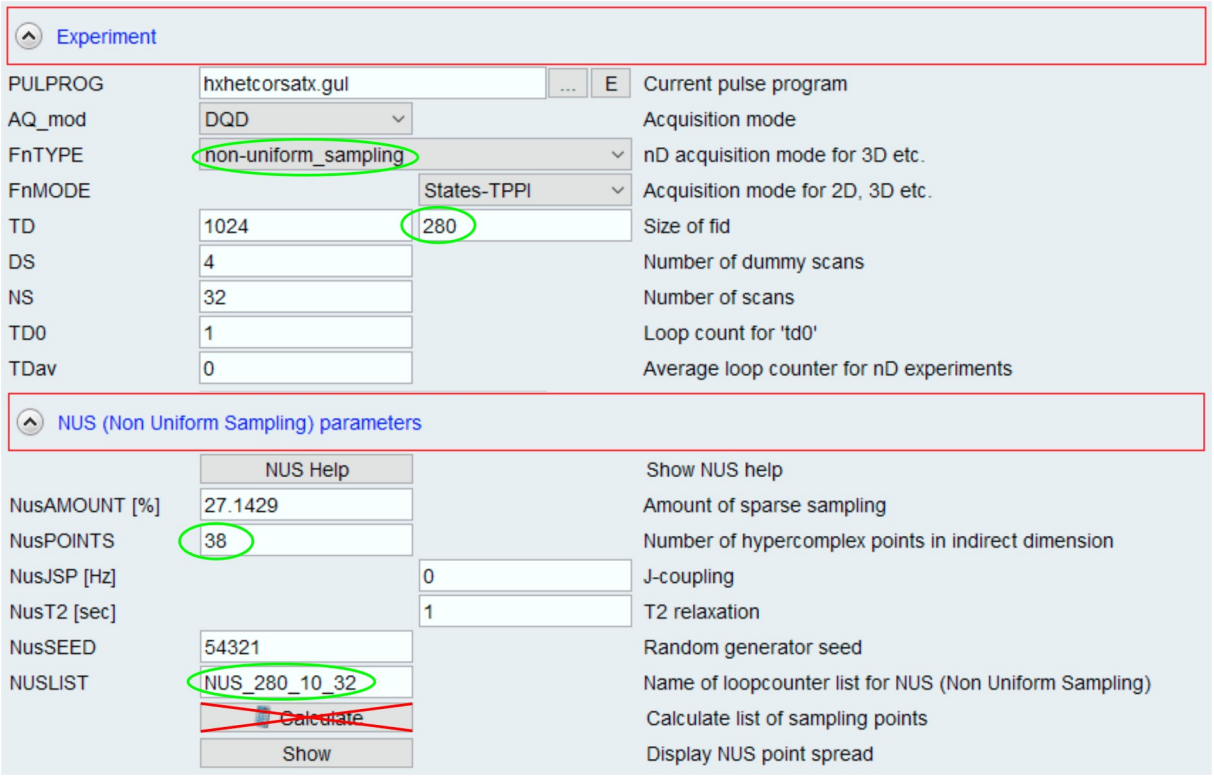


Figure IV.32: TopSpin acquisition parameters. Top: experiment part, bottom: NUS part.

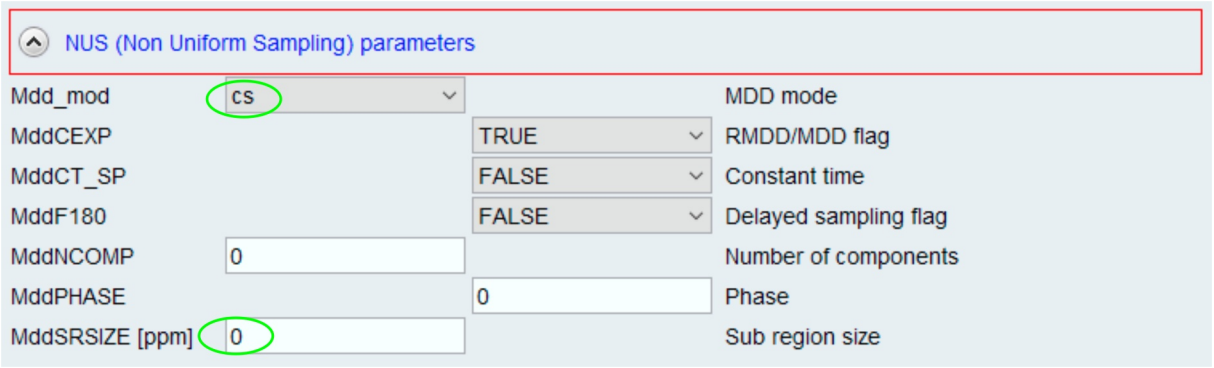


Figure IV.33: TopSpin processing parameters.

Subchapter F focussed on practical aspects to use NUS. We showed how to create an optimised sampling scheme with NUSscore, how to resample a full acquisition with MDDnmr and how to implement its usage in Topspin. The complete NUS chain is thus made available.

G. Chapter conclusion

In this chapter, we highlighted that standard nD NMR experiments suffer from long acquisition time, and of truncation of time domain in indirect dimensions (Subchapter B). It induces a low resolution if $AQ < 3\ T_2^*$, and even a low sensitivity if $AQ < 1.26\ T_2^*$. Apodisation and forward linear prediction can only partially compensate truncation. Sparse

sampling is a very powerful technique to circumvent these drawbacks, as it is possible to acquire up to the resolution limit in much less time (Subchapter C.1). In this family, radial sampling digitalise uniformly a few tilted planes, leaving sparse a high number of the (n-1)D indirect space. Peaks are projected perpendicularly to the acquired planes and are reconstructed using inverse Radon transform. Another sparse solution is to use NUS, which selects only a few points all over the (n-1)D indirect space, leaving all others blanks. Sampling scheme has a strong impact on processing artefacts, which can be reduced with Poisson sampling and scoring functions. In addition, sensitivity can be improved with a decreasing probability of sampling points above $AQ > 1.26\ T_2^*$. Processing is a non-trivial task, which has to be optimised (Subchapter D). Many algorithms are available, of which compressed sensing and maximum entropy are the most used. Each one has many variants with their advantages and drawbacks. Up to now, no clear consensus has emerged on the best reconstruction method, which may depend on applications. For each family of samples, calibration could be done by resampling of a US spectrum, before applying the NUS scheme and processing workflow to the entire family. For instance, after calibration on a substituted hydroxyapatite, other substitutions could be investigated with NUS. Application of NUS to solid-state NMR was depicted in Subchapter E. We proved that narrow lines in MAS (50-500 Hz) were suitable to use with NUS, but that default TopSpin sampling scheme has to be avoided. Poisson sampling gave much better results. In case of a mixture of narrow and broad lines, hybrid sampling with a mixed uniform/non-uniform scheme revealed to be a suitable option. Our results suggest different optimal reconstruction algorithms for NUS and hybrid sampling. Practical details were provided in Subchapter F in order to select a good sampling scheme, resample full data and apply it in everyday acquisitions and processing.

NUS is still a fast moving domain. New developments will probably be done in next years. In particular, it would be useful to compare all compressed sensing and maximum entropy algorithms in a fully automated way. It is especially necessary to quantify overfitting, which decreases reconstruction quality (*130*). Efficient resampling and reconstruction of direct spectrum could also be an efficient way to increase sensitivity, by decreasing noise introduced above $AQ > 1.26\ T_2^*$ (*146*). Finally, NUS and wavelets could be combined to optimise acquisition time of broad lines (*163*).

H. Chapter bibliography

1. M. Mobli, J. C. Hoch, *Prog. Nucl. Magn. Reson. Spectrosc.* **83**, 21–41 (2014).
2. E. D. Becker, *Anal. Chem.* **65**, 295A-302A (1993).
3. R. R. Ernst, W. A. Anderson, *Rev. Sci. Instrum.* **37**, 93–102 (1966).
4. J. Jeener, in *Ampere international summer school II* (Basko Polje, Yugoslavia, 1971).
5. M. P. Williamson, T. F. Havel, K. Wüthrich, *J. Mol. Biol.* **182**, 295–315 (1985).
6. W. Boucher, E. D. Laue, S. Campbell-Burk, P. J. Domaille, *J. Am. Chem. Soc.* **114**, 2262–2264 (1992).
7. L. Frydman, T. Scherf, A. Lupulescu, *Proc. Natl. Acad. Sci.* **99**, 15858–15862 (2002).
8. K. Kazimierczuk, V. Y. Orekhov, *Angew. Chem. Int. Ed.* **50**, 5556–5559 (2011).
9. P. Schanda, B. Brutscher, *J. Am. Chem. Soc.* **127**, 8014–8015 (2005).
10. Ē. Kupče, T. D. W. Claridge, *Angew. Chem. Int. Ed.* **56**, 11779–11783 (2017).
11. M. W. Maciejewski, H. Z. Qui, I. Rujan, M. Mobli, J. C. Hoch, *J. Magn. Reson.* **199**, 88–93 (2009).
12. V. Y. Orekhov, presented at the Biomolecular NMR: modern tools for data processing and interpretation dynamics (Gothenburg, Sweden, 2017).
13. J. C. Hoch, *Concepts Magn. Reson. Part A*. **46A**, e21459, 1–6 (2017).
14. W. P. Aue, E. Bartholdi, R. R. Ernst, *J. Chem. Phys.* **64**, 2229–2246 (1976).
15. G. Bodenhausen, R. Freeman, *J. Magn. Reson. 1969*. **28**, 471–476 (1977).
16. H. Fraga *et al.*, *ChemPhysChem*. **18**, 2697–2703 (2017).
17. G. Laurent, Master thesis (Sorbonne Université, Paris, France, 2012).
18. V. Jaravine, I. Ibraghimov, V. Y. Orekhov, *Nat. Methods*. **3**, 605–607 (2006).
19. D. Rovnyak, M. Sarcone, Z. Jiang, *Magn. Reson. Chem.* **49**, 483–491 (2011).
20. D. W. Tufts, R. Kumaresan, *Proc. IEEE*. **68**, 419–420 (1980).
21. Y. Zeng, J. Tang, C. A. Bush, J. R. Norris, *J. Magn. Reson. 1969*. **83**, 473–483 (1989).
22. P. Koehl, *Prog. Nucl. Magn. Reson. Spectrosc.* **34**, 257–299 (1999).
23. T. Szyperski, D. C. Yeh, D. K. Sukumaran, H. N. B. Moseley, G. T. Montelione, *Proc. Natl. Acad. Sci.* **99**, 8009–8014 (2002).
24. J. W. Yoon, S. Godsill, E. Kupče, R. Freeman, *Magn. Reson. Chem.* **44**, 197–209 (2006).
25. S. Hiller, C. Wasmer, G. Wider, K. Wüthrich, *J. Am. Chem. Soc.* **129**, 10823–10828 (2007).
26. K. Nagayama, P. Bachmann, K. Wuthrich, R. R. Ernst, *J. Magn. Reson. 1969*. **31**, 133–148 (1978).
27. G. Bodenhausen, R. R. Ernst, *J. Magn. Reson. 1969*. **45**, 367–373 (1981).
28. T. Szyperski, G. Wider, J. H. Bushweller, K. Wuethrich, *J. Am. Chem. Soc.* **115**, 9307–9308 (1993).
29. B. E. Coggins, R. A. Venters, P. Zhou, *Prog. Nucl. Magn. Reson. Spectrosc.* **57**, 381–419 (2010).
30. S. Hiller, F. Fiorito, K. Wüthrich, G. Wider, *Proc. Natl. Acad. Sci.* **102**, 10876–10881 (2005).
31. E. Kupče, R. Freeman, *J. Magn. Reson.* **173**, 317–321 (2005).
32. S. Kim, T. Szyperski, *J. Am. Chem. Soc.* **125**, 1385–1393 (2003).
33. E. Kupče, R. Freeman, *J. Biomol. NMR*. **27**, 383–387 (2003).
34. E. Kupče, R. Freeman, *Concepts Magn. Reson. Part A*. **22A**, 4–11 (2004).
35. R. N. Bracewell, *Aust. J. Phys.* **9**, 198–217 (1956).
36. G. N. Hounsfield, *Br. J. Radiol.* **46**, 1016–1022 (1973).

37. R. Freeman, Ē. Kupče, in *Novel sampling approaches in higher dimensional NMR*, M. Billeter, V. Y. Orekhov, Eds. (Springer Berlin Heidelberg, 2012), vol. 316 of *Topics in Current Chemistry*, pp. 1–20.
38. S. Hiller, G. Wider, in *Novel sampling approaches in higher dimensional NMR*, M. Billeter, V. Y. Orekhov, Eds. (Springer Berlin Heidelberg, 2012), vol. 316 of *Topics in Current Chemistry*, pp. 21–47.
39. J. C. J. Barna, E. D. Laue, M. R. Mayger, J. Skilling, S. J. P. Worrall, *J. Magn. Reson. 1969*. **73**, 69–77 (1987).
40. M. A. Davenport, J. N. Laska, J. R. Treichler, R. G. Baraniuk, *IEEE Trans. Signal Process.* **60**, 4628–4642 (2012).
41. V. Y. Orekhov, presented at the Workshop on novel reconstruction strategies in NMR and MRI (Goettingen, Germany, 2010), available at http://www.math.uni-goettingen.de/nmr-mri-workshop/orekhov.pdf.
42. H. Nyquist, *Trans. Am. Inst. Electr. Eng.* **47**, 617–644 (1928).
43. C. E. Shannon, *Proc. IRE*. **37**, 10–21 (1949).
44. E. J. Candès, J. Romberg, T. Tao, *IEEE Trans. Inf. Theory*. **52**, 489–509 (2006).
45. R. Dass, W. Koźmiński, K. Kazimierczuk, *Anal. Chem.* **87**, 1337–1343 (2015).
46. D. Donoho, J. Tanner, *Philos. Trans. R. Soc. Math. Phys. Eng. Sci.* **367**, 4273–4293 (2009).
47. J. C. Hoch, *J. Magn. Reson.* **283**, 117–123 (2017).
48. A. S. Stern, K.-B. Li, J. C. Hoch, *J. Am. Chem. Soc.* **124**, 1982–1993 (2002).
49. M. R. Palmer *et al.*, *J. Phys. Chem. B*. **119**, 6502–6515 (2015).
50. P. Schmieder, A. S. Stern, G. Wagner, J. C. Hoch, *J. Biomol. NMR*. **3**, 569–576 (1993).
51. A. D. Schuyler, M. W. Maciejewski, H. Arthanari, J. C. Hoch, *J. Biomol. NMR*. **50**, 247–262 (2011).
52. A. Kumar, S. C. Brown, M. E. Donlan, B. U. Meier, P. W. Jeffs, *J. Magn. Reson. 1969*. **95**, 1–9 (1991).
53. V. A. Jaravine, V. Yu. Orekhov, *J. Am. Chem. Soc.* **128**, 13421–13426 (2006).
54. M. Mayzel, J. Rosenlöw, L. Isaksson, V. Y. Orekhov, *J. Biomol. NMR*. **58**, 129–139 (2014).
55. M. Urbańczyk, A. Shchukina, D. Gołowicz, K. Kazimierczuk, *Magn. Reson. Chem.* **57**, 4–12 (2019).
56. A. D. Schuyler, M. W. Maciejewski, A. S. Stern, J. C. Hoch, *J. Magn. Reson.* **254**, 121–130 (2015).
57. K. Rossmann, *Radiology*. **93**, 257–272 (1969).
58. M. Lustig, D. Donoho, J. M. Pauly, *Magn. Reson. Med.* **58**, 1182–1195 (2007).
59. A. D. Schuyler, M. W. Maciejewski, A. S. Stern, J. C. Hoch, *J. Magn. Reson.* **227**, 20–24 (2013).
60. K. Kazimierczuk, M. Misiak, J. Stanek, A. Zawadzka-Kazimierczuk, W. Koźmiński, in *Novel sampling approaches in higher dimensional NMR*, M. Billeter, V. Y. Orekhov, Eds. (Springer Berlin Heidelberg, 2012), vol. 316 of *Topics in Current Chemistry*, pp. 79–124.
61. K. Aggarwal, M. A. Delsuc, *Magn. Reson. Chem.* **35**, 593–596 (1997).
62. M. W. Maciejewski, M. Mobli, A. D. Schuyler, A. S. Stern, J. C. Hoch, in *Novel sampling approaches in higher dimensional NMR*, M. Billeter, V. Y. Orekhov, Eds. (Springer Berlin Heidelberg, 2012), vol. 316 of *Topics in Current Chemistry*, pp. 49–77.
63. K. Kazimierczuk, A. Zawadzka, W. Koźmiński, *J. Magn. Reson.* **192**, 123–130 (2008).

64. S. G. Hyberts, H. Arthanari, G. Wagner, in *Novel sampling approaches in higher dimensional NMR*, M. Billeter, V. Y. Orekhov, Eds. (Springer Berlin Heidelberg, 2012), vol. 316 of *Topics in Current Chemistry*, pp. 125–148.
65. M. T. Eddy, D. Ruben, R. G. Griffin, J. Herzfeld, *J. Magn. Reson.* **214**, 296–301 (2012).
66. M. Mobli, *J. Magn. Reson.* **256**, 60–69 (2015).
67. D. L. Craft, R. E. Sonstrom, V. G. Rovnyak, D. Rovnyak, *J. Magn. Reson.* **288**, 109–121 (2018).
68. M. Mobli, T. M. Miljenović, *J. Magn. Reson.* **300**, 103–113 (2019).
69. P. C. Aoto, R. B. Fenwick, G. J. A. Kroon, P. E. Wright, *J. Magn. Reson.* **246**, 31–35 (2014).
70. S. G. Hyberts, K. Takeuchi, G. Wagner, *J. Am. Chem. Soc.* **132**, 2145–2147 (2010).
71. J. G. Ables, *Astron. Astrophys. Suppl. Ser.* **15**, 383–393 (1974).
72. M. Shatsky, R. J. Hall, S. E. Brenner, R. M. Glaeser, *J. Struct. Biol.* **166**, 67–78 (2009).
73. E. J. Candès, M. B. Wakin, *IEEE Signal Process. Mag.* **25**, 21–30 (2008).
74. G. Del Galdo, presented at the MacSeNet/SpaRTan Spring School on Sparse Representations and Compressed Sensing (Ilmenau, Germany, 2016), available at http://www.macsenet.eu/SpringSchool/Slides/DelGaldo_CompressedSensingTheory_Ilmenau_2016.ppsx.
75. H. Kudo, T. Suzuki, E. A. Rashed, *Quant. Imaging Med. Surg.* **3**, 147–161 (2013).
76. K. Kazimierczuk, V. Y. Orekhov, *J. Magn. Reson.* **223**, 1–10 (2012).
77. K. Kazimierczuk, O. Lafon, P. Lesot, *Analyst*. **139**, 2702–2713 (2014).
78. K. Kazimierczuk, in *eMagRes* (American Cancer Society, 2018), vol. 7 (1), pp. 1–8.
79. P. Lesot, K. Kazimierczuk, J. Trébosc, J.-P. Amoureux, O. Lafon, *Magn. Reson. Chem.* **53**, 927–939 (2015).
80. Y. Matsuki, M. T. Eddy, J. Herzfeld, *J. Am. Chem. Soc.* **131**, 4648–4656 (2009).
81. H. Barkhuijsen, R. De Beer, W. M. M. J. Bovee, J. H. N. Creyghton, D. Van Ormondt, *Magn. Reson. Med.* **2**, 86–89 (1985).
82. R. Kumaresan, D. Tufts, *IEEE Trans. Acoust. Speech Signal Process.* **30**, 833–840 (1982).
83. C. D. Eads, I. Noda, *J. Am. Chem. Soc.* **124**, 1111–1118 (2002).
84. R. G. Spencer, *Concepts Magn. Reson. Part A*. **36A**, 255–265 (2010).
85. D. Marion, *J. Biomol. NMR*. **32**, 141–150 (2005).
86. N. Pannetier, K. Houben, L. Blanchard, D. Marion, *J. Magn. Reson.* **186**, 142–149 (2007).
87. Y. Manassen, G. Navon, C. T. W. Moonen, *J. Magn. Reson. 1969*. **72**, 551–555 (1987).
88. G. McGeorge *et al.*, *J. Magn. Reson.* **129**, 134–144 (1997).
89. G. Mali, J.-P. Amoureux, V. Kaučič, *Phys. Chem. Chem. Phys.* **2**, 5737–5742 (2000).
90. V. Y. Orekhov, V. A. Jaravine, *Prog. Nucl. Magn. Reson. Spectrosc.* **59**, 271–292 (2011).
91. V. Y. Orekhov, I. Ibraghimov, M. Billeter, *J. Biomol. NMR*. **27**, 165–173 (2003).
92. X. Qu, M. Mayzel, J.-F. Cai, Z. Chen, V. Y. Orekhov, *Angew. Chem. Int. Ed.* **54**, 852–854 (2015).
93. M. Mayzel, K. Kazimierczuk, V. Y. Orekhov, *Chem. Commun.* **50**, 8947–8950 (2014).
94. E. T. Jaynes, *Phys. Rev.* **106**, 620–630 (1957).
95. J. P. Burg, in *37th Annual International SEG Meeting* (Oklahoma City, OK, USA, 1967).
96. R. Lacoss, *Geophysics*. **36**, 661–675 (1971).
97. S. J. Wernecke, L. R. D'Addario, *IEEE Trans. Comput.* **C–26**, 351–364 (1977).
98. S. F. Gull, G. J. Daniell, *Nature*. **272**, 686–690 (1978).

99. G. Minerbo, *Comput. Biol. Med.* **9**, 29–37 (1979).
100. D. Van Ormondt, K. Nederveen, *Chem. Phys. Lett.* **82**, 443–446 (1981).
101. S. Sibisi, *Nature*. **301**, 134–136 (1983).
102. M. A. Delsuc, in *Maximum Entropy and Bayesian Methods* (Springer, Dordrecht, 1989), *Fundamental Theories of Physics*, pp. 285–290.
103. J. C. Hoch, M. W. Maciejewski, M. Mobli, A. D. Schuyler, A. S. Stern, *Acc. Chem. Res.* **47**, 708–717 (2014).
104. P. J. Hore, *J. Magn. Reson. 1969*. **62**, 561–567 (1985).
105. D. L. Donoho, I. M. Johnstone, A. S. Stern, J. C. Hoch, *Proc. Natl. Acad. Sci.* **87**, 5066–5068 (1990).
106. S. Paramasivam *et al.*, *J. Phys. Chem. B*. **116**, 7416–7427 (2012).
107. S. G. Hyberts *et al.*, *J. Am. Chem. Soc.* **129**, 5108–5116 (2007).
108. D. L. Donoho, *IEEE Trans. Inf. Theory*. **52**, 1289–1306 (2006).
109. S. Foucart, H. Rauhut, *A Mathematical Introduction to Compressive Sensing* (Springer New York, New York, NY, 2013), *Applied and Numerical Harmonic Analysis*.
110. A. Shchukina, P. Kasprzak, R. Dass, M. Nowakowski, K. Kazimierczuk, *J. Biomol. NMR*. **68**, 79–98 (2017).
111. J. A. Högbom, *Astron. Astrophys. Suppl. Ser.* **15**, 417–426 (1974).
112. J. C. J. Barna, S. M. Tan, E. D. Lade, *J. Magn. Reson. 1969*. **78**, 327–332 (1988).
113. B. E. Coggins, P. Zhou, *J. Biomol. NMR*. **42**, 225–239 (2008).
114. S. G. Mallat, Z. Zhang, *IEEE Trans. Signal Process.* **41**, 3397–3415 (1993).
115. I. Daubechies, M. Defrise, C. D. Mol, *Commun. Pure Appl. Math.* **57**, 1413–1457 (2004).
116. A. S. Stern, D. L. Donoho, J. C. Hoch, *J. Magn. Reson.* **188**, 295–300 (2007).
117. I. Drori, *EURASIP J. Adv. Signal Process.* **2007**, 20248, 1–10 (2007).
118. V. Y. Orekhov, V. Jaravine, M. Mayzel, K. Kazimierczuk (2011), available at http://mddnmr.spektrino.com/.
119. S. G. Hyberts, A. G. Milbradt, A. B. Wagner, H. Arthanari, G. Wagner, *J. Biomol. Nmr*. **52**, 315–327 (2012).
120. S. Becker, J. Bobin, E. Candès, *SIAM J. Imaging Sci.* **4**, 1–39 (2011).
121. S. Sun, M. Gill, Y. Li, M. Huang, R. A. Byrd, *J. Biomol. NMR*. **62**, 105–117 (2015).
122. R. Otazo, E. Candès, D. K. Sodickson, *Magn. Reson. Med.* **73**, 1125–1136 (2015).
123. S. H. Amroabadi, Master thesis (Ryerson University, Toronto, Ontario, Canada, 2010), available at https://digital.library.ryerson.ca/islandora/object/RULA:1488/.
124. T. Blumensath, M. E. Davies, *J. Fourier Anal. Appl.* **14**, 629–654 (2008).
125. B. Natarajan, *SIAM J. Comput.* **24**, 227–234 (1995).
126. J. S. Toll, *Phys. Rev.* **104**, 1760–1770 (1956).
127. H. A. Kramers, in *Estratto dagli Atti del Congresso Internazionale de Fisici* (Como, Italy, 1927), vol. 2, pp. 545–557.
128. R. de L. Kronig, *JOSA*. **12**, 547–557 (1926).
129. R. R. Ernst, *J. Magn. Reson. 1969*. **1**, 7–26 (1969).
130. Q. Wu, B. E. Coggins, P. Zhou, *Nat. Commun.* **7**, 12281, 1–8 (2016).
131. M. A. Zambrello, M. W. Maciejewski, A. D. Schuyler, G. Weatherby, J. C. Hoch, *J. Magn. Reson.* **285**, 37–46 (2017).
132. N. S. Astrof, C. E. Lyon, R. G. Griffin, *J. Magn. Reson.* **152**, 303–307 (2001).
133. J. Leppert, B. Heise, O. Ohlenschläger, M. Görlach, R. Ramachandran, *J. Biomol. NMR*. **28**, 185–190 (2004).
134. W. T. Franks, H. S. Atreya, T. Szyperski, C. M. Rienstra, *J. Biomol. NMR*. **48**, 213–223 (2010).
135. M. Huber *et al.*, *ChemPhysChem*. **12**, 915–918 (2011).

136. Y. Matsuki, M. T. Eddy, R. G. Griffin, J. Herzfeld, *Angew. Chem. Int. Ed.* **49**, 9215–9218 (2010).
137. M. Strohmeier, D. M. Grant, *J. Magn. Reson.* **168**, 296–306 (2004).
138. P. R. Bodart, J. P. Amoureux, F. Taulelle, *Solid State Nucl. Magn. Reson.* **21**, 1–20 (2002).
139. B. Hu, P. Zhou, I. Noda, G.-Z. Zhao, *Anal. Chem.* **77**, 7534–7538 (2005).
140. C. Kaiser, J. J. Lopez, W. Bermel, C. Glaubitz, *Biochim. Biophys. Acta BBA - Biomembr.* **1768**, 3107–3115 (2007).
141. B. Hu, J.-P. Amoureux, J. Trébosc, M. Deschamps, G. Tricot, *J. Chem. Phys.* **128**, 134502, 1–8 (2008).
142. E. C. Lin, S. J. Opella, *J. Magn. Reson.* **239**, 57–60 (2014).
143. Y. Li *et al.*, *J. Magn. Reson.* **217**, 106–111 (2012).
144. Y. Qian, M. Shen, J.-P. Amoureux, I. Noda, B. Hu, *Solid State Nucl. Magn. Reson.* **59–60**, 31–33 (2014).
145. D. Rovnyak *et al.*, *J. Magn. Reson.* **161**, 43–55 (2003).
146. D. H. Jones, S. J. Opella, *J. Magn. Reson.* **179**, 105–113 (2006).
147. S. Sun *et al.*, *J. Phys. Chem. B*. **116**, 13585–13596 (2012).
148. C. L. Suiter *et al.*, *J. Biomol. NMR*. **59**, 57–73 (2014).
149. S. Xiang, V. Chevelkov, S. Becker, A. Lange, *J. Biomol. NMR*. **60**, 85–90 (2014).
150. E. C. Lin, S. J. Opella, *J. Magn. Reson.* **237**, 40–48 (2013).
151. G. Puentes, G. Waldherr, P. Neumann, G. Balasubramanian, J. Wrachtrup, *Sci. Rep.* **4**, 4677, 1–6 (2014).
152. R. Linser *et al.*, *J. Am. Chem. Soc.* **136**, 11002–11010 (2014).
153. J. C. Hoch, A. S. Stern (1996), available at http://rnmrtk.uchc.edu/rnmrtk/RNMRTK.html.
154. J. C. Hoch, A. S. Stern, *NMR data processing* (Wiley-Liss, New York, 1996).
155. S. G. Hyberts, S. A. Robson, G. Wagner, *J. Biomol. NMR*. **55**, 167–178 (2013).
156. A. Le Guennec, J.-N. Dumez, P. Giraudeau, S. Caldarelli, *Magn. Reson. Chem.* **53**, 913–920 (2015).
157. N. Baccile, G. Laurent, C. Bonhomme, P. Innocenzi, F. Babonneau, *Chem. Mater.* **19**, 1343–1354 (2007).
158. P. J. Sidebottom, *Magn. Reson. Chem.* **54**, 689–694 (2016).
159. M. W. Maciejewski *et al.*, *Biophys. J.* **112**, 1529–1534 (2017).
160. F. Delaglio *et al.*, *J. Biomol. NMR*. **6**, 277–293 (1995).
161. J. J. Helmus, C. P. Jaroniec, *J. Biomol. NMR*. **55**, 355–367 (2013).
162. S. Robson (2014), available at http://gwagner.med.harvard.edu/intranet/hmsIST/pulseprog.html.
163. M. Yon, PhD dissertation (Université d'Orléans, Orléans, France, 2017), available at http://theses.fr/2017ORLE2024.

Chapter V. Singular value decompostion

‘Donnez-moi cent paramètres et je vous ferai un éléphant. donnez-m'en un cent-unième et je lui ferai remuer la queue !’

Jacques Hadamard (1865-1963), mathematician

A. Chapter introduction

Previous chapters describe first how to increase signal as close as possible from the sample with microcoils, and second how to decrease acquisition time of the NMR spectra, either in direct dimension with CPMG or in indirect dimensions with NUS. Signal processing, presented in this chapter and the following one, intervenes as a third step to improve NMR sensitivity and resolution. Once raw data are obtained, this key process allows to analyse them and to get all the available information before publication.

Standard signal processing in NMR was presented in Subchapter C of Chapter I. In this chapter, we focussed on denoising with Singular Value Decomposition (SVD). Denoising of data corrupted by noise is a challenge to improve results and to deepen their analysis. Solid-state NMR is especially concerned due to its intrinsic low sensitivity. In addition to apodisation, multiple approaches have been developed so far, for instance maximum entropy (*1*) or Bayesian analysis (*2*, *3*). Both tend to convert the time-domain FID to a spectrum with minimum amount of noise and improved resolution. Maximum entropy was presented in Section D.3 of Chapter IV. Bayesian analysis is a statistical method, that alternates a prediction step and an update step (*4*). Jeol Delta NMR software uses it through Complete Reduction to Amplitude Frequency Table (CRAFT) (*5*). In a recent study, Matviychuk *et al.* compared Bayesian analysis to SVD. Despite a higher computational cost, the former was more sensitive than the latter (*6*). However, the authors used SVD as a black box, as they said themselves. They also proved that their Bayesian algorithm can take into account lineshape distortions, including asymmetric ones. Nevertheless, the starting hypothesis concerning peaks shape is the main drawback of Bayesian analysis and has to be defined carefully.

SVD is a totally different denoising tool on which we focussed in this chapter. In this paragraph, we present SVD denoising from a mathematical point of view. Further details will be found in Subsection B.3.a. Many datasets can be presented into a matrix, with rows corresponding to equations and columns to variables, describing a linear system of equations. If the matrix is horizontally elongated, there are less equations than variables and the matrix is said undetermined. On the contrary, if the matrix is vertically elongated, it is overdetermined. Moreover, the number of linearly independent lines is limited and is called the matrix rank. Undetermined matrices are ill-posed problems, which can be solved by SVD (*7*). This mathematical tool converts a matrix, whatever its shape, into its singular values, *i.e.*, the

individual inequivalent variables that it contains. Real spectroscopic signals are corrupted by noise, which is equivalent to add errors on the linear system of equations. Noise is converted to small singular values after SVD. By doing a low-rank approximation, it is possible to discriminate signals from noise, *i.e.*, to denoise data. A recent study compared denoising process on Hankel matrices and partially circulant matrices (*8*). An innovative approach was developed by Sheberstov *et al.* who used Signal/Artifact/Noise plots (SAN plots) to discriminate signals from noise (*9*). Interestingly, their tool gave very similar figures to the ones we obtained with singular values (Figure V.6c).

This work is a follow up of the investigations made by Pascal Man (2014) from the Institut des Matériaux de Paris Centre (IMPC, FR 2482). He developed a SVD program under Java using Nvidia Graphics Processing Units (GPU) (*10*). This led to an intense thought about SVD, which is described in the following two-parts article. Part (I) (*11*) is presented in Subchapter B of this chapter and focussed on concept and limits of this denoising technique. SVD will be applied to 2D experiments in Subchapter C. Part (II) of the article (*12*) will be devoted to SVD computation time (Subchapter B of Chapter VI). These two articles are presented as published, except typos, formatting, and additional footnotes.

B. Denoising applied to spectroscopies – part I: concept and limits, *Appl. Spectrosc. Rev.* 54, 602–630 (2019)

Some spectroscopies are intrinsically poorly sensitive, such as Nuclear Magnetic Resonance (NMR) and Raman spectroscopy. This drawback can be overcome by using Singular Value Decomposition (SVD) and low-rank approximation to denoise spectra and consequently increase sensitivity. However SVD limits have not been deeply investigated until now[1]. We applied SVD to NMR and Raman spectra and showed that best results were obtained with a square data set in time domain. Automatic thresholding was applied using Malinowski's indicators. 6 × 7380 noisy spectra with 41 signal-to-noise ratios were compared to their non-noisy counterparts, highlighting that SVD induces a systematic error for Gaussian peaks but faithfully reproduces shape of Lorentzian peaks, thus allowing quantification. Used carefully, SVD can decrease experimental time by a factor of 2.3 for spectroscopies. This study may help scientists to apply SVD to denoise spectra in a more efficient way, without falling into pitfalls.

[1] More precisely, experimental limits were much less studied than theoretical ones.

Keywords: spectroscopy, sensitivity, signal processing, Cadzow denoising, Singular Value Decomposition (SVD)

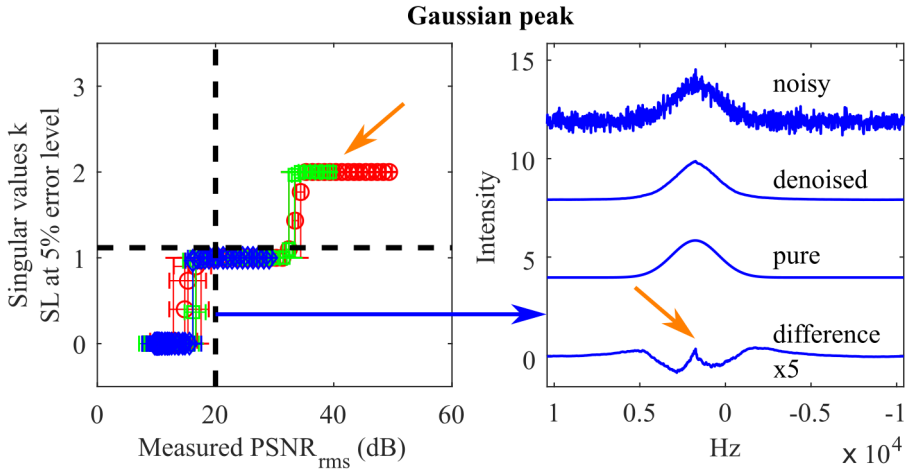


Figure V.1: Graphical abstract of (*11*).

B.1. Introduction

Spectroscopic techniques are of outermost importance in the field of materials analysis. Among them, Nuclear Magnetic Resonance (NMR) (*13*) and Raman (*14*) spectroscopies are very powerful local probes of the chemical structure. Especially, they allow to analyse liquid-state, solid-state, or even gas-state samples. While NMR informs about chemical and magnetic environment of atomic nuclei, Raman spectroscopy provides vibrational and rotational information on chemical entities. Unfortunately, these techniques suffer from a major drawback, namely their intrinsic low sensitivity. Although focused on NMR and Raman approaches, this work can be easily extended to other spectroscopies.

In the case of NMR, only one nucleus over 10^5 is detected under usual conditions (*15*). This is due to the low population difference between nuclear spin energy levels, which results from Boltzmann equilibrium. Many factors influence NMR sensitivity: magnetic field strength, sample volume, sample temperature, electronics temperature, radio-frequency coil quality factor and coil filling factor (*16*). When studying a solid-state sample, situation gets worse due to spectral line broadening, which results either from environment distribution or from relaxation (*17*). Indeed, NMR relevant anisotropic interactions, as chemical shift anisotropy, dipolar coupling and quadrupolar coupling, are no longer averaged to zero by fast and isotropic molecular motions, leading to spectra spreading over hundreds of ppm or a few

megahertz (*18*). During the last decades, numerous technical progress has allowed to increase sensitivity of solid-state NMR: Magic Angle Spinning (MAS) up to 110 kHz (*19–21*), Cross Polarisation (CP) (*22*), high performance heteronuclear decoupling (*23*) and NMR magnetic field strength increase (*24*). Additionally, new very sensitive techniques were developed, for instance micro-coils and microresonators adapted to MAS (*25*) or Dynamic Nuclear Polarisation (DNP) (*26–28*), with which a gain of up to 320 per unit of time was achieved (*29*).

In the case of Raman spectroscopy, only one photon over 10^6 is detected (*30*), due to its low scattering cross-section of $\approx 10^{-30}\,cm^2$ per molecule, to be compared with $\approx 10^{-20}\,cm^2$ for infrared absorption spectroscopy and $\approx 10^{-16}\,cm^2$ for fluorescence spectroscopy (*31*). Furthermore, some samples are fluorescent, thus hiding their Raman spectrum, or can be locally damaged by the laser light during analysis (*32*). A major advancement in Raman analysis has been achieved with the discovery of Surface-Enhanced Raman Scattering (SERS) (*33*, *34*) with an enhancement factor of 10^4-10^6 (*35*). However, this effect applies only to transition metals (*36*) and strongly depends on surface roughness (*37*). To circumvent these limitations, Raman equipment has been hyphenated with an atomic force microscope or a tunnel effect microscope, leading to Tip-Enhanced Raman Spectroscopy (TERS) (*38*, *39*). Another important progress has been achieved through the use of non-linear light sources and picosecond lasers (*40*, *41*). Thanks to such improvements, it is now possible to study artworks (*42*), to map a surface (*43*) and even to follow cure kinetics of an epoxy resin (*44*).

In parallel with these instrumental and methodological developments, mathematical and computer tools have become increasingly widespread in the field of data processing (*45*). In addition to Fourier transform revolution (*46*, *47*), other treatments have emerged in NMR spectroscopy and Magnetic Resonance Imaging (MRI). As examples, one may cite Hadamard transform (*48*), compressed sensing (*49*), non-uniform sampling (*50*), or quantitative signal reconstruction from multiple echoes (*51*). In Raman spectroscopy, pre-processing became of paramount importance to obtain quantitative measurements (*52*): it is now recommended to suppress fluorescence background (*53*), to correct cosmic ray spikes (*54*) and to normalise spectra (*55*).

Furthermore, a very important mathematical tool family concerns noise reduction (*56*). Indeed, the above mentioned sensitivity-enhanced spectroscopic equipments are costly and are not accessible to all laboratories, hence the need to denoise. Moreover, even with

hardware and methodological improvements, spectra can still be noisy, especially when studying amorphous materials (*57*, *58*), for which distribution of bond lengths and bond angles broadens signals and reduce sensitivity. In order to decrease experimental time or in case of unstable samples, signal processing is mandatory to get a reasonable Signal-to-Noise Ratio (SNR). The easiest way to perform noise reduction is smoothing. In NMR, apodisation, especially exponential multiplication, is used prior to Fourier transform (*59*). In Raman spectroscopy, a polynomial algorithm named Savitzky–Golay is preferred (*60*). Other options to reduce noise are maximum entropy (*61*), covariance matrix (*62*), Wiener's estimation (*63*), wavelet transform (*64*), uncoiled random QR denoising (*65*), and the method initially proposed by Tufts *et al.* (*66*) and generalised by Cadzow (*67*).

Singular Value Decomposition (SVD) is an important part of Cadzow's denoising algorithm and its related low-rank approximation (*68*). History of SVD can be found in (*69*). It has been discovered independently by Beltrami in 1873 (*70*), by Jordan in 1874 (*71*) and rediscovered by Lanczos in 1958 (*72*). It is preferred over eingenvalue decomposition, which is less precise, as demonstrated by Läuchli (*73*). Though quite old, Cadzow's procedure is currently a research domain of vivid interest (*74*), especially using sparse data, *i.e.*, partially empty matrix (*75*). SVD is widely used in several domains including acoustics (*76*), geophysics (*77*), air quality (*78*), electrocardiograms (*79*), image compression (*80*), video surveillance (*81*), MRI (*82*), data mining (*83*) and even on Facebook (*84*). Nevertheless, SVD is still not so commonly used in spectroscopies like NMR (*85–88*) and Raman spectroscopy (*89–92*), despite its use could significantly reduce experimental time and be of particular interest for the scientific community.

Recently, Man *et al.* developed a new SVD application for NMR (*10*). It was programmed under Java and two versions are currently available: one for processors (*93*) and the other one for Nvidia graphic cards (*94*) using CUDA (*95*). Indeed, graphic cards allow very efficient parallel computations. However, the limits of this approach are not clear: (i) which matrix shape should be preferred? (ii) what is the minimal experimental SNR? (iii) are denoised spectra quantitative? (iv) is it suitable for other spectroscopies?

Following a previous communication (*96*), we tried to address these questions. This work is divided into two parts. In this first contribution (I), we focus on SVD concept and limits. Experimental details will be provided in Subsection B.2. Theoretical background on SVD and low-rank approximation concepts is developed in Subsection B.3.a. Hankel and

Toeplitz matrices are explored in Subsection B.3.b. SNR definitions are given in Subsection B.3.c. Subsection B.4.a is devoted to experimental results by applying SVD to solid-state NMR and Raman spectroscopies. The influence of both matrix shape and thresholding are studied in Subsections B.4.b and B.4.c, respectively. Time and frequency denoising are compared in Subsection B.4.d. The minimum SNR needed to have accurate results is investigated in Subsection B.4.e. The impact of SVD on peak shape is considered in Subsection B.4.f. Finally, denoising on a real NMR spectrum is analysed in Subsection B.4.g.

In a second part (II) (*12*) we will benchmark SVD using Java, Matlab and Python, on various processors and nvidia graphic cards ranging over 10 and 6 years, respectively. We will try to optimise algorithms, software libraries and hardware capabilities to achieve the fastest possible denoising computation.

B.2. Materials and methods

B.2.a. Synthesis of the 50:50 MTEOS:TEOS sample

This sample is representative of typical materials obtained by sol-gel chemistry (*97–99*). This soft chemistry synthetic approach is a suitable route to design hybrid materials that contain both organic (methyltriethoxysilane, MTEOS, T species) and inorganic functions (tetraethylorthosilicate, TEOS, Q species), combining for instance hydrophobicity and high mechanical stability (*100*). T and Q stand for the number of oxygen on each silicon, namely Tri (3) and Quadri (4), for MTEOS and TEOS, respectively. The letter is associated with a superscript indicating the number of condensed Si-O-Si bridges. It is important to quantify the ratio T/Q and the condensation degree, mainly by ^{29}Si MAS NMR, in order to properly characterise such hybrid materials. This nucleus suffers from a low natural abundance of 4.7 % and an intermediate resonating frequency at 1/5th of ^{1}H one, both lowering SNR. It is thus current to average noise over one night or one weekend for a single spectrum. Using denoising is an interesting approach to decrease acquisition time.

Every chemical was used as received with no further purification. The solution was prepared by adding 10.18 g of MTEOS (98 %, Alfa Aesar; M = 178.30 g.mol^{-1}, 57.1 mmol) and 11.89 g of TEOS (>99 %, Aldrich; M = 208.33 g.mol^{-1}, 57.1 mmol), 29 mL of milliQ water and 50 mg of a 37 % w/w aqueous solution of hydrochloric acid (HCl, VWR; M = 36.46 g.mol^{-1}). The solution was stirred at room temperature for at least one hour at

500 rpm to ensure hydrolysis of the precursors. Controlled condensation occurred during spray drying of the sample, performed using a mini spray dryer B-290 (BUCHI) fitted with an atomiser (nozzle tip diameter = 0.7 mm) and a peristaltic pump. The temperatures at the inlet and outlet of the spray dryer were fixed at 220 °C and within the range of 95-120 °C, respectively. Polydisperse spherical particles of hybrid organic/inorganic amorphous silica (characteristic size: 1-10 μm) were obtained.

B.2.b. Solid-state NMR experiments

^{29}Si solid-state NMR experiments were performed on a Bruker Avance III spectrometer operating at 300.29 MHz for ^{1}H and 59.65 MHz for ^{29}Si with 4 mm zirconia rotors spun at 14 kHz (MAS broadband dual probe). Unless otherwise stated, CP was used with a contact time of 5 ms, a relaxation delay of 1 s, $NS = 2048$ scans. Low-power ^{1}H SPINAL-64 decoupling ($\nu_{1H} = 2.4$ kHz) (*101*) was checked to be sufficient and was used to protect the probe as the total acquisition time of 197 ms[2] was too long for high-power decoupling. During acquisition, 4096 complex points were acquired with 24 Carr-Purcell-Meiboom-Gill (CPMG) echoes (*51*) and a full echo delay of 8 ms.

SVD was applied on Free Induction Decay (FID, time domain) after removal of the first 68 points corresponding to oversampled digitisation. Zero-filling to 16384 complex points and cosine multiplication were applied after SVD. This apodisation limits both signal truncation and broadening effects. One may note that SVD was not directly applied to spectra (SPC, frequency domain) because zero-filling increases matrix size and thus computation time.

B.2.c. Simulation of kinetics studied by Raman spectroscopy under Python

2000 Raman spectra of 2000 points each were calculated *in silico* using four Gaussian lines at 450, 510, 750 and 900 cm^{-1}, respectively. Full Widths at Half Maximum (FWHM) ranged from 118 to 212 cm^{-1}. Such high FWHM are typical of amorphous materials like glasses (*57*). In order to reflect a kinetic evolution, the amplitude of peaks at 510 and 750 cm^{-1} were linearly decreased across the series while amplitude of peaks at 450 and 900 cm^{-1} were linearly increased. This series of spectra may mimic ageing of a material for instance. Homoscedastic white Gaussian noise was added on each spectrum. SVD was applied using

[2] A long acquisition time was needed to acquire as many CPMG echoes as possible.

Principal Component Analysis (PCA) function from Python Scikit-learn package (*102*). Computation took only a few seconds under Python Anaconda 3.5. The source code is available in file Figure_I.4a.py of (*103*).

B.2.d. Simulation of NMR spectra with known noise under Matlab

NMR complex FID were simulated *in silico* under Matlab[3] (The MathWorks, Inc., Natick, MA, USA) with a complex exponential at the expected frequency v and either an exponential decay (Equation V.1) or a Gaussian decay (Equation V.2), leading to a Lorentzian or a Gaussian peak on spectra after Fourier transform, respectively.

$$y = e^{i2\pi\nu t} \cdot \mathrm{e}^{\frac{-t}{T_2}} \quad \text{V.1}$$

$$y = e^{i2\pi\nu t} \cdot \mathrm{e}^{\frac{-t^2}{2\sigma^2}} \quad \text{V.2}$$

While the former is typical of a relaxation-driven shape, the latter highlights a distribution of chemical environments (*104*) or more complex relaxation phenomena, *e.g.*, strong dipoar coupling. To obtain a Gaussian peak with the same FWHM as a Lorentzian peak, the shape is defined according to Equation V.3.

$$\sigma = T_2\sqrt{2\ln(2)} \quad \text{V.3}$$

7380 NMR FID were simulated, grouped as follows:

- 2 shapes for decay: exponential and Gaussian;
- 3 T_2 values of 10, 1.0 and 0.10 ms corresponding to *narrow* (32 Hz), *intermediate* (320 Hz) and *broad* peaks (3200 Hz), respectively, which are typical values obtained by ^{13}C, ^{29}Si or ^{31}P solid-state NMR, for various types of crystalline or amorphous materials;
- 41 levels of homoscedastic white Gaussian noise ranging from -20 dB to +20 dB;
- 30 random noise patterns at the same noise level.

Additionally, each data set was repeated 6 times with different processing parameters:

- 2 with truncation or not, at 5 T_2, time above which signal is almost no longer existent;
- 3 Significance Level (*SL*, see Subsection B.4.c) for SVD automatic thresholding at error level of 5, 7.5 or 10 %.

[3] Raman and NMR spectra were not simulated by the same article author, which explains the different software used.

Each FID was composed of 1024 points for a duration of 49 ms. SVD was applied before zero-filling (if truncation was applied) and Fourier transform. As peak was not at the middle of the spectrum, *signal region* was defined as the 512 points centred at peak frequency. *Noise region* corresponded to the other 512 points. Baseline zero-order offset was preliminary corrected by subtracting the mean value of noise region. This step was essential to avoid spectrum aliasing due to Fourier transform, especially for broad peaks. The source codes of SVD automatic thresholding and FID simulations are available in files sfa.m and Figure_I.7_I.8_I.S2.m of (*103*), respectively. Computation of the full set of $6 \times 7380 = 44280$ spectra took 30 minutes with an overclocked Intel Core i5 4670K @ 4.4 GHz processor with Matlab R2016b.

B.3. Theoretical background

In this subsection, SVD and low-rank approximation are first developed. Hankel and Toeplitz matrices are then presented. Finally, SNR is defined.

B.3.a. SVD and low-rank approximation

SVD is a mathematical tool used to decompose a matrix X with m rows and n columns, whatever its size or shape, into the product of three other matrices U, Σ and V^T (Equation V.4).

$$X = U \cdot \Sigma \cdot V^T \qquad \text{V.4}$$

This is illustrated in Figure V.2 by orange hatched rectangles. U and V are unitary square matrices, of size $m \times m$ and $n \times n$, respectively. If complex numbers are used, V^T, the transpose of matrix V, is replaced by V^*, its conjugate transpose. SVD can indifferently be applied on real or complex matrices, the only difference being a double computation time for complex matrices (see part (II) of this work (*12*)). The central matrix Σ has the same shape as the original matrix X. Nevertheless, it has values only on its main diagonal (green rectangle), sorted by amplitude. These diagonal entries are called singular values and are the non-negative square roots of the eigenvalues of X^TX or XX^T (*105*). One can notice that the more elongated is matrix X, the less singular values it has.

Using low-rank k, matrix X can be approximated to X_k according to Equation V.5, where U_k, Σ_k and V_k^T are the matrices U, Σ and V^T truncated at k values, represented as blue filled rectangles in Figure V.2.

$$X \approx X_k = U_k \Sigma_k V_k^T \qquad \text{V.5}$$

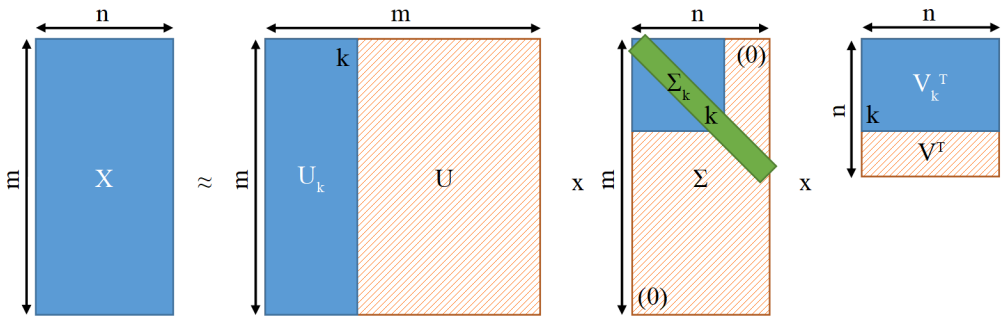


Figure V.2: Singular value decomposition (orange hatched rectangles) and low-rank approximation (blue filled rectangles). The *k* singular values are sorted by size (in green).

This process is really useful to compress large data sets (*80*). In the case of noise-containing data, true signals correspond to low-*k* values while noise signals are related to high-*k* values. Thus, selecting the correct *k*-limit allows to keep all true signals while rejecting noise, including t_1-noise (*85*). Additionally, when a baseline distortion is present, its intense signal corresponds to the first singular value which may be removed (*106*). The following subsection describes how to apply SVD on one-dimensional (1D) data.

B.3.b. Hankel and Toeplitz matrices

As stated above, SVD can only be applied to matrices. However, 1D data form only a (complex) row or column but not a matrix. In such a case, a transformation step is required. This can be performed thanks to Hankel matrix, or similarly to Toeplitz matrix. The former is defined by its first row and its last column. All anti-diagonal values are filled identically to the first ones (Figure V.3). Toeplitz matrix is defined by its first column and its first row, all diagonal values being identical (*107*). Circulant matrices are a special case of Hankel or Toeplitz matrices, where every row of the matrix is a cyclic shift of the row above (*108*). However, in general case, these are semi-circulant matrices, with one value being replaced from one row to the next one.

Hankel and Toeplitz matrices are not necessarily square, and are the vertical reflection of each other. From a programming point of view, extracting anti-diagonals needs a matrix vertical reflection. Although this step is computationally inexpensive, Toeplitz matrices were preferred for the sake of simplicity when using Python and Matlab. The Java application developed by Man *et al.* (*10*, *93*, *94*) used a Hankel matrix. It should be noted that after denoising, (anti-)diagonals values are no longer identical and averaging is needed as

highlighted by Hansen and Jensen (*76*). They indeed stated that 'simply extract (and transpose) an arbitrary row of the matrix […] lacks a solid theoretical justification'.

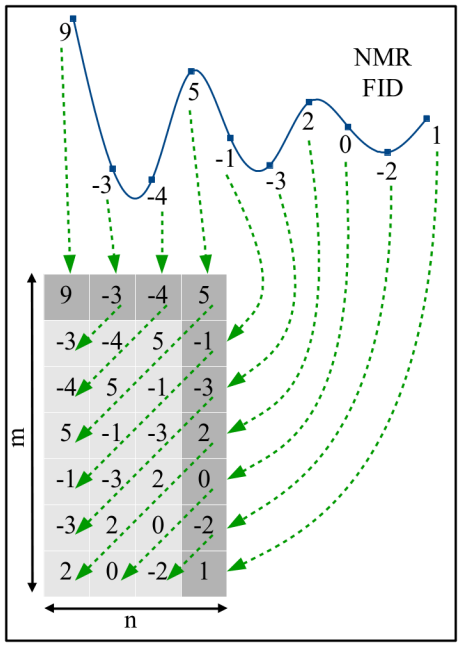


Figure V.3: Hankel matrix applied to a NMR FID. Each point of the FID defines either one point of the first row or of the last column of the Hankel matrix. All anti-diagonal points are filled identically to the first point.

B.3.c. Signal-to-noise ratio

Two definitions of *SNR* are used. The first one corresponds to the mathematical formula (Equation V.6), used in electronics, where y_i are the individual values while σ_{signal} and σ_{noise} are the standard deviations for signal region and noise region, respectively.

$$SNR = \frac{\sigma^2_{signal}}{\sigma^2_{noise}} = \frac{\left(\frac{1}{n-1}\sum_i (y_i)^2\right)_{signal}}{\left(\frac{1}{n-1}\sum_i (y_i)^2\right)_{noise}} \qquad \text{V.6}$$

Normalisation over *(n-1)* points is preferred to avoid a bias in standard deviation (*109*). This formula is valid only when signal can be measured without any noise, so-called *pure* signal. However, the only observable parameter on a real noisy signal is the signal-plus-noise-to-noise ratio (*SNNR*) (*110*) defined by replacing σ_{signal} by $\sigma_{signal+noise}$ in Equation V.6. *SNR* can then be deducted from *SNNR*, following Equation V.7.

$$SNR = \frac{\sigma^2_{signal+noise} - \sigma^2_{noise}}{\sigma^2_{noise}} = SNNR - 1 \qquad \text{V.7}$$

The other possible definition is the analytical chemistry formula (Equation V.8), where *PSNR* is the *SNR* based on peak amplitude (*110*).

$$PSNR_{max} = \frac{H_{signal}}{h_{noise\_max}} = \frac{H_{signal}}{h_{noise\_peak\_peak}/2} = \frac{2\,max(y_i)_{signal}}{(max(y_i) - min(y_i))_{noise}} \qquad \text{V.8}$$

Signal height (H_{signal}) is measured from maximum of peak to mean of noise, whereas noise is measured on a region of 20 times the signal FWHM (*111*). Additionally, variants exist, depending on the way noise is defined, either as maximum noise (h_{noise_max}), mean noise (h_{noise_mean}) or Root Mean Square (RMS) noise (h_{noise_rms}) (*112*). Following this nomenclature, *SNR* used in NMR (*112*) and Raman spectroscopy (*63*) should rather be called $PSNR_{rms}$ (Equation V.9)[4].

$$PSNR_{rms} = \frac{H_{signal}}{h_{noise\_rms}} = \frac{H_{signal}}{\sigma_{noise}} = \frac{max(y_i)_{signal}}{\sqrt{\left(\frac{1}{n-1}\sum_i (y_i)^2\right)_{noise}}} \qquad \text{V.9}$$

While *SNR* is related to the area of the studied peak, *PSNR* is related to its height, leading to different results. They can be expressed in decibels (Equation V.10), which is more convenient to explore a wide variation range.

$$SNR^{dB} = 10\log_{10}(SNR) \quad , \quad PSNR^{dB}_{rms} = 10\log_{10}(PSNR^2_{rms}) \qquad \text{V.10}$$

Additionally, Currie defined a Critical Level (L_c), a Detection Limit (L_d) and a Quantitative Limit (L_q), at 1.64, 3.29 and 10 σ_{noise}, respectively (*114*). The corresponding levels, SNR and PSNR measurements are drown in Figure V.4.

[4] $PSNR_{rms}$ is halved by tradition in NMR literature, including Bruker TopSpin software (*113*).

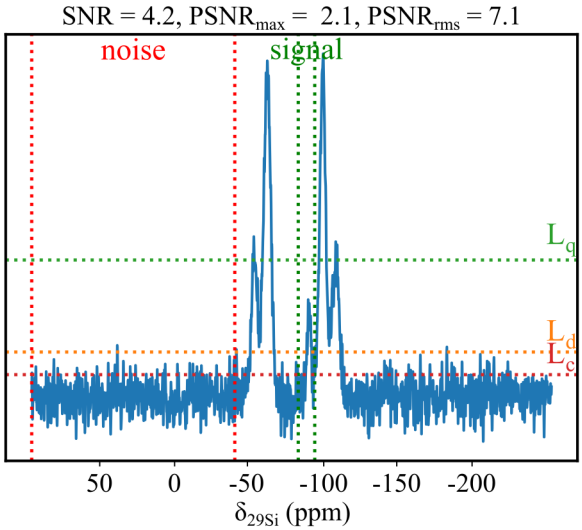


Figure V.4: Measurement of Signal-to-Noise Ratio (*SNR*) and Peak SNR (*PSNR*) based on $noise_{max}$ and $noise_{rms}$ for the ^{29}Si MAS solid-state NMR spectrum of the 50:50 MTEOS:TEOS sample. Signal and noise regions are highlighted with dotted vertical green and red lines, respectively. Critical (L_c), detection (L_d) and quantitative (L_q) limits (*114*) are shown with red, orange and dotted horizontal green lines, respectively.

B.4. Results and discussion

In this subsection, SVD is applied to NMR and Raman spectra. We then focus on practical aspects of denoising, namely the impact of the matrix shape and the number of components used for thresholding. Time and frequency denoising are compared. The minimum experimental SNR needed for valid use of SVD and its impact on peak shape are thoroughly investigated. Finally, a limit case is evaluated.

B.4.a. Denoising of NMR and Raman spectra

Images are already matrices and SVD can directly be applied on them. Two-dimensional (2D) spectra can be treated similarly. However one-dimensional (1D) spectra are not directly suitable for SVD. If a series of spectra is available, one just need to stack the successive spectra to obtain a 2D data set. Figure V.5a shows such a stack of spectra simulating a reaction kinetics studied by Raman spectroscopy, as described in Subsection B.2.c. A similar stack can be obtained when a surface is mapped to analyse species in a sample region (*43*). The spectrum consisted of four overlapped peaks with varying intensities. Without SVD, the two peaks at 510 and 900 cm^{-1} were difficult to detect due to the amount of noise (red arrows). However, after SVD, these components were identified as evidenced by green arrows. One should note the very low residual noise.

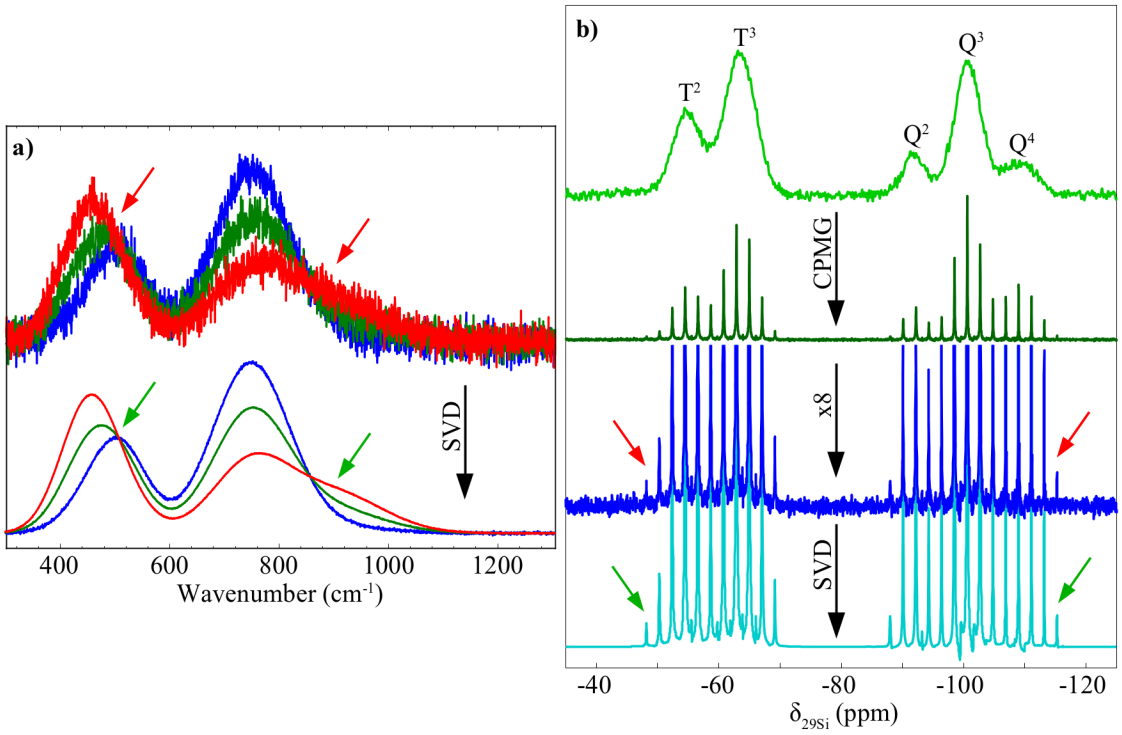


Figure V.5: a) SVD applied to a set of spectra mimicking a kinetic reaction, as probed by Raman spectroscopy. Four overlapping bands are simulated on 2000 spectra with 2000 points each, resulting into a matrix of 2000 × 2000 points. Only three representative spectra are shown. b) ^{29}Si MAS solid-state NMR spectra of the 50:50 MTEOS:TEOS sample. From top to bottom: standard spectrum, CPMG spectrum, vertically zoomed CPMG spectrum, spectrum obtained after SVD applied on time domain signal of the above transformed to an Hankel matrix of 2015 × 2014 points. Red and green arrows show difficult to detect and enhanced signals, respectively.

It is also possible to use SVD on a 1D data set, by way of a Hankel or a Toeplitz matrix, as described in Subsection B.3.b. This feature was used in the context of ^{29}Si MAS solid-state NMR (Figure V.5b) using the GPU Java application (*94*). The ^{29}Si MAS NMR spectrum of the 50:50 MTEOS:TEOS sample (see Subsection B.2.a) is shown in the top of the figure. It should be noted that this spectrum already displays a very good SNR. The signal has been enhanced using CP, allowing a non-quantitative spectrum to be acquired in 40 minutes. The second spectrum shows the same sample analysed using CPMG echoes (*51*), which led to numerous spikelets. The overall shape of the CPMG spectrum is qualitatively similar to the above spectrum. Such an approach was proved to be a mean to increase SNR during acquisition step by discretising broad peaks (*115*). The original shape with improved SNR can be recovered by summing echoes, but nevertheless this leads to relaxation distortions. The third spectrum of Figure V.5b is a vertical zoom to highlight noise level and much less intense spikelets marked by red arrows. By comparison with the same spectrum

after SVD processing, such small peaks were highly enhanced and noise has disappeared. This ^{29}Si CPMG MAS NMR spectrum was used hereafter as a reference to evaluate the performances of the SVD process.

B.4.b. Matrix shape

It is sometimes argued that efficient denoising can be obtained using an iterative process on a rectangular matrix, with a number of columns higher than the number of signals of interest (*116*), *i.e.*, roughly the number of peaks. The iteration consists in converting into a matrix, applying SVD, and reverting to a 1D set of denoised data. Due to (anti-)diagonals averaging (see Subsection B.3.b), the matrix at the beginning of the second iteration is not exactly the one at the end of the first iteration, which explains that multiple iterations give different results. However in our case, such a procedure led to some residual noise (Figure S.V.1). Even with $m \times n = 3901 \times 128$ points and 10 iterations, noise was only marginally reduced. Nevertheless, this procedure has the advantage of a low computation time *per* iteration, as Hankel or Toeplitz matrix size is smaller for a rectangular shape than a square shape, when starting with the same number of points in 1D data set (see part (II) of this work (*12*)).

In a second step, Figure V.6a depicts the efficiency of denoising on various matrix shapes, either rectangular ones ($m \times n = 3997 \times 32$ points) or square ones ($m \times n = 2015 \times 2014$ points). While noise was strongly present for elongated matrices, it decreased when the number of columns increased and finally disappeared for a square matrix. Square matrices were used hereafter. More generally, our results suggest to tend to a square data matrix before applying SVD, as also recommended by Van Huffel *et al.* (*117*). For $n = 512$ and $n = 2014$, small peaks seem to be missing, highlighted by orange arrows. This feature is explained in next subsection.

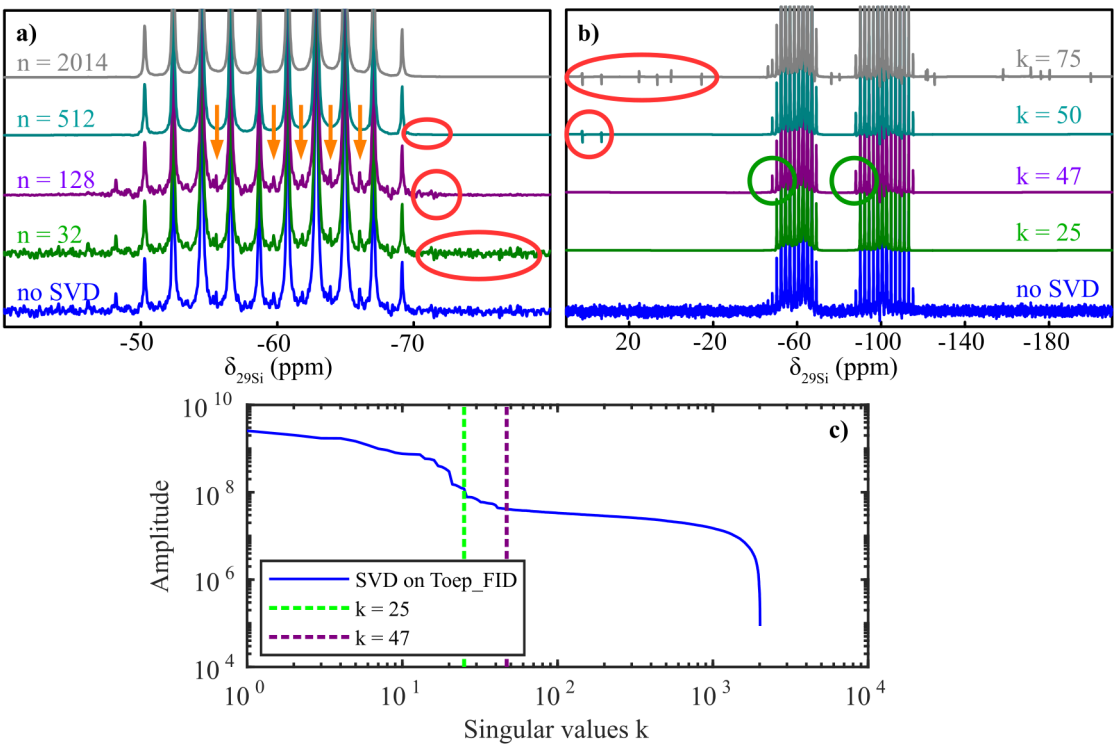


Figure V.6: a) Influence of the number of columns (n) for Toeplitz FID (Toep_FID) matrix construction, with $m + n = 4028$ points and $k = 22$ singular values, for a ^{29}Si CPMG solid-state NMR spectrum of the 50:50 MTEOS:TEOS sample. b) influence of the number of singular values (k) and c) singular values in logarithmic scale for a matrix of 2015 × 2014 points ($n = 2014$). 22 major spikelets and 47 in total are present. Red and green circles show artefacts and enhanced small signals, respectively; orange arrows are minor spikelets visible at $k = 47$.

B.4.c. Thresholding

Another parameter to be varied in order to optimise denoising is the number of singular values k (corresponding to signals) used for low-rank approximation (Figure V.6b). In a first attempt, k was set to the number of peaks present, $k = 22$. However, $k = 25$ resulted in a better shape for the three more intense spikelets, corresponding to Q^3 and T^3 peaks (Figure V.5b) and $k = 47$ was necessary to select all small peaks that were missing in previous subsection (green circles and orange arrows). Above this value, isolated artefacts were observed as shown for $k = 50$ and $k = 75$ (red ellipses). They are usually narrow and out of phase, which make them easy to detect. It is thus necessary to carefully adjust k to discriminate signals from noise.

Another approach to manually select the correct number of singular values was to plot singular values in logarithmic scale (Figure V.6c). Up to $k = 25$, singular value amplitude was

strongly decreasing. Between $k = 25$ and $k = 47$ (dashed vertical lines), a slow slope was present and it was hard to distinguish the optimal value, because these singular values had a too low SNR. Above $k = 47$, the curve exhibited a plateau and finally a cliff for last indexes. The plateau and the cliff are characteristic of noise values (*118*).

In order to select automatically the proper number of singular values, Malinowski developed an INDicator function (*IND*) and a Significance Level (*SL*) function (*119*), which are available under Matlab in file sfa.m of (*103*). While IND is based on the residual standard deviation, SL is a Fisher variance test giving the probability for a singular value to correspond to noise. In the former case, the minimum of *IND* reflects the number of singular values to select. In the latter case, the singular value is rejected if it has a probability of being noise higher than a desired level. The influence of this level will be investigated in Subsection B.4.e.ii. When applying these functions to the spectrum reported in Figure V.6b, we found $k = 53$ with *IND* and $k = 31$, 36 and 39 with *SL* at 5, 7.5 and 10 % error level, respectively. As stated before, $k = 47$ was the manual optimum for the spectrum considered here, selecting all singular values attributed to signals while rejecting artefacts. While *IND* overestimated k and introduces artefacts, *SL* underestimated k, thus ignoring small signals. Nevertheless, *SL* was the tool of choice for automatic thresholding as the result was close from the expected one and did not display artefacts. It should be noted that singular value thresholding is also available (*120*) but it needed to adjust too many parameters and to sparsify data.

B.4.d. Time and frequency domains

In this subsection, spectra resulting from a same FID with SVD denoising applied either before (Figure V.7, Toep_FID, in cyan) or after Fourier transform (Toep_SPC, in green) are compared. Both were simulated under Matlab and converted into a square Toeplitz matrix, according to Subsection B.4.b, before applying SVD. Additionally, multiple similar spectra with identical signal and random noise were stacked (mult_SPC, in red), to simulate a mapping of a homogeneous region in MRI or in Raman. For mult_SPC, we chose a number of spectra identical to the number of points per spectrum, again to get a square matrix. Mult_SPC was also useful to compare the influence of multiple sampling *vs.* a single sampling converted to a semi-circulant matrix. Singular values are presented in Figure V.7a. By construction, the size of Toep_FID and Toep_SPC was half the one of mult_SPC, which explains that their maximum singular value indexes were lower. Singular value plot was very similar for Toep_FID and mult_SPC, but amplitude of the latter was higher. The difference

between signals and noise singular values was also more pronounced on mult_SPC. On the contrary, Toep_SPC had a very different signature, with 17 singular values corresponding to signals, instead of only one that was expected.

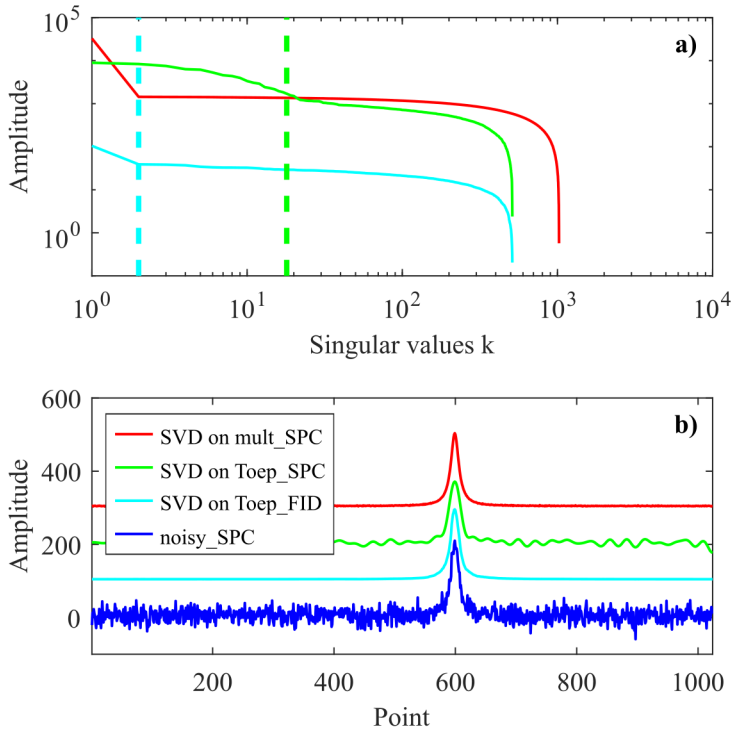


Figure V.7: a) Singular values and b) spectra obtained with and without SVD. noisy_SPC: Fourier transform (SPC) of noisy FID simulated with 1024 points; Toep_FID and Toep_SPC: Toeplitz matrix of FID and SPC, respectively, with 513 × 512 points; mult_SPC: matrix of stacked multiple identical spectra with random noise, characterised by 1024 × 1024 points. Dashed lines: corresponding Significance Level (*SL*) noise limit.

Noisy and denoised spectra are shown in Figure V.7b. Interestingly, very different behaviours were observed. Satisfactory denoising was achieved using mult_SPC or Toep_FID. This result explains why SVD was efficient in the case of the afore mentioned Raman spectra mimicking a kinetic process (Figure V.5a) and the NMR FID (Figure V.5b). On the contrary, for Toep_SPC, noise –though reduced– was still present. This clearly denotes that in case of 1D frequency domain, a reverse Fourier transform before converting data into a Toeplitz matrix and applying SVD should be preferred[5] (*121*).

[5] More simply, prefer to use the time domain data if available.

B.4.e. Minimum signal-to-noise ratio

B.4.e.i. Comparison of SNR^{dB} and $PSNR_{rms}{}^{dB}$

An additional question is the sensitivity of SVD, that is to say the minimum SNR needed to get a proper signal detection. The corresponding indicators, SNR^{dB} and $PSNR_{rms}{}^{dB}$, were defined in Subsection B.3.c. 7380 noisy spectra were simulated with either Lorentzian or Gaussian shape and with a peak width of either 32 Hz, 320 Hz or 3200 Hz (see Subsection B.2.d). First, we compared in Figures V.8a and V.8b the SNR^{dB} determined using the pure signal and the separated desired noise (*calculated* SNR^{dB}) to the $PSNR_{rms}{}^{dB}$ measured on signal and noise regions of the noisy data (*measured* $PSNR_{rms}{}^{dB}$). *Calculated* SNR^{dB} was a simulating tool, close to the theoretical SNR^{dB}, whereas *measured* $PSNR_{rms}{}^{dB}$ reflected an experimental assessment, which was directly obtained on noisy spectra. Comparing them was a way to check how the noise itself could influence $PSNR_{rms}{}^{dB}$. For narrow peaks (in red) a linear relationship between SNR^{dB} and $PSNR_{rms}{}^{dB}$ was obtained. However, for intermediate and broad peaks (in green and blue, respectively), the evolution of SNR^{dB} with $PSNR_{rms}{}^{dB}$ displayed a steeper increase, with a vertical asymptote at $PSNR_{rms}{}^{dB} = 10$ dB (dashed black line). This value reflected an undetectable signal with $H_{signal} \leq h_{noise\_max}$, as $\sigma_{noise} = h_{noise\_max} / 3.3$ with a probability of 99.9 % for Gaussian noise. The increase resulted from spreading of peak area over a wider range. This implied a lower amplitude for broad peaks than for narrow peaks and consequently a lower $PSNR_{rms}{}^{dB}$ value.

The 7380 noisy spectra were then truncated to 5 T_2 and the same comparison was performed between calculated SNR^{dB} and measured $PSNR_{rms}{}^{dB}$ (Figures S.V.2a and S.V.2b). This truncation removed the vertical asymptote observed at low $PSNR_{rms}{}^{dB}$, as less points were defining noise, and intense noise peaks were less probable. This led to an artificial increase[6] of SNR and to a vertical shift of $SNR^{dB} = f(PSNR_{rms}{}^{dB})$ evolution for intermediate and broad lines, pictured in green and blue, respectively, in Figures S.V.2a and S.V.2b. Moreover, a broader distribution of SNR values was obtained, especially for broad peaks (blue line). An additional feature was observed for broad Lorentzian peaks: a vertical asymptote at $PSNR_{rms}{}^{dB} = 37$ dB. This feature was a consequence of the spectral extension of the wings of Lorentzian peaks that were contributing to amplitude within noise region. On the contrary, Gaussian peak wings are much less intense and in this case, this vertical asymptote was not observed on $SNR^{dB} = f(PSNR_{rms}{}^{dB})$ evolution.

[6] The term ‘apparent increase’ would be more accurate.

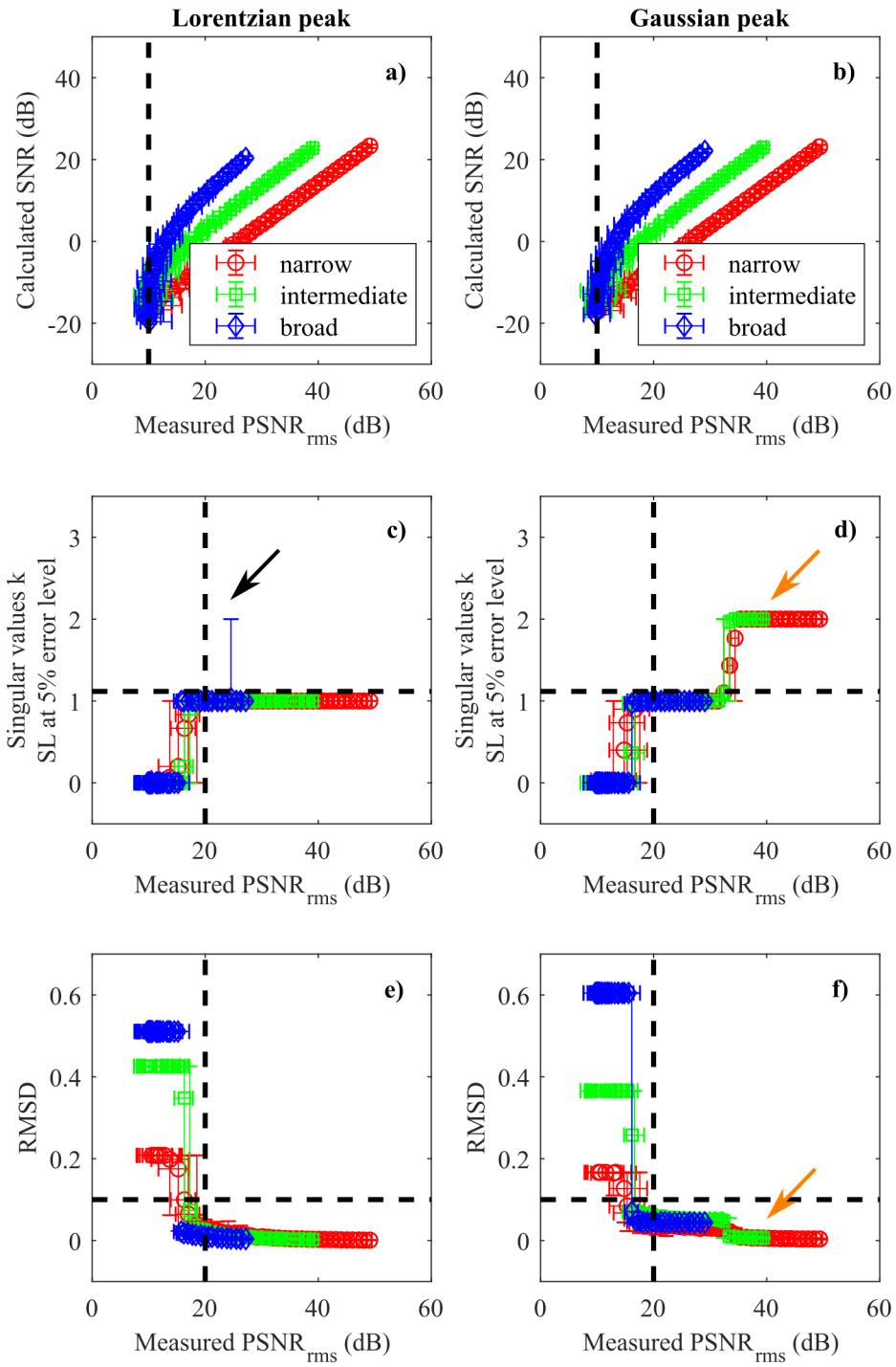


Figure V.8: 7380 simulated spectra with known added homoscedastic white Gaussian noise. a), c) and e) Lorentzian peak; b), d) and f) Gaussian peak; a) and b) comparison of *SNR* definitions; c) and d) automatic thresholding with *SL* at 5% error level; e) and f) root mean square deviation of denoised spectra. Error bars correspond to the repetition of 30 simulated spectra with the same level of added noise. 41 levels of noise were used ranging from -20 dB to +20 dB. SVD was applied on time data. *SNR* and *PSNR* were obtained on frequency data. Dashed black lines represent characteristic values (see text for more details); black arrow shows an artefact detection; orange arrows highlight the second singular value for Gaussian peaks. Data set is available in file Data_SL5.mat of (*103*).

B.4.e.ii. Automatic thresholding

SVD and Malnowski's *SL* automatic thresholding (see Subsection B.4.c) were applied to these simulated NMR FID corresponding to only one peak. When *SL* error level was set to 5 % and above $PSNR_{rms}^{dB} = 20$ dB this single peak was detected with $k = 1$ singular value, whatever peak width and shape (Figures V.8c-d and S.V.3a-b, dashed vertical black line). However, this was only an upper limit and many peaks were detected around $PSNR_{rms}^{dB} = 17$ dB, between the detection limit of 3.3 σ_{noise} and the quantification limit of 10 σ_{noise}, as defined by Currie (*114*). Thus, to be detected through Malinowski's algorithm, a signal has to be enough different from noise, *i.e.*, between two to three times higher than $noise_{max}$. Surprisingly, a second singular value was detected with $k = 2$ for Gaussian shapes above $PSNR_{rms}^{dB} = 36$ dB, (Figures V.8d and S.V.3b). The amplitude of this second component was significant and improved the resulting shape of denoised spectrum. This second singular value will be explained in Subsection B.4.f.iii. On the full data set of 7380 spectra, only one false detection was observed, as indicated by the black arrow in Figure V.8c. Thus, the amount of artefacts was negligible. On the truncated FID (Figures S.V.2c and S.V.2d), this limit of $PSNR_{rms}^{dB} = 20$ dB was not so abrupt, due to lack of accuracy on noise measurement. However, an advantage of truncation was a much faster computation for broad peaks, thanks to the smaller matrix used.

Table V.1: *PSNR* needed for SVD with *SL* automatic thresholding depending on the desired error level.

***SL* error level**	$\boldsymbol{PSNR_{rms}^{dB}}$	$\boldsymbol{PSNR_{rms}}$	$\boldsymbol{PSNR_{max}}$	**Artefacts**
5 %	17	7.1	2.1	no
7.5 %	16	6.3	1.9	small
10 %	15	5.6	1.7	strong

When *SL* error level was set to 7.5 or 10 %, the minimum SNR to get a peak detection was decreasing (Figures S.V.3c-f). However, the number of false detections also increased noticeably, as indicated by the black arrow. In some rare cases, evidenced with lines being higher than the figure vertical limit (black arrow on Figure S.V.3d), *SL* was unable to distinguish signal from noise, resulting in a noisy spectrum after SVD. While an error level of 5 % is really safe, a level of 7.5 % may be necessary to detect tiny peaks. A value of 10 % seems too high to avoid artefacts. Results are summarised in Table V.1. *SNR* is not presented in this table as it is not a relevant parameter, that is too much depending on peak width and shape.

B.4.e.iii. Error measurement

The difference between denoised signal and simulated non-noisy signal (pure signal), was measured using the Root Mean Square Deviation (*RMSD*), defined on Equation V.11, where $y_i^{denoised}$ and y_i^{pure} are the individual values for denoised and pure SPC, respectively (Figures V.8e and V.8f).

$$RMSD = \sqrt{\frac{\sum_{i=1}^{n} (y_i^{denoised} - y_i^{pure})^2}{n}} \qquad \text{V.11}$$

A high RMSD was obtained below $PSNR_{rms}^{dB}$ = 17 dB. As no peak was detected in this range, the obtained value corresponded to RMSD of pure signal compared to zero, which was higher for broad peaks, due to its wide spread range. Above this $PSNR_{rms}^{dB}$ value, RMSD displayed a steep decrease under 0.1 (dashed horizontal black line). These results emphasised the very good agreement between denoised and pure data. However, *RMSD* was higher for Gaussian than for Lorentzian peaks (Figure V.8f). Above the second threshold of $PSNR_{rms}^{dB}$ = 36 dB, *RMSD* exhibited a further decrease down to the level obtained for Lorentzian peaks. This confirms the significance of the second singular value. A similar trend was observed on truncated data (Figures S.V.2e and S.V.2f).

B.4.f. Quantification

B.4.f.i. Pure and denoised spectra

The next step that was investigated concerns the possibility to use denoised spectra for the sake of quantification. For each peak width and shape at $PSNR_{rms}^{dB}$ = 20 dB, *i.e.*, at quantification limit, the spectrum with the worst RMSD is presented in Figure V.9. These spectra were fitted with a Voigt function with error estimation implemented in Matlab (*122*). Amplitude, position, shape and width were automatically adjusted. Results are reported in Tables V.2 and V.3 for Lorentzian and Gaussian peaks, respectively. A very large uncertainty occurred on fitting parameters derived from noisy spectra (top traces). For denoised spectra (middle top traces), uncertainty decreased significantly, roughly by a factor of 10, but was still higher than for pure spectra (middle bottom traces). Difference between denoised and pure spectra is presented in bottom trace of Figure V.9.

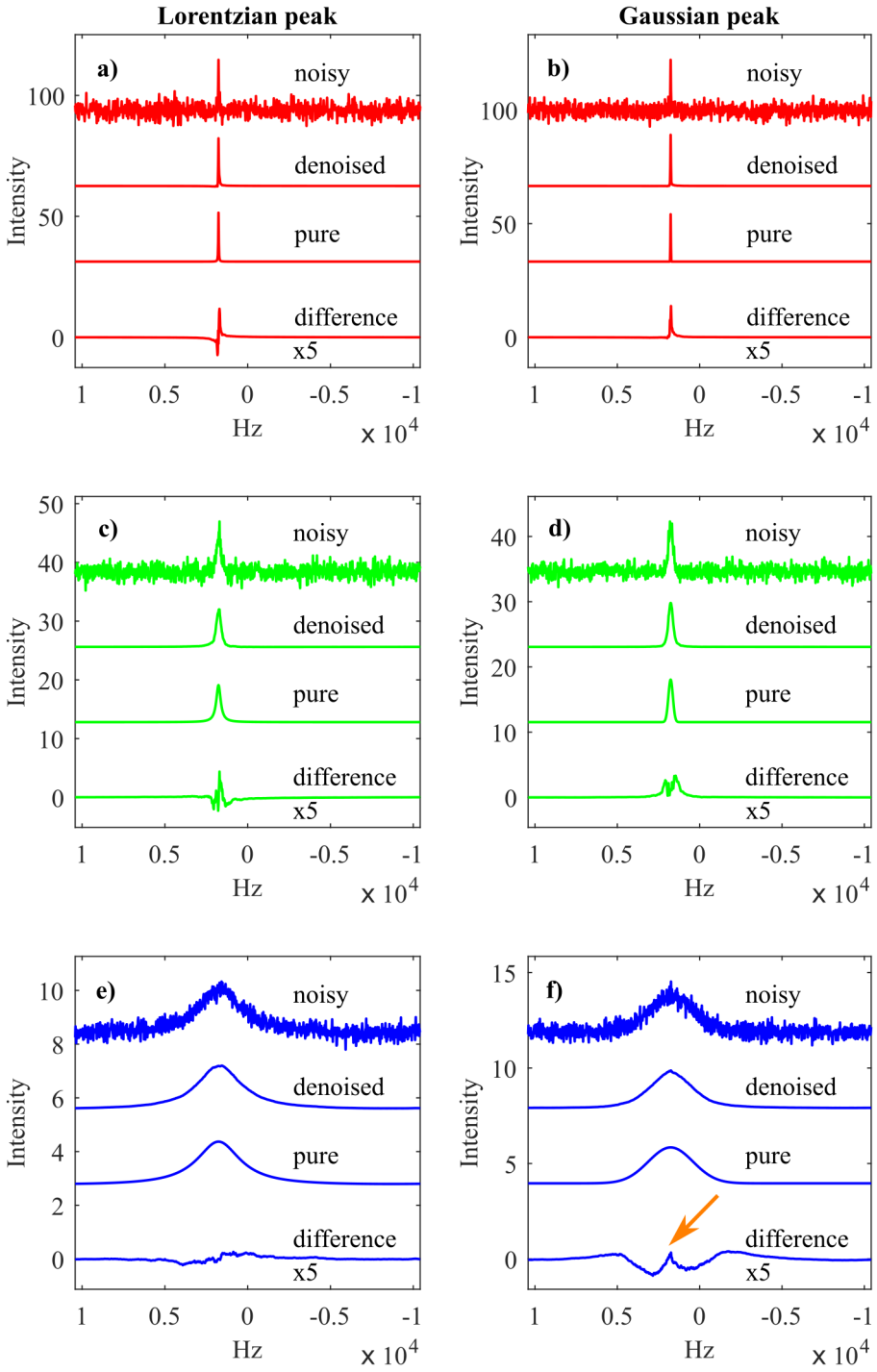


Figure V.9: Spectra obtained in Figure V.8 at $PSNR_{rms}^{dB} = 20$ dB with worst *RMSD* for a), c) and e) Lorentzian peak; b), d) and f) Gaussian peak; a) and b) narrow peak in red; c) and d) intermediate peak in green; e) and f) broad peak in blue. On each subfigure, top spectrum, middle top, middle bottom, bottom spectra correspond to noisy, denoised, pure spectra and difference between denoised and pure spectra, respectively. Orange arrow highlight the Gaussian peak distortion.

Table V.2: Modelling of pure, noisy and denoised **Lorentzian peak** with narrow, intermediate and broad widths (in red, green and blue, respectively), for low, intermediate and high $PSNR_{rms}^{dB}$ (dark grey, light grey and white rows, respectively). a: Matlab file exchange 52321 (*122*); b: root mean square deviation to pure spectrum (Equation V.11); c: Percent error on pure spectrum area (Equation V.12).

				Modelling [a]					
Shape	**T_2 (ms)**	**Signal**	**RMSD** [b]	**Ampli-tude**	**Position (Hz)**	**G/L**	**Width (Hz)**	**Area**	**PE_{area} (%)** [c]
Lorentzian low $PSNR_{rms}^{dB}$	10	Noisy 19.6 dB	2.78	21 ± 5	1763 ± 6	0.0 ± 0.8	40 ± 10	1185	+ 17.3
		Denoised	0.05	19.7 ± 0.3	1760.3 ± 0.4	0.00 ± 6e-2	35 ± 1	1096	+ 8.5
		Pure	-	20.191 ± 3e-3	1761.87 ± 4e-3	9.2e-3 ± 7e-4	31.99 ± 1e-2	1010	-
	1	Noisy 19.5 dB	1.12	6.1 ± 0.8	1710 ± 20	0.3 ± 0.4	320 ± 60	2693	- 14.4
		Denoised	0.04	6.27 ± 7e-2	1729 ± 2	0.27 ± 4e-2	342 ± 6	2992	- 4.9
		Pure	-	6.299 ± 3e-3	1761.89 ± 6e-2	4e-3 ± 1e-3	318.5 ± 0.2	3145	-
	0.1	Noisy 20.2 dB	0.22	1.54 ± 6e-2	1520 ± 50	0.3 ± 0.2	3200 ± 200	6665	+ 2.5
		Denoised	0.02	1.52 ± 1e-2	1590 ± 10	0.35 ± 4e-2	3190 ± 40	6529	+ 0.4
		Pure	-	1.52 ± 1e-2	1762 ± 8	0.34 ± 3e-2	3160 ± 30	6502	-
Lorentzian interm $PSNR_{rms}^{dB}$	10	Noisy 30.1 dB	0.79	20 ± 1	1763 ± 2	0.1 ± 0.3	31 ± 4	952	- 5.7
		Denoised	0.01	20.56 ± 7e-2	1762.8 ± 0.1	0.00 ± 2e-2	30.8 ± 0.3	995	- 1.5
		Pure	-	20.191 ± 3e-3	1761.87 ± 4e-3	9.2e-3 ± 7e-4	31.99 ± 1e-2	1010	-
Lorentzian high $PSNR_{rms}^{dB}$	10	Noisy 36.2 dB	0.40	19.8 ± 0.6	1761.4 ± 0.8	0.1 ± 0.1	32 ± 2	963	- 4.7
		Denoised	0.01	19.90 ± 1e-2	1761.60 ± 2e-2	0.061 ± 3e-3	32.40 ± 4e-2	984	- 2.6
		Pure	-	20.191 ± 3e-3	1761.87 ± 4e-3	9.2e-3 ± 7e-4	31.99 ± 1e-2	1010	-

Table V.3: Modelling of pure, noisy and denoised **Gaussian peak** with narrow, intermediate and broad widths (in red, green and blue, respectively), for low, intermediate and high $PSNR_{rms}^{dB}$ (dark grey, light grey and white rows, respectively). a: Matlab file exchange 52321 (*122*); b: root mean square deviation to pure spectrum (Equation V.11); c: Percent error on pure spectrum area (Equation V.12).

				Modelling [a]					
Shape	**σ (ms)**	**Signal**	**RMSD** [b]	**Ampli-tude**	**Position (Hz)**	**G/L**	**Width (Hz)**	**Area**	**PE_{area} (%)** [c]
Gaussian low $PSNR_{rms}^{dB}$	12	Noisy 19.9 dB	2.81	22 ± 4	1762 ± 4	1 ± 1	37 ± 9	927	+ 31.3
		Denoised	0.08	22.51 ± 8e-2	1761.00 ± 9e-2	0.42 ± 2e-2	34.2 ± 0.2	1007	+ 42.6
		Pure	-	20.827 ± 9e-5	1761.89 ± 9e-5	1.000 ± 3e-5	31.861 ± 2e-4	706	-
	1.2	Noisy 20.4 dB	0.88	6.7 ± 0.5	1790 ± 10	1.0 ± 0.5	330 ± 30	2329	+ 5.3
		Denoised	0.07	6.78 ± 2e-2	1757.3 ± 0.4	0.46 ± 1e-2	331 ± 1	2891	+30.8
		Pure	-	6.5193 ± 2e-5	1761.89 ± 5e-4	1.000 ± 2e-5	318.61 ± 1e-3	2211	-
	0.12	Noisy 20.5 dB	0.32	1.86 ± 8e-2	1810 ± 50	0.9 ± 0.3	3200 ± 100	6499	+ 1.9
		Denoised	0.05	1.839 ± 7e-3	1734 ± 4	0.62 ± 2e-2	2980 ± 10	6637	+ 4.1
		Pure	-	1.8800 ± 2e-5	1761.89 ± 1e-2	1.000 ± 8e-5	3186.0 ± 3e-2	6376	-
Gaussian intem $PSNR_{rms}^{dB}$	12	Noisy 29.7 dB	0.89	21 ± 1	1761 ± 2	0.2 ± 0.4	30 ± 4	899	+ 27.3
		Denoised	0.04	22.24 ± 4e-2	1761.80 ± 4e-2	0.38 ± 1e-2	30.77 ± 9e-2	913	+ 29.3
		Pure	-	20.827 ± 9e-5	1761.89 ± 9e-5	1.000 ± 3e-5	31.861 ± 2e-4	706	-
Gaussian high $PSNR_{rms}^{dB}$	12	Noisy 36.3 dB	0.40	21.4 ± 0.6	1761.7 ± 0.6	1.0 ± 0.2	31 ± 1	715	+ 1.3
		Denoised	0.02	21.32 ± 6e-2	1761.85 ± 6e-2	1.00 ± 2e-2	31.4 ± 0.1	713	+ 1.0
		Pure	-	20.827 ± 9e-5	1761.89 ± 9e-5	1.000 ± 3e-5	31.861 ± 2e-4	706	-

B.4.f.ii. Lorentzian and Gaussian peaks

Surprisingly, despite a *RMSD* lower than 0.1, the area Percent Error (PE_{area} given by Equation V.12) could be as high as 8.5 % and 42.6 % for Lorentzian and Gaussian peaks, respectively.

$$PE_{area}(\%) = \frac{Area^{denoised} - Area^{pure}}{Area^{pure}} \times 100 \qquad \text{V.12}$$

While the former was acceptable at detection limit, the latter evidenced an overestimation. Although PE_{area} decreased after SVD denoising on Lorentzian peaks, it increased for Gaussian peaks. Moreover, difference spectrum on Gaussian peak exhibited a mix of narrow and wide components with opposite amplitudes (Figure V.9f). Such a shape modification was not observed for Lorentzian peaks. Besides, the Gaussian/Lorentzian ratio was around 0.5 instead of 1.0 after denoising on Gaussian peaks (dark grey and light grey rows of Table V.3). This result highlighted that SVD induced a change in peak shape from Gaussian peaks to more Lorentzian ones. Above the second threshold of $PSNR_{rms}^{dB} = 36$ dB, the shape was corrected thanks to the second singular value, giving a pure Gaussian peak after denoising (white rows of Table V.3).

B.4.f.iii. Real and extracted errors

The error on the measured area can origin from two sources: first the error introduced by the added noise, known as the *real error*, and second the error coming from the denoising itself, so-called the *extracted error* (*119*). An example of real error is presented in Figure S.V.4a. Two successive measurements with $NS = 120$ scans gave a different amplitude for Q^2 peak at -91 ppm (red arrow). With a higher noise averaging at $NS = 360$ (not shown), amplitude ratios were similar to $NS = 840$. Thus, *SNR* at $NS = 120$ was too low and amplitude was tainted by error. A strong apodisation has been used here to artificially improve *SNR* (see Subsection B.4.g.i). Automatic thresholding was unable to correctly discriminate signals from noise and manual thresholding with $k = 5$ singular values was preferred. Nevertheless, the real error was kept after denoising (Figure S.V.4b), which demonstrated that manual thresholding is a dangerous tool. Failure of automatic thresholding is thus an indication that SNR has to be improved.

The extracted error was especially present for Gaussian spectra, for which the Gaussian/Lorentzian ratio modification (see Subsection B.4.f.ii) led to an increase of peak area. Indeed, SVD fits time decays with a sum of exponential (*67*). When fitting a Gaussian decay with a single exponential component, corresponding to one singular value at low $PSNR_{rms}^{dB}$, the peak area is correspondingly overestimated by 20 % (*123*). Our results were consistent with this value. Gaussian and exponential decays are very different, as Gaussian is flatter around its maximum. Above the second threshold, the Gaussian decay was fitted with two exponential decays, improving peak area value.

Unfortunately, when studying solid-state samples by using spectroscopic approaches, peaks are most of the time not Lorentzian. In such cases, SVD quantitative results are difficult to obtain. A workaround would be to model the resulting spectrum with pseudo-Voigt functions. For peaks with a Gaussian/Lorentzian ratio around 0.5, dividing their area by 1.2 (20 %) should improve quantification. Taking this precaution into account for analysis of our data, PE_{area} was found to be similar between denoised spectra at $PSNR_{rms}^{dB} = 20$ dB (dark grey rows in Tables V.2 and V.3) (*111*) and noisy spectra at $PSNR_{rms}^{dB} = 30$ dB, (light grey rows in Tables V.2 and V.3). In solid-state NMR, another possibility to avoid Gaussian peak error relies on use of CPMG echoes (*51*) as in Figure V.5b. This technique transforms a peak driven by chemical shift distribution (Gaussian shape, inhomogeneous interaction) into multiple narrow peaks driven by relaxation (Lorentzian shapes, homogeneous interaction) (*124*, *125*), which are very suitable for SVD, being both sensitive and quantitative.

B.4.g. Limit case on a real NMR spectrum

B.4.g.i. Pre-processing

A pre-processing step called apodisation can be applied on FID before SVD and Fourier transform. The aim is first to reduce noise, and second to remove truncation artefacts leading to oscillations at peak foot, hence the name. In NMR, one can use for instance either exponential, cosine, or (shifted-)Gaussian decays. Their shapes were compared in (*126*). While exponential is convex, cosine is concave and Gaussian is intermediate. In Figure S.V.4c, we compared the influence of apodisation on initial noise. The resulting denoised SPC with automatic thresholding at an *SL* error level of 7.5 % are presented in Figure S.V.4d. Without apodisation, SNR was too low to detect Q^2 peak at -91 ppm (red circle) and $k = 4$ singular values were found. With cosine apodisation, the correct number of

peaks was obtained with $k = 5$ singular values (green ellipse). With an exponential decay of 20 Hz, corresponding to the intrinsic SPC resolution, a similar result was obtained, but with $k = 6$ singular values, leading to small baseline distortions. Surprisingly, with an exponential decay of 50 Hz, a much higher number of $k = 94$ singular values was found, with almost no denoising (orange circle). An explanation was that apodisation changed the amplitude of noise values, especially at the end of the FID. By this way, noise became heteroscedastic, decreasing efficiency of SVD and Malinowski's criterion. When plotting singular values in logarithmic scale (Figure S.V.4e), the slope moved from a plateau for cosine (purple curve) to a decay for exponential with 50 Hz (blue curve). In such a case, it is harder to discriminate signals from noise, as the slope is similar. Cosine is thus a good compromise before SVD as it decreases noise without changing too much singular values[7]. An alternative would be to combine SVD and Savitzky–Golay smoothing filter (*127*), which process noise the same way all over the FID[8].

B.4.g.ii. Denoising

On the sol-gel 50:50 MTEOS:TEOS sample, four hours and $NS = 240$ scans with cosine pre-processing, were needed to have a spectrum with a sufficient SNR to apply SVD and to detect Q^2 peak with automatic thresholding at *SL* error level of 10 % (not shown). If *SL* error level was limited to 7.5 %, six hours and $NS = 360$ scans were necessary (Figure V.10 top trace). The corresponding denoised spectrum (middle top trace) was very close to the reference spectrum acquired in fourteen hours and $NS = 840$ (middle bottom trace), with $PSNR_{rms} = 9.7$, *i.e.*, at quantification limit. Their difference (bottom trace) was comparable to noise. This was confirmed by peaks integration (Table V.4) where very good agreement was obtained between the denoised spectra at $NS = 240$ or $NS = 360$ and the reference noisy spectrum at $NS = 840$. In particular, Q^2/Q^3 ratio was very consistent. Time gain was thus between 2.3 to 3.5, depending on the *SL* error level allowed.

[7] Square cosine gave similar results with a slightly more pronounced slope.

[8] More precisely, Savitzky–Golay filter can be applied on singular vectors but not on the FID.

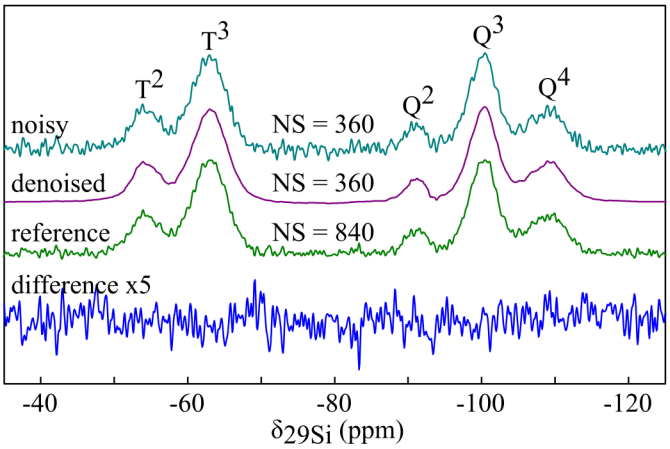


Figure V.10: ^{29}Si MAS solid-state NMR spectrum of the 50:50 MTEOS:TEOS sample. From top to bottom: noisy spectrum at *NS* = 360 scans; denoised spectrum at *NS* = 360 scans with *k* = 5 singular values at *SL* error level 7.5 %; reference spectrum at *NS* = 840; difference between denoised and reference spectra. The spectrum is quantitative with an impulsion of 30° and a relaxation delay of 60 s.

Table V.4: Peaks integration on noisy (grey rows) and denoised (white rows) spectra of the 50:50 MTEOS:TEOS sample for various number of scans (*NS*). The spectrum is quantitative with an impulsion of 30° and a relaxation delay of 60 s. Spectra were modelled using Dmfit (*128*).

NS	**Pre-proces-sing**	**PSNR rms**	***SL* error level**	**% T^2**	**% T^3**	**% Q^2**	**% Q^3**	**% Q^4**	**Q^2/Q^3**
120	cosine	3.9	no	11.5	38.8	3.3	29.1	17.2	0.11
			10 %	12.9	39.8	0	29.2	18.1	0
240	cosine	5.2	no	12.0	37.2	4.7	29.6	16.5	0.16
			10 %	12.5	36.3	5.5	28.8	16.8	0.19
360	cosine	7.1	no	12.3	36.5	4.8	30.2	16.2	0.16
			7.5 %	12.8	35.6	5.5	29.5	16,6	0.19
840	cosine	9.7	no	13.2	35.8	5.4	29.7	15.9	0.18
			7.5 %	13.7	35.2	5.8	29.1	16.1	0.20

B.5. Conclusion

Singular Value Decomposition is of crucial importance in many mathematical treatments involved in spectrocopies. In this first part (I), SVD with low-rank approximation was successfully applied to denoise NMR and Raman spectra. This approach can easily be generalised to other spectroscopies. We have shown that a better denoising was obtained with square matrices and with SVD applied to time domain signal rather than to the corresponding frequency spectrum. Automatic thresholding was used thanks to Malinowski's Significant

Level indicator and a 7.5 % error value was a good compromise between sensitivity and unwanted artefacts. 6 × 7380 SVD were carried out to compare pure, noisy and denoised spectra with SNR^{dB} ranging over 41 dB. Our results proved that this technique can detect signals as low as twice $noise_{max}$, *i.e.*, with $PSNR_{max} = 2.0$ and $PSNR_{rms} = 6.6$, whatever the peak width. A systematic shape modification has been highlighted for Gaussian peaks with an overestimation of peak area by 20 %. This overestimation for Gaussian peaks is a major result as peak shape is often neglected when denoising, which can give misinterpreted data. A correction step is thus needed if Gaussian/Lorentzian ratio of denoised peak is around 0.5.

When used carefully, SVD can lead to similar results between denoised spectra at $PSNR_{rms} = 6.6$ and noisy spectra at quantification limit ($PSNR_{rms} = 10$). As $PSNR_{rms}$ is increasing with the square root of time, this difference is equivalent to a considerable gain on acquisition time of 2.3, which is of paramount importance for low sensitivity experiments.

In a second part (II) (*12*), we will focus on the computation time needed for SVD treatment under Java, Matlab and Python, using both processors and graphic cards. We will check the influence of algorithms, especially the divide and conquer one, as well as the influence of single precision calculation will be investigated. Software libraries such as MKL (Intel Math Kernel Library) and hardware capabilities such as SSE3 (Streaming SIMD Extensions) (*129*) will be evaluated. All these optimisations will decrease computation time by a factor of 100.

B.6. Acknowledgements

Sylvie Masse and Cedric Lorthioir are thanked for fruitful discussions.

B.7. Supplementary material

Figure S.V.1: Influence of SVD iterations; Figure S.V.2: SVD applied to 7380 simulated spectra with truncation; Figure S.V.3: Influence of significance level for automatic thresholding; Figure S.V.4: influence of number of scans and preprocessing; simulated data sets and source codes are available online in (*103*).

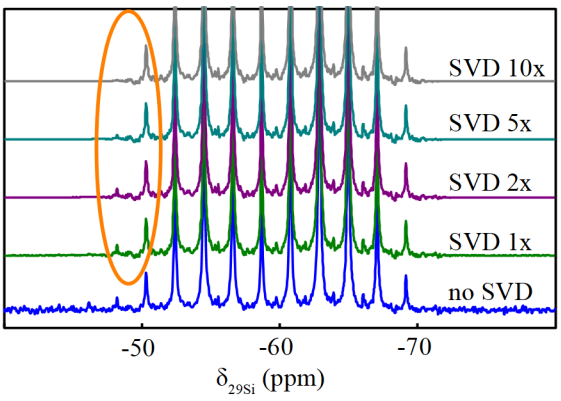


Figure S.V.1: Influence of SVD iterations for a matrix of 3901 × 128 points (n = 128) with k = 22 singular values. 4028 points ^{29}Si CPMG MAS solid-state NMR spectrum of the 50:50 MTEOS:TEOS sample. Orange ellipse shows slight decrease of noise.

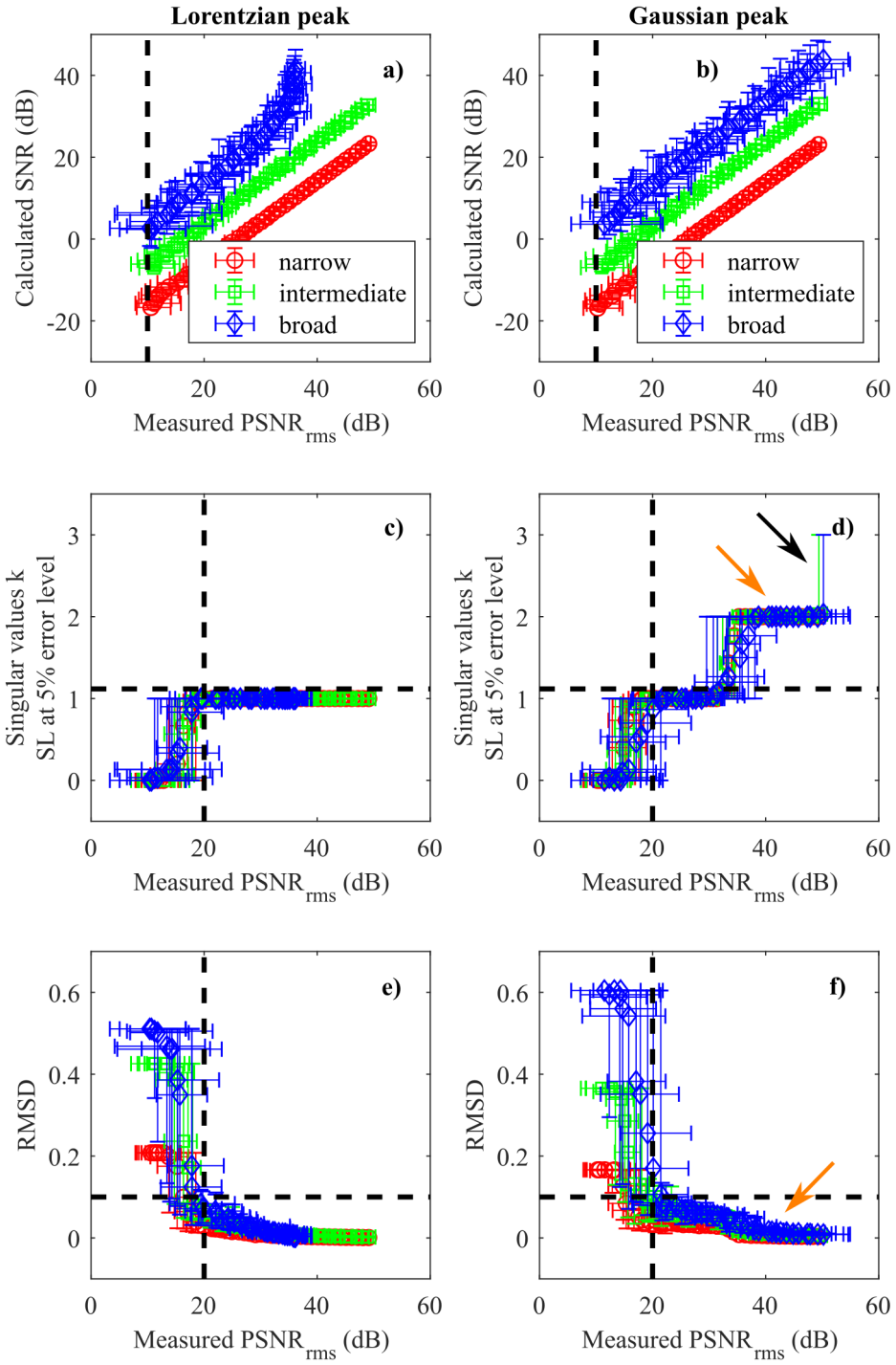


Figure S.V.2: 7380 simulated spectra with known added homoscedastic white Gaussian noise, **with truncation of FID at 5 T_2**. a), c) and e) Lorentzian peak; b), d) and f) Gaussian peak; a) and b) comparison of SNR definitions; c) and d) automatic threshold value with *SL* at 5% error level; e) and f) 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.

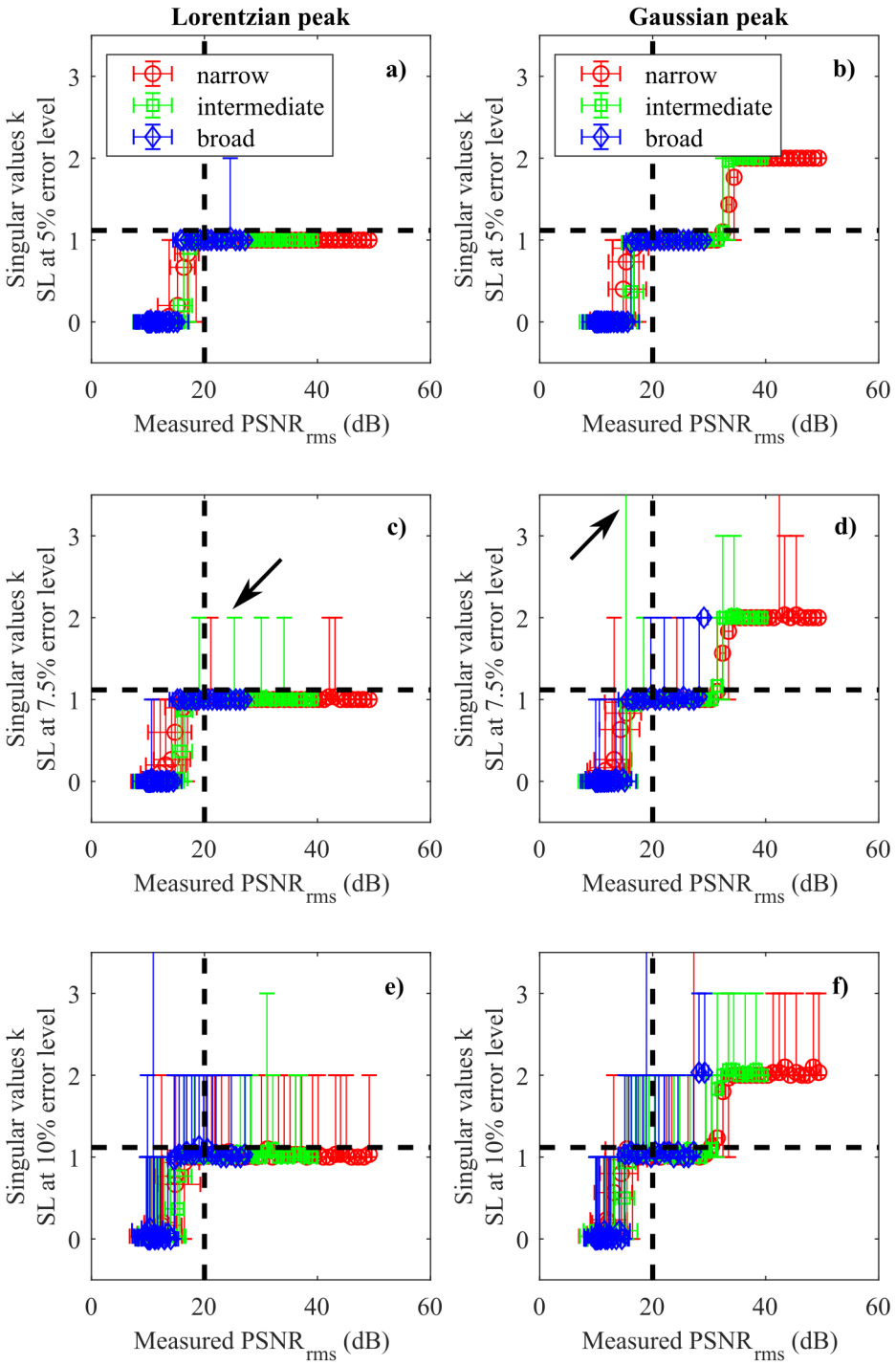


Figure S.V.3: 3 × 7380 simulated spectra with known added homoscedastic white Gaussian noise. a), c) and e) Lorentzian peak; b), d) and f) Gaussian peak; automatic threshold value with a) and b) *SL*=5 %; c) and d) *SL*=7.5 %; e) and f) *SL*=10 %. Error bars correspond to the repetition of 30 simulated 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 (*103*)

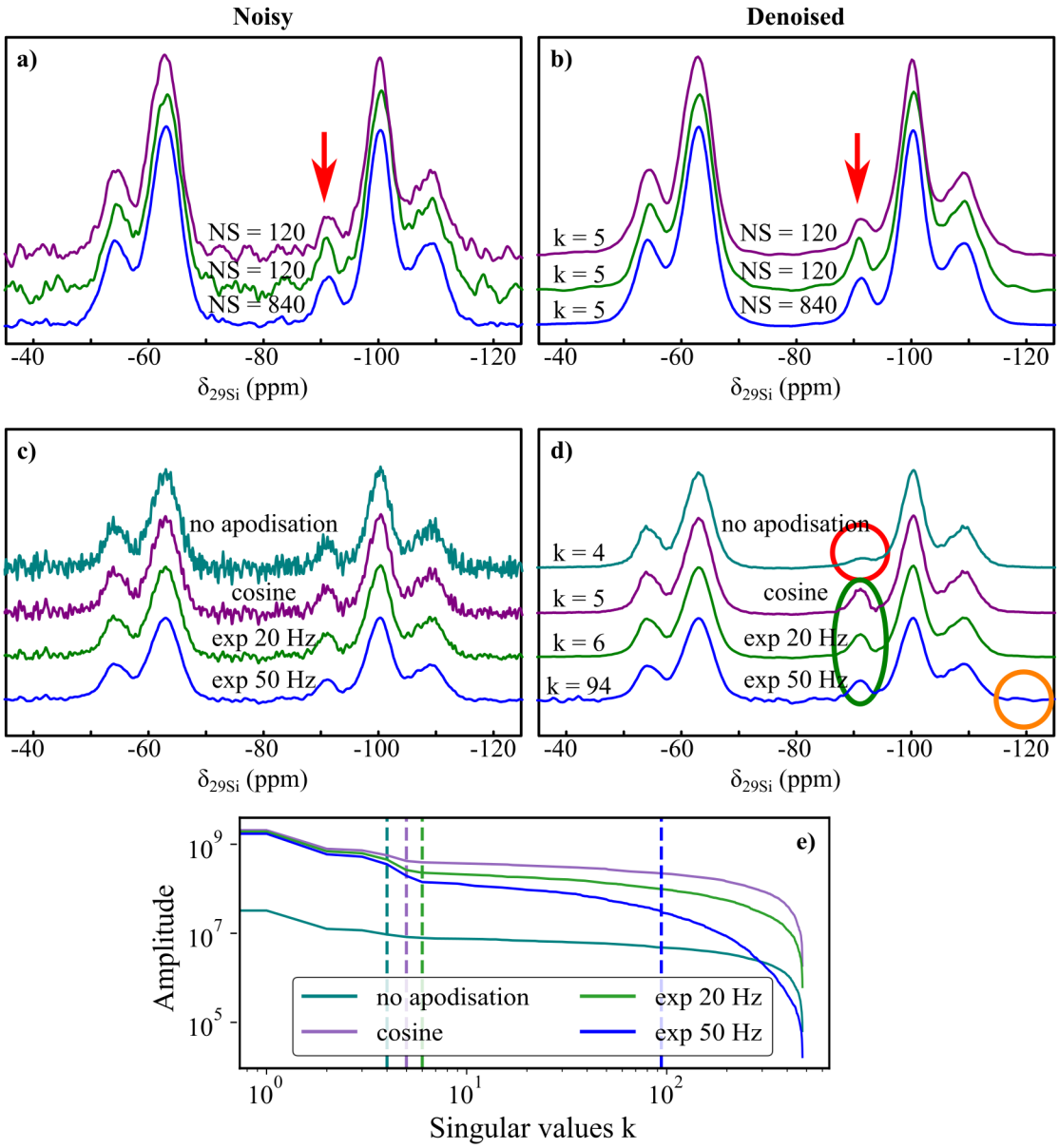


Figure S.V.4: a) and b) Influence of number of scans (NS) on SVD with manual thresholding and 50 Hz exponential apodisation; c) and d) influence of pre-processing before SVD with automatic thresholding and apodisation as annotated on spectra; a) and c) noisy spectra; b) and d) denoised spectra; e) singular 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 ellipses highlight not detected and detected peaks, respectively; orange circle shows ineffective denoising; dashed vertical lines correspond to automatic thresholding of data with the same colour, at an *SL* error level of 7.5 %.

C. SVD on two-dimensional spectra

C.1. Processing workflow

In order to apply SVD with apodisation under Bruker TopSpin software, four files were used (Figure V.11). The first file was raw FID, on which oversampled points were suppressed using the *convdta* command (*94*). This created a second file on which the desired apodisation was applied. However, this processed file has to be converted to a pseudo-FID (third file) using the *genfid* command, on which SVD was applied, giving the last denoised FID file usable for FT. We recently wrote a program to automatise these steps (*130*). TopSpin (in Java) called a first Jython program (Python for Java) that itself launched a CPython script (standard Python, based on C language). Julien Trebosc from UCCS laboratory at Lille University is aknowledged for providing this subprocess call.

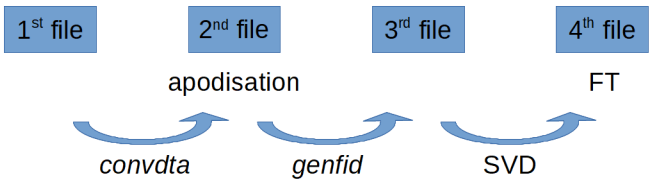


Figure V.11: Process to apply SVD denoising with apodisation on an NMR spectrum. *convdta* and *genfid* are Bruker TopSpin commands.

C.2. Denoising of 2D matrix

In the continuity of the article in previous subchapter, we applied SVD on a 2D NMR spectrum, directly on the entire 2D matrix in time domain. The corresponding spectrum was a $^{1}H \rightarrow ^{29}Si \rightarrow ^{1}H$ double Cross-Polarisation (CP) HETeronuclear CORrelation (HETCOR) (Figure V.12a) (*131*) of a partially dehydrated SBA-15 mesoporous silica (Section E.7 of Chapter I). This sample can be used for instance to entrap pharmaceutical molecules for controlled delivery into human body.

Unfortunately, *convdta* in 2D introduced an artefact with a negative region for each peak in indirect dimension, highlighted by a red arrow (Figure V.12b). However, *convdta* was a necessary step before denoising on time domain, which gave better results than frequency domain, as we confirmed on 2D spectra, similarly to 1D spectra (Subsection B.4.d). A not tested workaround could be to apply an inverse FT on the spectrum, before SVD. This would avoid *convdta* step.

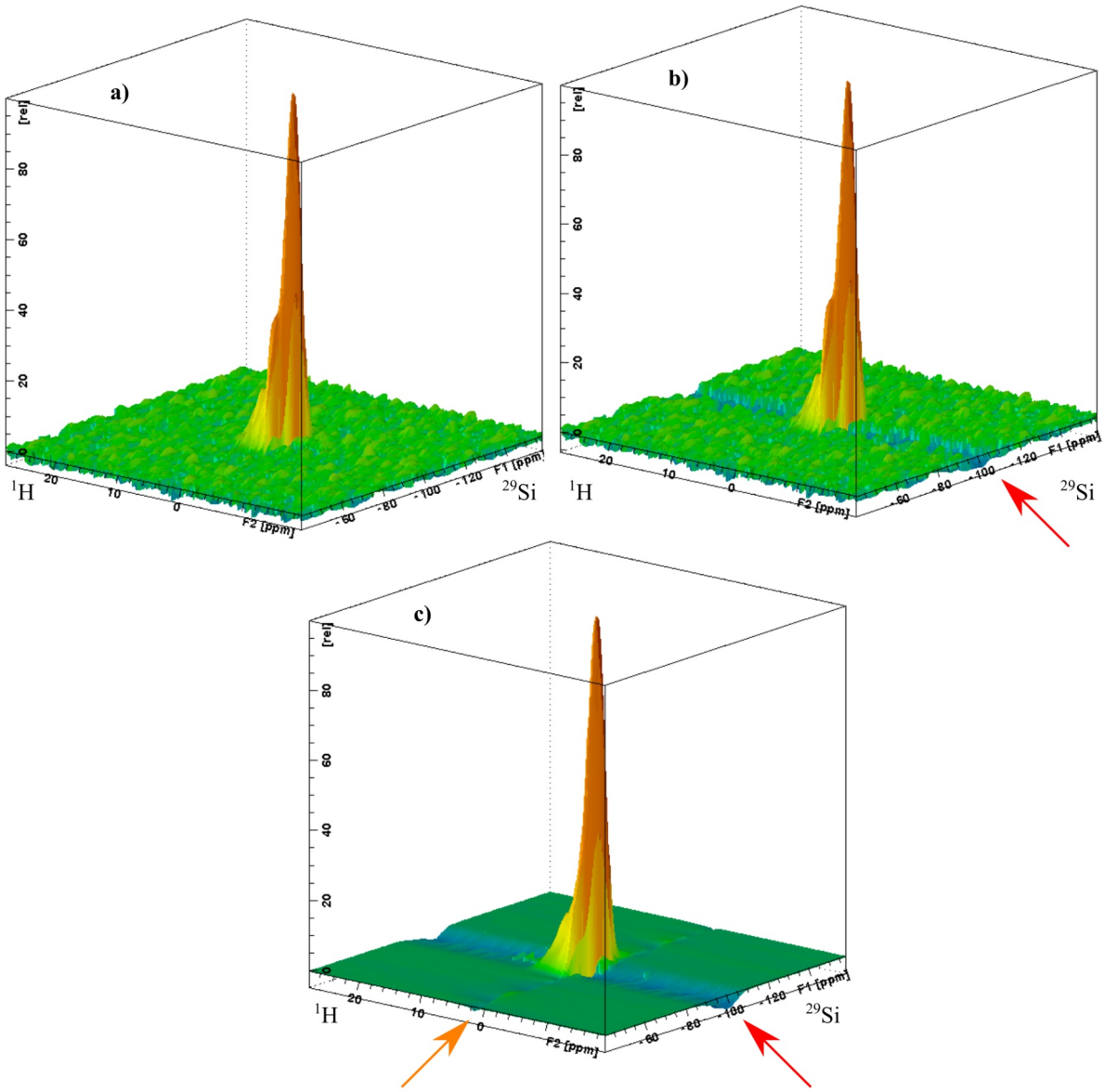


Figure V.12: Influence of *convdta* on a 2D spectrum: a) FT applied on noisy spectrum, b) FT applied after *convdta*, c) FT applied after *convdta* and SVD, 512 × 32 points, $k = 1$ singular value. 2D ^{1}H→^{29}Si→^{1}H HETCOR spectrum of partially dehydrated SBA-15 silica. Red and orange arrows: artefacts. Parameters: $B_0 = 7.0$ T, *MAS* = 14 kHz, *RD* = 1s, *NS* = 2496, $TD(^{29}Si)$ =16 complex slices, $t_{exp} = 22$ h, $SW(^{29}Si) = 7.0$ kHz, $t_{cp1} = 3$ ms, $t_{cp2} = 5$ ms, no decoupling, cosine apodisation.

A similar but lower artefact was present in direct dimension, evidenced after SVD (orange arrow on Figure V.12c). An explanation could be a disturbance of SVD algorithm due to the double oscillation of FID in t_1 and t_2 dimensions. Furthermore, despite good overall denoising, detailed analysis of slices revealed that noise was proportional to signal intensity, similarly to t_1-noise. Hence, noise was only marginally reduced on the highest peak (Figure V.13a), with $PSNR_{max} = 4.4$ and 5.0 for noisy and denoised spectrum, respectively, measured on ^{1}H shoulder (dotted vertical black line). Moreover, signal intensities were distorted, as

highlighted by red ellipses on difference spectra between denoised and noisy spectra. Negative values indicated a decrease of signal intensities, that were especially visible in indirect dimension (Figure V.13b).

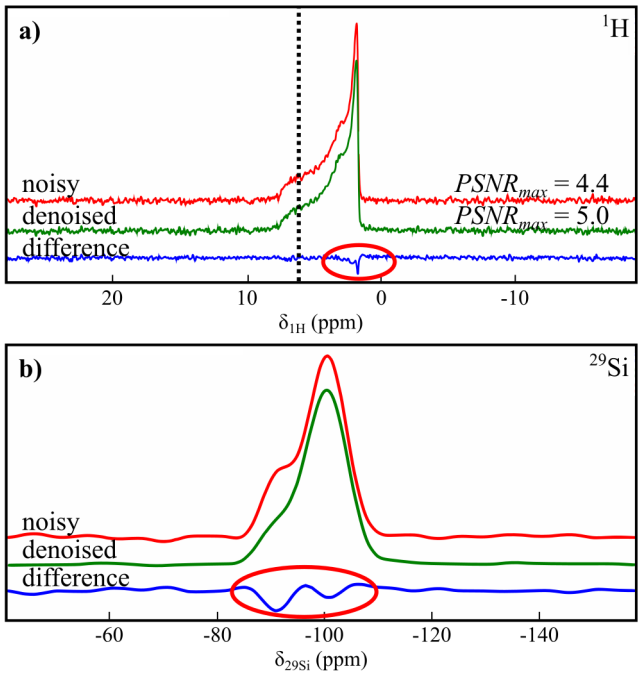


Figure V.13: Extracted slices from Figure V.12c on highest peak in a) ^{1}H direct and b) ^{29}Si indirect dimensions. Red ellipses: distortions; dotted vertical black line: peak used for ^{1}H $PSNR_{max}$ measurement. Parameters: B_0 = 7.0 T, MAS = 14 kHz, RD = 1s, NS = 2496, $TD(^{29}Si)$ =16 complex slices, t_{exp} = 22 h, $SW(^{29}Si)$ = 7.0 kHz, t_{cp1} = 3 ms, t_{cp2} = 5 ms, no decoupling, cosine apodisation.

Resampling of the dataset to a square, instead of the original 512 × 32 matrix, was tested but it did not improved results. Even worth, triangular oscillations appeared on 2D peak shape (not shown). These results depicted that additional optimisations are needed before confident use of 2D spectra denoising in our program. A solution could be to apply SVD to each column of the data matrix after FT along direct dimension. This solution was used by Brissac *et al.* to remove t_1-noise (*85*), and by Qu *et al.* for Non-Uniform Sampling (NUS) low-rank reconstruction (*132*).

In Subchapter C, we detailed the processing workflow to apply SVD under TopSpin. We evidenced an artefact due to *condta* and that 2D denoising degraded quantification.

D. Chapter conclusion

In Subchapter B, we analysed signal processing in the context of denoising by SVD and low-rank approximation. However, these mathematical operations have to be taken with caution, especially in solid-state NMR, in order to avoid pitfalls. We highlighted that denoising is more efficient in time domain, on square matrices, and with automatic thresholding. Gaussian peaks are especially tricky, as SVD systematically overestimated their area by ~ 20 %. Applied to solid-state NMR, quantitative results without any loss of information are obtained at $PSNR_{max} = 2$ with SVD instead of $PSNR_{max} = 3$ with standard processing, corresponding to a time gain of ~ 2.3. This observation on ^{29}Si is generalisable to any nucleus studied by NMR. Further sensitivity gain could be obtained either by combining SVD with Savitzky–Golay filtering (*127*) or with deep neural network (*133*). However, their robustness to real and extracted errors has still to be characterised. In Subchapter C, we applied SVD on 2D NMR spectra on the entire matrix in time domain. We showed that analogic data conversion introduced an artefact and that t_1-noise-like distortions were present after SVD processing, with peaks deformations, attributed to the double time oscillation. This could be solved by applying SVD in indirect dimension after FT in direct dimension.

E. Chapter bibliography

1. S. Sibisi, J. Skilling, R. G. Brereton, E. D. Laue, J. Staunton, *Nature*. **311**, 446–447 (1984).
2. P. C. Kelly, G. Horlick, *Anal. Chem.* **46**, 2130–2136 (1974).
3. A. Rouh, A. Louis-Joseph, J.-Y. Lallemand, *J. Biomol. NMR*. **4**, 505–518 (1994).
4. G. L. Bretthorst, *J. Magn. Reson. 1969*. **88**, 571–595 (1990).
5. K. Krishnamurthy, *Magn. Reson. Chem.* **51**, 821–829 (2013).
6. Y. Matviychuk, E. von Harbou, D. J. Holland, *J. Magn. Reson.* **285**, 86–100 (2017).
7. V. Klema, A. Laub, *IEEE Trans. Autom. Control*. **25**, 164–176 (1980).
8. X. C. Chen, Yu. A. Litvinov, M. Wang, Q. Wang, Y. H. Zhang, *Phys. Rev. E*. **99**, 063320 (2019).
9. K. F. Sheberstov, E. S. Guardiola, M. Pupier, D. Jeannerat, *Magn. Reson. Chem.* **58**, 466–472 (2020).
10. P. P. Man, C. Bonhomme, F. Babonneau, *Solid State Nucl. Magn. Reson.* **61–62**, 28–34 (2014).
11. G. Laurent, W. Woelffel, V. Barret-Vivin, E. Gouillart, C. Bonhomme, *Appl. Spectrosc. Rev.* **54**, 602–630 (2019).
12. G. Laurent, P.-A. Gilles, W. Woelffel, V. Barret-Vivin, E. Gouillart, C. Bonhomme, *Appl. Spectrosc. Rev.* **55**, 173–196 (2020).
13. C. Bonhomme, C. Gervais, D. Laurencin, *Prog. Nucl. Magn. Reson. Spectrosc.* **77**, 1–48 (2014).
14. S. Lohumi, M. S. Kim, J. Qin, B.-K. Cho, *TrAC Trends Anal. Chem.* **93**, 183–198 (2017).

15. M. H. Levitt, *Spin dynamics: basics of nuclear magnetic resonance* (John Wiley & Sons Ltd, Chichester, England, 2nd ed., 2008).
16. D. I. Hoult, R. E. Richards, *J. Magn. Reson. 1969*. **24**, 71–85 (1976).
17. C. Heitz, G. Laurent, R. Briard, E. Barthel, *J. Colloid Interface Sci.* **298**, 192–201 (2006).
18. D. D. Laws, H.-M. L. Bitter, A. Jerschow, *Angew. Chem. Int. Ed.* **41**, 3096–3129 (2002).
19. E. R. Andrew, A. Bradbury, R. G. Eades, *Nature*. **182**, 1659–1659 (1958).
20. I. J. Lowe, *Phys. Rev. Lett.* **2**, 285–287 (1959).
21. D. Cala-De Paepe *et al.*, *Solid State Nucl. Magn. Reson.* **87**, 126–136 (2017).
22. A. Pines, M. G. Gibby, J. S. Waugh, *J. Chem. Phys.* **56**, 1776–1777 (1972).
23. V. S. Mithu, S. Pratihar, S. Paul, P. K. Madhu, *J. Magn. Reson.* **220**, 8–17 (2012).
24. L. Frydman, *J. Magn. Reson.* **242**, 256–264 (2014).
25. J. A. Lehmann-Horn, J.-F. Jacquinot, J. C. Ginefri, C. Bonhomme, D. Sakellariou, *J. Magn. Reson.* **271**, 46–51 (2016).
26. K.-N. Hu, C. Song, H. Yu, T. M. Swager, R. G. Griffin, *J. Chem. Phys.* **128**, 052302, 1–17 (2008).
27. A. J. Rossini *et al.*, *Acc. Chem. Res.* **46**, 1942–1951 (2013).
28. J. R. Yarava, S. R. Chaudhari, A. J. Rossini, A. Lesage, L. Emsley, *J. Magn. Reson.* **277**, 149–153 (2017).
29. H. Takahashi, B. Viverge, D. Lee, P. Rannou, G. De Paëpe, *Angew. Chem. Int. Ed.* **52**, 6979–6982 (2013).
30. R. Gautam *et al.*, *Curr. Sci.* **108**, 341–356 (2015).
31. N. Hayazawa, Y. Inouye, Z. Sekkat, S. Kawata, *J. Chem. Phys.* **117**, 1296–1301 (2002).
32. P. Vítek *et al.*, *Spectrochim. Acta. A. Mol. Biomol. Spectrosc.* **86**, 320–327 (2012).
33. M. Fleischmann, P. J. Hendra, A. J. McQuillan, *Chem. Phys. Lett.* **26**, 163–166 (1974).
34. U. K. Sur, J. Chowdhury, *Curr. Sci.* **105**, 923–939 (2013).
35. R. S. Das, Y. K. Agrawal, *Vib. Spectrosc.* **57**, 163–176 (2011).
36. B. Ren, G.-K. Liu, X.-B. Lian, Z.-L. Yang, Z.-Q. Tian, *Anal. Bioanal. Chem.* **388**, 29–45 (2007).
37. U. Laor, G. C. Schatz, *Chem. Phys. Lett.* **82**, 566–570 (1981).
38. R. M. Stöckle, Y. D. Suh, V. Deckert, R. Zenobi, *Chem. Phys. Lett.* **318**, 131–136 (2000).
39. Z.-Q. Tian, B. Ren, J.-F. Li, Z.-L. Yang, *Chem. Commun.*, 3514–3534 (2007).
40. M. D. Duncan, J. Reintjes, T. J. Manuccia, *Opt. Lett.* **7**, 350–352 (1982).
41. H. Kano, H. Segawa, P. Leproux, V. Couderc, *Opt. Rev.* **21**, 752–761 (2014).
42. J. L. Perez-Rodriguez, M. D. Robador, M. A. Centeno, B. Siguenza, A. Duran, *Spectrochim. Acta. A. Mol. Biomol. Spectrosc.* **120**, 602–609 (2014).
43. C. Krafft, M. Kirsch, C. Beleites, G. Schackert, R. Salzer, *Anal. Bioanal. Chem.* **389**, 1133–1142 (2007).
44. R. Hardis, J. L. P. Jessop, F. E. Peters, M. R. Kessler, *Compos. PartA - Appl. Sci. Manuf.* **49**, 100–108 (2013).
45. C. L. Byrne, lecture notes (Department of Mathematical Sciences, University of Massachusetts Lowell, USA, 2013), available at http://faculty.uml.edu/cbyrne/SP1text.pdf.
46. R. R. Ernst, W. A. Anderson, *Rev. Sci. Instrum.* **37**, 93–102 (1966).
47. R. Freeman, *Anal. Chem.* **65**, 743A-753A (1993).
48. E. Kupče, T. Nishida, R. Freeman, *Prog. Nucl. Magn. Reson. Spectrosc.* **42**, 95–122 (2003).

49. M. Lustig, D. L. Donoho, J. M. Santos, J. M. Pauly, *IEEE Signal Process. Mag.* **25**, 72–82 (2008).
50. M. Mobli, J. C. Hoch, *Prog. Nucl. Magn. Reson. Spectrosc.* **83**, 21–41 (2014).
51. W. J. Malfait, W. E. Halter, *J. Non-Cryst. Solids*. **354**, 4107–4114 (2008).
52. K. H. Liland, A. Kohler, N. K. Afseth, *J. Raman Spectrosc.* **47**, 643–650 (2016).
53. A. E. Kandjani *et al.*, *J. Raman Spectrosc.* **44**, 608–621 (2013).
54. O. Ryabchykov *et al.*, *Chemom. Intell. Lab. Syst.* **155**, 1–6 (2016).
55. H. G. Schulze *et al.*, *J. Raman Spectrosc.* **43**, 360–369 (2012).
56. P. Chatterjee, P. Milanfar, *IEEE Trans. Image Process.* **19**, 895–911 (2010).
57. A. Perriot *et al.*, *J. Am. Ceram. Soc.* **89**, 596–601 (2006).
58. M. C. Davis *et al.*, *J. Phys. Chem. A*. **114**, 5503–5508 (2010).
59. J. C. Lindon, A. G. Ferrige, *Prog. Nucl. Magn. Reson. Spectrosc.* **14**, 27–66 (1980).
60. M. Člupek, P. Matějka, K. Volka, *J. Raman Spectrosc.* **38**, 1174–1179 (2007).
61. D. L. Donoho, I. M. Johnstone, A. S. Stern, J. C. Hoch, *Proc. Natl. Acad. Sci.* **87**, 5066–5068 (1990).
62. K. Takeda, in *Annual Reports on NMR Spectroscopy*, G. A. Webb, Ed. (Academic Press, 2015), vol. 84, pp. 77–113.
63. S. Chen *et al.*, *Opt. Express*. **22**, 12102–12114 (2014).
64. D. Barache, J.-P. Antoine, J.-M. Dereppe, *J. Magn. Reson.* **128**, 1–11 (1997).
65. L. Chiron, M. A. van Agthoven, B. Kieffer, C. Rolando, M.-A. Delsuc, *Proc. Natl. Acad. Sci.* **111**, 1385–1390 (2014).
66. D. W. Tufts, R. Kumaresan, I. Kirsteins, *Proc. IEEE*. **70**, 684–685 (1982).
67. J. A. Cadzow, *IEEE Trans. Acoust. Speech Signal Process.* **36**, 49–62 (1988).
68. D. W. Tufts, A. A. Shah, *IEEE Trans. Signal Process.* **41**, 1716–1721 (1993).
69. G. W. Stewart, *SIAM Rev.* **35**, 551–566 (1993).
70. E. Beltrami, *G. Mat. Ad Uso Degli Stud. Delle Univ.* **11**, 98–106 (1873).
71. C. Jordan, *J. Mathématiques Pures Appliquées*. **19**, 35–54 (1874).
72. C. Lanczos, *Am. Math. Mon.* **65**, 665–679 (1958).
73. P. Läuchli, *Numer. Math.* **3**, 226–240 (1961).
74. J. A. K. Suykens, *Appl. Comput. Harmon. Anal.* **40**, 600–609 (2016).
75. L. Condat, A. Hirabayashi, *Sampl. Theory Signal Image Process.* **14**, 17–47 (2015).
76. P. C. Hansen, S. H. Jensen, *EURASIP J. Adv. Signal Process.* **2007**, 092953, 1–24 (2007).
77. Y. Chen *et al.*, *Geophys. J. Int.* **206**, 1695–1717 (2016).
78. E. S. Park, R. C. Henry, C. H. Spiegelman, Technical report, NRCSE-TRS No. 034, *NRCSE technical report series* (The National Research Center for Statistics and the Environment, USA, 1999), available at http://www.nrcse.washington.edu/pdf/trs34_receptor.pdf.
79. G. D. Clifford, lecture notes (HST.582J Biomedical Signal and Image Processing. Massachusetts Institute of Technology: MIT OpenCourseWare, 2005), available at http://www.mit.edu/~gari/teaching/6.222j/ICASVDnotes.pdf.
80. M. Aharon, M. Elad, A. Bruckstein, *IEEE Trans. Signal Process.* **54**, 4311–4322 (2006).
81. M. Baumann, Master thesis (Institut für Kommunikationstechnik, Eidgenössische Technische Hochschule, Zürich, Switzerland, 2010), available at http://dx.doi.org/10.3929/ethz-a-007215003.
82. H. M. Nguyen, PhD dissertation (Electrical & Computer Eng, University of Illinois, Urbana-Champaign, USA, 2011), available at http://hdl.handle.net/2142/29753.
83. H. G. Schulze, R. F. B. Turner, *Appl. Spectrosc.* **69**, 643–664 (2015).
84. A. Tulloch (2014), available at https://research.fb.com/fast-randomized-svd/.

85. C. Brissac, T. E. Malliavin, M. A. Delsuc, *J. Biomol. NMR*. **6**, 361–365 (1995).
86. P. Stoica, N. Sandgren, Y. Selén, L. Vanhamme, S. Van Huffel, *J. Magn. Reson.* **165**, 80–88 (2003).
87. O. Nikel *et al.*, *J. Phys. Chem. C*. **116**, 6320–6331 (2012).
88. Y. Zou, R. Xie, *Comput. Geosci.* **19**, 389–401 (2015).
89. N. Uzunbajakava *et al.*, *Biophys. J.* **84**, 3968–3981 (2003).
90. J. Palacký, P. Mojzeš, J. Bok, *J. Raman Spectrosc.* **42**, 1528–1539 (2011).
91. A. Khmaladze *et al.*, *Appl. Spectrosc.* **68**, 1116–1122 (2014).
92. M. C. Caraher *et al.*, *Biochim. Biophys. Acta - Mol. Basis Dis.* **1864**, 398–406 (2018).
93. P. P. Man (2012), available at http://pascal-man.com/navigation/faq-java-browser/SVD-Java-application2012.shtml.
94. P. P. Man (2012), available at http://www.pascal-man.com/navigation/faq-java-browser/SVD-Java-application-GPU.shtml.
95. J. Nickolls, I. Buck, M. Garland, K. Skadron, *Queue*. **6**, 40–53 (2008).
96. G. Laurent, W. Woelffel, V. Barret-Vivin, E. Gouillart, C. Bonhomme, in *Actes du C2i-2016 : 7ème Colloque Interdisciplinaire en Instrumentation* (Saint-Nazaire, France, 2016), pp. 1–8.
97. R. B. Figueira, I. R. Fontinha, C. J. R. Silva, E. V. Pereira, *Coatings*. **6**, 12, 1–19 (2016).
98. G. J. Owens *et al.*, *Prog. Mater. Sci.* **77**, 1–79 (2016).
99. N. Ben Ahmed *et al.*, *Anal. Bioanal. Chem.* **410**, 1205–1216 (2018).
100. M. Faustini *et al.*, *Chem. Mater.* **22**, 4406–4413 (2010).
101. B. M. Fung, A. K. Khitrin, K. Ermolaev, *J. Magn. Reson.* **142**, 97–101 (2000).
102. F. Pedregosa *et al.*, *J. Mach. Learn. Res.*, in press.
103. G. Laurent, P.-A. Gilles, W. Woelffel, V. Barret-Vivin, E. Gouillart, C. Bonhomme (2018), available at http://doi.org/10.5281/zenodo.1406172.
104. L. Petrakis, *J. Chem. Educ.* **44**, 432–436 (1967).
105. G. Golub, W. Kahan, *J. Soc. Ind. Appl. Math. Ser. B Numer. Anal.* **2**, 205–224 (1965).
106. D. E. Brown, T. W. Campbell, *J. Magn. Reson. 1969*. **89**, 255–264 (1990).
107. G. Heinig, K. Rost, in *Algebraic methods for Toeplitz-like matrices and operators* (Birkhäuser Basel, 1984), vol. 13 of *Operator Theory: Advances and Applications*, pp. 9–135.
108. H. Karner, J. Schneid, C. W. Ueberhuber, *Linear Algebra Its Appl.* **367**, 301–311 (2003).
109. R. Ross, *IEEE Trans. Dielectr. Electr. Insul.* **3**, 28–42 (1996).
110. N. Priyadarshani, S. Marsland, I. Castro, A. Punchihewa, *PLoS ONE*. **11**, e0146790, 1–26 (2016).
111. A. Shrivastava, V. B. Gupta, *Chron. Young Sci.* **2**, 21–25 (2011).
112. S. G. Hyberts, S. A. Robson, G. Wagner, *J. Biomol. NMR*. **55**, 167–178 (2013).
113. M. E. Lacey, R. Subramanian, D. L. Olson, A. G. Webb, J. V. Sweedler, *Chem. Rev.* **99**, 3133–3152 (1999).
114. L. A. Currie, *Anal. Chem.* **40**, 586–593 (1968).
115. F. H. Larsen, J. Skibsted, H. J. Jakobsen, N. Chr. Nielsen, *J. Am. Chem. Soc.* **122**, 7080–7086 (2000).
116. H. Hassani, *J. Data Sci.* **5**, 239–257 (2007).
117. S. Van Huffel, H. Chen, C. Decanniere, P. Vanhecke, *J. Magn. Reson. A*. **110**, 228–237 (1994).
118. N. M. Faber, L. M. C. Buydens, G. Kateman, *Chemom. Intell. Lab. Syst.* **25**, 203–226 (1994).

119. E. R. Malinowski, *Factor analysis in chemistry* (Wiley, Hoboken, NJ, USA, 3rd ed., 2002).
120. J.-F. Cai, E. J. Candès, Z. Shen, *SIAM J. Optim.* **20**, 1956–1982 (2010).
121. K. Y. Sanliturk, O. Cakar, *Mech. Syst. Signal Process.* **19**, 615–631 (2005).
122. Maxim (2016), available at http://fr.mathworks.com/matlabcentral/fileexchange/52321-peak-fitting-to-either-voigt-or-lognormal-line-shapes.
123. Z. Dong, W. Dreher, D. Leibfritz, B. S. Peterson, *Am. J. Neuroradiol.* **30**, 1096–1101 (2009).
124. A. N. Garroway, *J. Magn. Reson. 1969*. **28**, 365–371 (1977).
125. A. S. Lipton, J. A. Sears, P. D. Ellis, *J. Magn. Reson.* **151**, 48–59 (2001).
126. M. R. Palmer *et al.*, *J. Phys. Chem. B*. **119**, 6502–6515 (2015).
127. H. Hassanpour, A. Zehtabian, S. J. Sadati, *Digit. Signal Process.* **22**, 786–794 (2012).
128. D. Massiot *et al.*, *Magn. Reson. Chem.* **40**, 70–76 (2002).
129. J. Hofmann, J. Treibig, G. Hager, G. Wellein, in *Proceedings of the 2014 Workshop on Programming Models for SIMD/Vector Processing* (ACM, New York, NY, USA, 2014), *WPMVP '14*, pp. 57–64.
130. G. Laurent, P.-A. Gilles (2017), available at https://github.com/gul916/NMR_post_proc.
131. N. Baccile, G. Laurent, C. Bonhomme, P. Innocenzi, F. Babonneau, *Chem. Mater.* **19**, 1343–1354 (2007).
132. X. Qu, M. Mayzel, J.-F. Cai, Z. Chen, V. Y. Orekhov, *Angew. Chem. Int. Ed.* **54**, 852–854 (2015).
133. S. K. Devalla *et al.*, (2018), available at http://arxiv.org/abs/1809.10589.

Chapter VI. Decreasing computation time

'The speed at which modern CPUs perform computations still blows my mind daily.'

Markus Persson (1979-), businessman

A. Chapter introduction

Despite being very powerful, Singular Value Decomposition (SVD) is a computationally expensive mathematical operation, as it can take up to thousands of seconds to complete. Indeed, it is using matrix-matrix operations with a computational complexity of $O(n^3)$, much higher than matrix-vector ($O(n^2)$) or vector-vector ($O(n)$) operations (*1*). Various approaches were developed to decrease computation time, such as random projection on a subspace (*2*). Randomized QR decomposition (*3*) and uncoiled random QR denoising (*4*) proved to be faster and more robust than SVD. Q is an orthogonal matrix whereas R is an upper triangular matrix (*5*). Here we rather focussed on computation time decrease of standard SVD, which can be generalised to other denoising methods.

Our preliminary tests using SVD evidenced the need first to properly define the suitable hardware, second to explain why different computation times were observed with different software (Java, Matlab and Python) for similar data sets. Trying to answer these two questions led to the article presented in Section B (*6*), second part of this two-steps article. We tried to decorrelate the influence of all elements between experimenter SVD call and hardware, namely algorithm, libraries and hardware instructions usage. After providing some useful benchmarks under Java for hardware ranging over ten years, we studied influence of software version, precision and algorithms under Matlab. Finally, we demonstrated under Python that proper call to hardware instructions could lead to a three-fold time reduction. By combining all these software optimisations, computation time was drastically decreased by 100, on the same hardware. We highlighted that, for matrices smaller than $4097 \times 4096 = 1.7.10^7$ points, SVD is faster on a Central Processing Unit (CPU) using MKL (*7*) than on a GPU using CULA (*8*). We provided an optimised Python program to denoise a matrix with SVD, combining MKL and CULA, with automatic CPU/GPU selection. Another Python program is in charge of importing NMR Bruker dataset and of converting it into a matrix before SVD. In Section C, additional computation tests will be undergone on precision and on heterogeneous computing.

B. Denoising applied to spectroscopies – part II: decreasing computation time, *Appl. Spectrosc. Rev.* 55, 173–196 (2020).

Spectroscopies are of fundamental importance but can suffer from low sensitivity. Singular Value Decomposition (SVD) is a highly interesting mathematical tool, which can be conjugated with low-rank approximation to denoise spectra and increase sensitivity. SVD is also involved in data mining with Principal Component Analysis (PCA). In this paper, we focussed on the optimisation of SVD duration, which is a time-consuming computation. Both Intel processors (CPU) and Nvidia graphic cards (GPU) were benchmarked. A 100 times gain was achieved when combining divide and conquer algorithm, Intel Math Kernel Library (MKL), SSE3 (Streaming SIMD Extensions) hardware instructions and single precision. In such a case, the CPU can outperform the GPU driven by CUDA technology. These results give a strong background to optimise SVD computation at the user scale.

Keywords: spectroscopy, signal processing, Cadzow denoising, Singular Value Decomposition (SVD), benchmarking

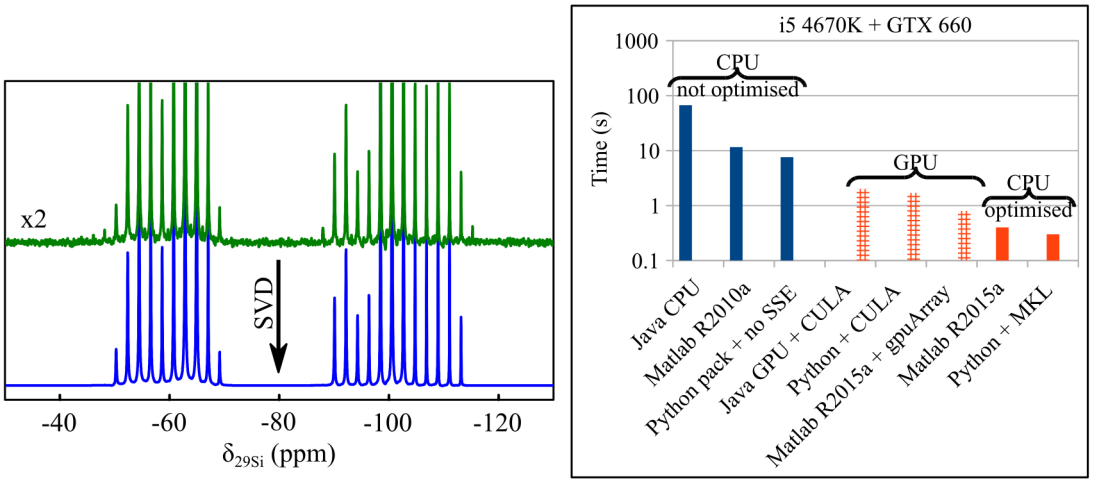


Figure VI.1: Graphical abstract of (*6*).

B.1. Introduction

Spectrocopies are very efficient tools to help scientists in various domains: physics, chemistry, biology or medecine. However, the intrinsic sensitivity is highly dependant of the technique used. In particular, Nuclear Magnetic Resonance (NMR) (*9*) and Raman spectroscopy (*10*) are poorly sensitive but are very precise local probes commonly used to

study materials. While NMR can detect only one nucleus over 10^5 in usual conditions (*11*), Raman spectroscopy can only detect one photon over 10^6 (*12*). Both NMR and Raman sensitivity were improved over the years, as was detailed in part (I) of this study (*13*), but still suffer from low signal-to-noise ratio (SNR), especially when studying amorphous or non-stable materials. Additionally, different physicochemical techniques can be hyphenated to obtain multiple signatures in a single experiment (*14*, *15*).

This sensitivity gain entails an increasing amount of data to analyse, especially in the domain of metabolomics (*16*, *17*). In such a case, it is necessary to combine statistics and chemistry, what is called chemometrics (*18*, *19*). A similar problem of data mining is present in social engineering (*20*) or with medical images dictionaries (*21*). Many tools are available to process these data: Principal Component Analysis (PCA) (*22*), Principal Component Regression (PCR) (*23*), Partial Least Squares (PLS) (*24*), Discriminant Analysis (PLS-DA) (*25*), Independent Component Analysis (ICA) (*26*), or Non-negative Matrix Factorisation (NMF) (*27*). The aim of these multivariate data analysis methods is to find relevant parameters in order to discriminate samples (ex wine from region A or B (*28*)). These tools are of paramount importance to apply data mining to spectroscopies (*29*). For instance, PCA was used recently with Gas Chromatography/Quadrupole Time-of-Flight (GC/Q-ToF) mass spectrometry (*30*), Mid-InfraRed (MIR) spectroscopy (*31*) and Inductively Coupled Plasma Optical Emission Spectroscopy (ICP-OES) (*32*). An important step of PCA and relative techniques is Singular Value Decomposition (SVD) (*33*, *34*).

Moreover, SVD was proposed as a method to denoise signals by Tufts *et al* (*35*), and was generalised by Cadzow in 1988 (*36*). In the context of low sensitivity and in the continuation of a previous communication (*37*), we thoroughly described SVD in part (I) of this work (*13*). We first gave theoretical background on SVD, low-rank approximation (*38*), Hankel or Toeplitz matrices (*39*) and SNR definitions. SVD was applied to Raman and NMR spectra. We highlighted that best results were obtained with square matrices and data in time domain rather than in frequency domain. Automatic thresholding was applied thanks to Malinowki's significant level indicator (*40*). $6 \times 7380 = 44280$ denoised spectra with known noise were compared to their non-noisy counterparts. It was evidenced that the minimum peak *SNR* measured on maximum of noise ($PSNR_{max}$) needed to have reliable results was $PSNR_{max} = 2$, leading to a gain on acquisition time of 2.3. Surprisingly, while Lorentzian peaks

were correctly denoised, SVD transformed Gaussian peaks into intermediate Gaussian/Lorentzian ones, which overestimated their peak area by 20 %.

The main disadvantage of SVD is its long computation time, especially for big data sets. It is thus essential to optimise the computation procedure. Different approaches have been chosen in literature: specialised processors (*41*), wavelet transformation before performing the SVD (*42*), divide and conquer method (*43*) or sparse matrices (*44*, *45*). Another approach is General Purpose computing on Graphics Processing Unit (GPGPU) (*46*). SVD has been applied multiple times using GPGPU (*47–50*). Two programming languages are available: CUDA (Compute Unified Device Architecture) for Nvidia graphic cards (*51*) and OpenCL (Open Computing Language) for all graphic cards (Graphic Processing Unit, GPU) and processors (Central Processing Unit, CPU) (*52*). In addition, Dongarra *et al.* developed very efficient algorithms combining CPU and GPU (*53*, *54*), what is called heterogeneous computing (*55*). Recently Man *et al* proposed a Java implementation of SVD for NMR (*56*) using CPU (*57*) or Nvidia GPU (*58*). Pending questions can be raised on these Java applications: (i) how powerful does the computer need to be? (ii) how long does computation take? (iii) how large can be the data set? (iv) is the used algorithm efficient? (v) will other programming languages give better results?

In this second part (II), after providing some experimental details in Subsection B.2, we benchmarked SVD using Java, on various CPU and Nvidia GPU ranging over 10 and 6 years, respectively (see Subsection B.3.a). We focussed on algorithms and precision under Matlab in Subsection B.3.b. We tried to decorrelate software libraries from hardware capabilities (Single Instruction Multiple Data, SIMD) (*59*) in Subsection B.3.c. Finally, we compared Java, Matlab and Python in Subsection B.3.d to reach the fastest possible denoising computation.

B.2. Materials and methods

B.2.a. Solid-state NMR experiments

Two solid-state NMR spectra were used to benchmark SVD. The first one was a ^{29}Si spectrum with 4096 complex points, used for matrices up to $2015 \times 2014 = 4.1e6$. For matrices above this limit, a ^{87}Sr spectrum with 30504 complex points, was chosen. The noisy and denoised spectra for ^{29}Si and ^{87}Sr are presented on Figures VI.2a and VI.2b, respectively.

SVD was applied on Free Induction Decay (FID, time domain) after removal of the first 68 points corresponding to oversampled digitisation .

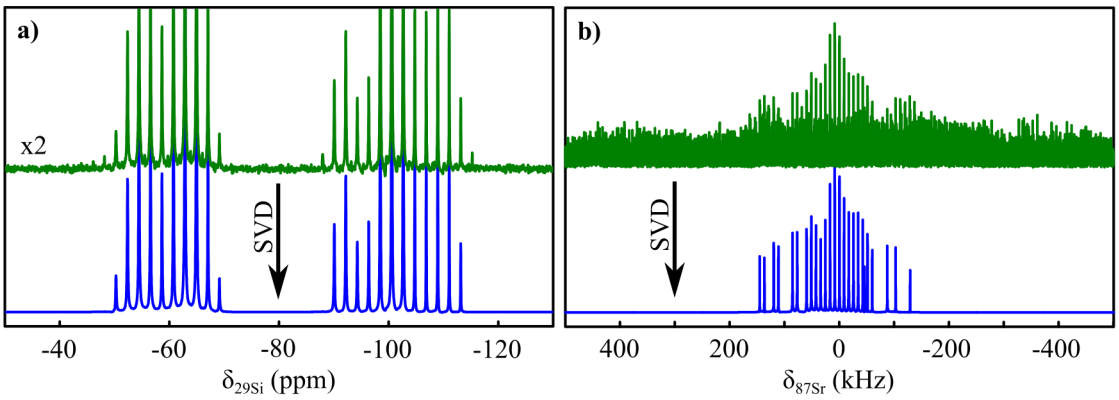


Figure VI.2: Spectra used to benchmark SVD; top: noisy spectra; bottom: denoised spectra. a) ^{29}Si CPMG MAS solid-state NMR spectrum of a 50:50 MTEOS:TEOS sample; processing: Java GPU application with 2015 × 2014 = 4.1e6 points, k = 25 singular values, computation time of 6 s on a GTX 660, cosine apodisation. b) ^{87}Sr DFS-WURST-CPMG solid-state static NMR spectrum of non-hydrated non-protonated $Sr(PO_3F)$; processing: Java GPU application with 4097 × 4096 = 1.7e7 points, k = 25 singular values, computation time of 31 s on a GTX 660, cosine apodisation, magnitude calculation.

The sample analysed by ^{29}Si solid state NMR and the experiments were presented in part (I) of this study (*13*). Briefly, it was representative of sol-gel chemistry, combining hydrophobicity and mechanical properties (*60*). A 50:50 mix of Methyltriethoxysilane (MTEOS) : tetraethylorthosilicate (TEOS) was prepared by spray drying giving spherical micrometer silica particles. Carr-Purcell-Meiboom-Gill (CPMG) Magic Angle Spinning (MAS) experiments (*61*) were performed in 40 minutes on a Bruker Avance III spectrometer operating at 300.29 MHz for ^{1}H and 59.65 MHz for ^{29}Si.

The sample analysed by ^{87}Sr solid state NMR was a model of biocompatible material in relation with bone substitutions (*62*). Non-hydrated non-protonated $Sr(PO_3F)$[1] was studied on a Bruker Avance III spectrometer operating at 699.98 MHz for ^{1}H and 30.34 MHz for ^{87}Sr in a 5 mm static probe. In order to enhance sensitivity, DFS (*63*), WURST (*64*) and CPMG were used with 58,000 transients and a relaxation delay of 300 ms, leading to a total acquisition time of 5.5 h. 260 echoes were acquired with a full echo delay of 0.12 ms.

[1] After publication of this article, we were aware that this spectrum was wrongly attributed. The correct description is a strontium bioactive glass (Section E.8 of Chapter I).

B.2.b. Measurement of SVD computation times

Special care has been taken to validate measured times: the computer was checked to be idle, without any update running and without an active internet browser window, which would use the graphic card through Adobe Flash module. Computations were systematically repeated and results were highly reproducible. Nvidia drivers ranging from 331.113 to 352.30 were used, either under Linux Ubuntu 14.04, CentOS 5, Fedora 21, 22 or Windows XP, 7 SP1 or 8.1. Under Windows it was necessary to modify the TdrLevel registry key to 0 in order to avoid graphics driver failure (*65*). Under Java (Oracle Corporation, Redwood Shores, CA, USA) and Matlab (The MathWorks, Inc., Natick, MA, USA), ^{29}Si FID was used up to 4028 points. Above this value, ^{87}Sr FID was used. The corresponding FID was truncated if needed to the desired data length. Under Python (*66*), data set was a simple list of increasing values with the corresponding length. Unless otherwise stated, $k = 25$ singular values were kept for low-rank approximation. This value corresponded to the major spikelets observed on ^{29}Si spectrum (Figure VI.2a) and was not changed for coherence along the series. The measured delays are the sum of decomposition and low-rank approximation steps, including all processor to graphic card latencies, if relevant.

B.2.b.i. Java

Two applications are available online: one for CPU (*57*) and the other one for Nvidia GPU (*56*, *58*). While CPU version calls JAMPACK library (*67*), GPU version calls CULA R15 (*8*). The CPU 32 bits version failed above a matrix size of 1025 × 1024 = 1.0e6 points, whatever the CPU used. The same problem was observed with GPU 32 bits version above a matrix size of 3005 × 1024 = 3.1e6 points. The reason is that Java heap space is limited to around 1.5 GB for 32 bits applications (*68*), while it is 16 exabytes for 64 bits applications. This memory amount corresponds not only to the data, but also to the program and its libraries. 64 bits applications are not compatible with 32 bits Java runtime environment and 32 bits operating systems. As the source code was not accessible and no internal timer was implemented in these Java applications, computation times were measured with a handheld chronograph giving a time resolution of 1 s for both SVD step and low-rank approximation step. For a same GPU, no computation time difference was observed between Windows and Linux.

B.2.b.ii. Matlab

Three versions were tested: R2010a, corresponding to the most recent compatible with CULA library free (in version R14) (*69*); R2014a, corresponding to the most recent for old graphic cards with Compute Capability (CC) less than 2.0, such as GTX 260; and R2015a, corresponding to a recent version. FID were imported thanks to matNMR (*70*). Under R2014a and R2015a, their respective Parallel Computing Toolbox was added to use gpuArray function. Computation times were measured with an internal timer. The source code is available in file Figure_II.4a_II.4b.m of (*71*).

B.2.b.iii. Python

No significant time difference was observed between Python 2.7 and 3.5 versions, neither between 32 and 64 bits versions. The source code is compatible with all of these options and is available in file Figure_II.5.py of (*71*) (CPU and GPU). Different SVD implementations were tested with NumPy 1.10.1 (*72*), SciPy 0.16.1 (*73*) and Scikit-Cuda 0.5.0 (*74*). The last one required PyCUDA 2015.1.3 (*75*) and CULA R18 free (*8*). Computation times were measured with an internal timer as well.

Under Linux, Automatically Tuned Linear Algebra Software (ATLAS) (*76*) and Open source Basic Linear Algebra Subprograms (OpenBlas) (*77*) development libraries, separately either one or the other, were installed as rpm packages. For each library, NumPy and SciPy were build with the pip mechanism[2]. Compiling these two latter packages with Intel Math Kernel Library (MKL) (*78*) was also tested. Under Windows, NumPy and SciPy superpack 32 bits with ATLAS library are available (*79*, *80*), and also pre-compiled packages with MKL library (*81*).

B.3. Results and discussion

As stated in the introduction, the main disadvantage of SVD is its long computation time. Of course the hardware itself is important but multiple steps are present between the human level function call and the hardware level implementation, namely the algorithm, the libraries and the use of hardware instructions (Figure VI.3a). Even on hardware, we can

[2] Forcing reinstallation of a specific python package can be done with ‘pip install --ignore-installed numpy==1.10.1’.

choose to compute either on CPU or on GPU. In the following, we checked the respective benefits of all these parameters.

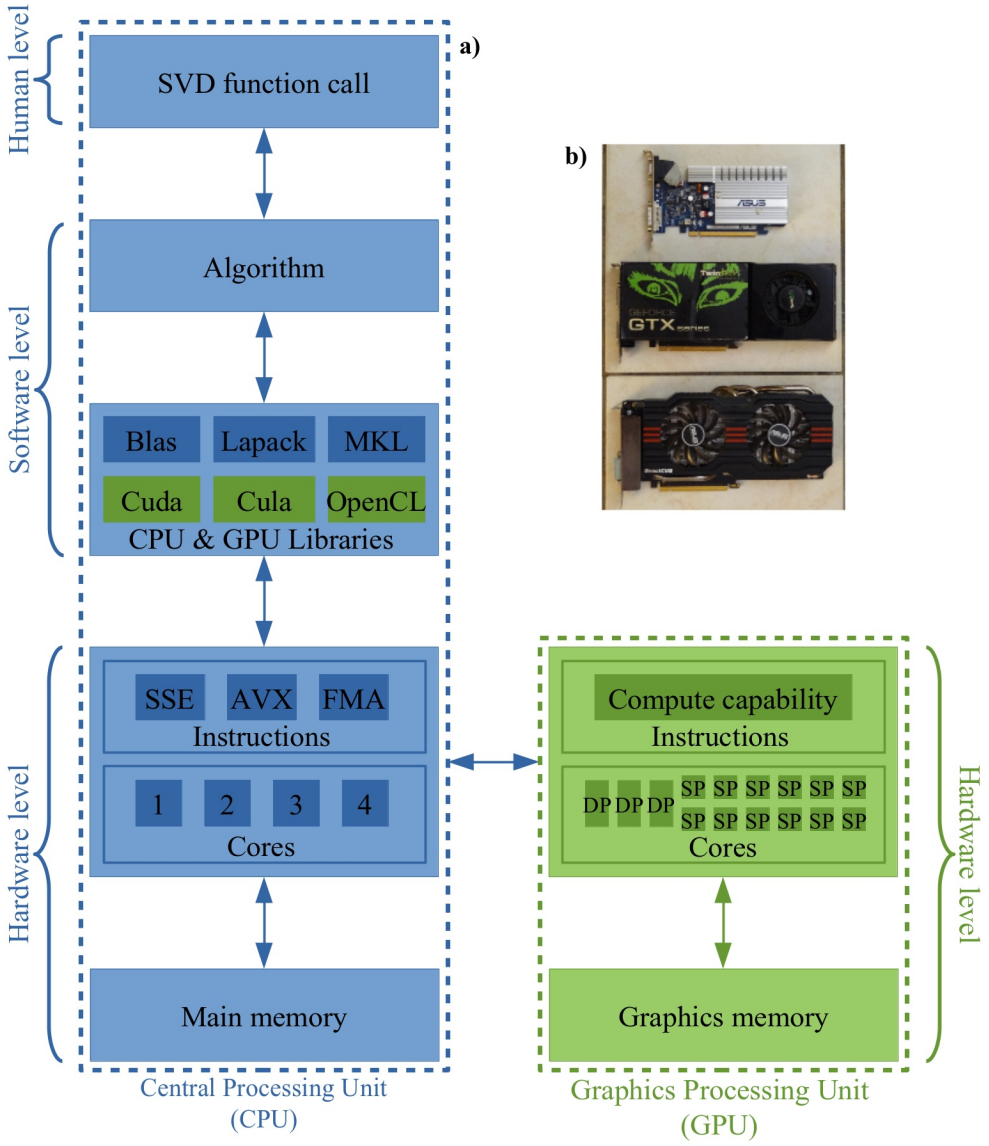


Figure VI.3: a) SVD function call diagram on CPU and GPU. MKL: Intel Math Kernel Library; SSE: Streaming SIMD Extensions; AVX: Advanced Vector Extensions; FMA: Fused Multiply Add; DP: double precision; SP: single precision. b) Some Nvidia GPU used. From top to bottom: 8400 GS, GTX 260, GTX 660.

B.3.a. Influence of hardware under Java CPU and GPU applications

B.3.a.i. Java CPU and GPU benchmarks

In an attempt to characterise the hardware needed to compute SVD, we used CPU ranging over 10 years and GPU ranging over 6 years, for both desktop and laptop computers (Tables VI.1 and VI.2 and Figure VI.3b). According to Figure VI.3a, we were changing

hardware level, supposing that everything was optimised at software level and measuring durations at human level. Study has been restricted to Intel CPU and Nvidia GPU because no AMD CPU was available in the laboratory and AMD GPU are not compatible with CUDA. Results are presented on Figure VI.4a.

We noticed first that computation were much faster for GPU than for CPU, with 2 to 23 s for GPU (dashed lines) and 67 to 500 s for CPU (top right end of plain lines), *i.e.*, 22 to 34 times speed increase, for a quite small matrix of 1025 × 1024 = 1.0e6 points. This almost square shape was due to construction of the Hankel matrix with a dataset containing an even number of points. Indeed, in order to not overwrite the corner point, it is necessary to add one row over columns. No significant time difference was seen with a true square matrix and thus this small shape difference will be neglected in the following.

Surprisingly, even a low-end GPU of 2008 (8400 GS) was surpassing a middle-range CPU of 2013 (Core i5 4670K) (with 23 and 67 s, respectively). This behaviour was really intriguing, as we would expect at least similar computation performance (*82*). However, when checking CPU activity, we observed that, with the CPU application, only one core was busy, what is called mono-threading. On the contrary, with the GPU application, not only GPU was fully busy, but also all the available CPU cores were used, what is called multi-threading. This difference between mono- and multi-threading is explored in Subsection B.3.b.

When increasing the square matrix size, a linear trend was visible in logarithmic scale, down-shifted for faster hardware (Figure VI.4a). However, computation time jumped when going from a rectangular matrix (diamond symbol) to a square one (square symbol), even if matrix size was not so different. This behaviour was observed for both CPU and GPU at 1537 × 512 = 7.9e5 *vs.* 1025 × 1024 = 1.0e6 points, and similarly for GPU at 3005 × 1024 = 3.1e6 *vs.* 2015 × 2014 = 4.1e6 points. This denoted that the number of mathematical operations dramatically increased for a square matrix. An explanation could be the use of reduced SVD for rectangular matrices, not computing the last rows and columns of U and V^T unitary matrices (see part (I) of this work (*13*)). This time jump is probably the reason why a rectangular matrix shape is chosen in most studies, despite a square matrix gives more precise singular values, *i.e.*, more performant denoising.

Table VI.1: Properties of **CPU** used for SVD under Java. Gray rows indicate hardware used for SVD under Matlab and Python. a: ark.intel.com; b: CPU-Z 1.72; c: Performance Test 8.0.

Central Processing Unit (CPU)	**Type**	**Year** [a]	**Fabrication (nm)** [b]	**Number of cores** [b]	**Cache (MB)** [a]	**Core frequency (MHz)** [b]	**Memory frequency (MHz)** [b]	**Memory size (MB)** [b]	**CPU Mark** [c]	**Mono-thread (Mops/s)** [c]	**Matrices (millions/s)** [c]
Intel Pentium M 745	laptop	2004	90	1	2	1800	133	1024	444.8	577	1.41
Intel Pentium 4 530	desktop	2004	90	2	1	3000	200	1024	335.2	726	0.29
Intel Core 2 Duo E6400	desktop	2006	65	2	2	2130	333	2048	1451	849	3.72
Intel Core 2 Quad Q8200	desktop	2008	45	4	4	2330	400	4096	2001	1004	5.8
Intel Core 2 Duo T9600	laptop	2008	45	2	6	2800	400	4096	2190	1161	4.67
Intel Core i3 4005U	laptop	2013	22	4	3	1700	800	4096	2551	1010	11.4
Intel Core i5 4670K	desktop	2013	22	4	6	4200	1000	8192	8824	2519	31.6

Table VI.2: Properties of **GPU** used for SVD under Java. Gray rows indicate hardware used for SVD under Matlab and Python. a: GPU-Z 0.8; b: Cuda-Z 0.9; #N/A: not available.

Graphics Processing Unit (GPU)	**Type**	**Year** [a]	**Fabrication (nm)** [a]	**Number of cores** [a]	**Bandwidth (GB/s)** [a]	**Core frequency (MHz)** [a]	**Memory frequency (MHz)** [a]	**Memory size (MB)** [a]	**Single precision float (GFLOPS)** [b]	**Double precision float (GFLOPS)** [b]	**CUDA compute capability** [b]
Nvidia Quadro FX 570	desktop	2007	80	16	12.8	460	400	256	29	#N/A	1.1
Nvidia GeForce 8400 GS	desktop	2008	65	8	6.4	567	400	512	21	#N/A	1.1
Nvidia Quadro NVS 160M	laptop	2008	65	8	11.2	580	700	256	23	#N/A	1.1
Nvidia Quadro FX 770M	laptop	2008	65	32	25.6	500	800	512	79	#N/A	1.1
Nvidia GeForce GTX 260	desktop	2008	65	216	111.9	576	1000	896	533	67	1.3
Nvidia GeForce 820M	laptop	2012	28	96	14.4	625	900	2048	315	31	2.1
Nvidia GeForce GTX 660	desktop	2012	28	960	144.2	1100	1500	2048	1707	88	3.0

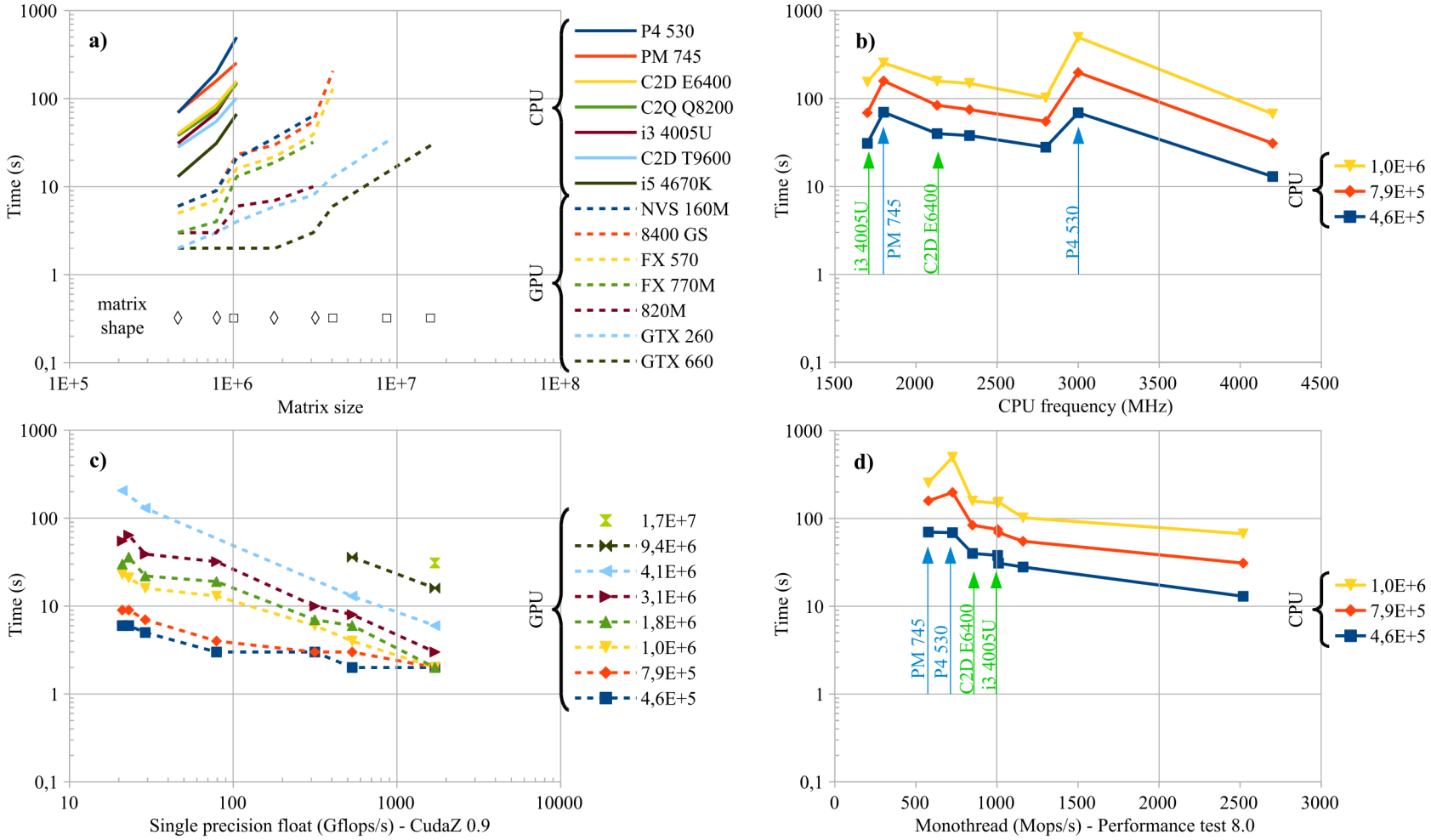


Figure VI.4: Computation times for SVD under **Java**, in logarithmic scale.
a) Comparison of CPU and GPU times against matrix size; ◊: rectangular matrix; □: square matrix. b) CPU times against CPU frequency. c) GPU times against single precision float performance. d) CPU times against mono-thread performance; light blue and light green arrows evidence CPU cache and CPU release year influence, respectively. CPU and GPU times are drawn with plain lines and dashed lines, respectively.

Thanks to the speed up obtained on GPU, it was possible to compute a matrix of 4097 × 4096 = 1.7e7 points in 31 s on a mid-range GPU of 2012 (GTX 660). By extrapolating the curve in Figure VI.4a, it may take around 6000 s on a mid-range CPU of 2013 (Core i5 4670K). For square matrices, the slope seems similar between graphic cards. As underlined in Subsection B.2.b.i, only 64 bits Java versions can handle matrix sizes above 3005 × 1024 = 3.1e6 points. However, all the processors and some graphic cards (NVS 160M, FX 770M and 820M) were benchmarked using the 32 bits Java applications, which explains the truncated curves for this hardware.

B.3.a.ii. Java CPU performance indicator

To better characterise the hardware needed for Java CPU SVD, we looked for a performance indicator. One would expect CPU frequency to be a good one but it was clearly not the case as shown on Figure VI.4b. However, mono-thread performance evidenced a trend and was thus a usable parameter as shown on Figure VI.4d. This is coherent with our observation of only one active processor core under the Java CPU application.

Characteristic points are visible on these curves (Figures VI.4b and VI.4d). Light blue arrows highlight Pentium M 745 and Pentium 4 530, which were two processors of 2004. Their frequencies were very different (1800 and 3000 Mhz, respectively) but their mono-thread performance were similar. Additionally, with a bigger matrix size (yellow line), the Pentium M 745 was faster than the Pentium 4 530, even if the latter had a higher frequency. The main difference between them was their cache size, of 2 and 1 MB, respectively. The CPU cache is the amount of quick memory directly available inside the processor. On the contrary, memory plugged into motherboard is at least 10 times slower. As SVD request many matrix-vector multiplication, memory access is limiting.

Light green arrows evidence Core 2 Duo E6400 and Core i3 4005U, which were two processors with a similar frequency but released in 2006 and 2013, respectively. No performance increase was observed using Java SVD CPU application between these two processors. That was also questioning as we would expect that some hardware optimisations happened in 7 years. These observations highlight that the best CPU for SVD under Java will not necessarily be recent or have a high frequency, but rather have a high memory cache and a high mono-thread performance. In other words, it is better to use an old high- or middle-range

CPU than a new low-range one. This explains why Core 2 Duo T9600 was faster than Core i3 4005U. The former processor is a good candidate to denoise all over the night a matrix of 8193 × 8192 = 6.7e7 points with the 64 bits CPU Java application.

B.3.a.iii. Java GPU performance indicator

Graphic card computing power is characterized by core frequency and number of cores. Despite frequency was not so different along the series, number of cores was strongly increasing over the years and with them the Single Precision Floating Point performance (SP or FP32), expressed in Giga FLoating-point Operations Per Second (GFLOPS) (*83*). Only high-end professional cards have a high Double Precision Floating Point performance (DP or FP64). General public GPU are more commonly devoted to games and lack DP. The precision is the number of bits used to store numbers, 32 and 64 for SP and DP, respectively (*84*). The higher the precision used, the lower the computed error. However, the errors initially present in the matrix can be larger than rounding errors (*85*). Moreover, CULA free (*8*), the library implemented on Java GPU application could only use SP. It was thus useless to invest money in professional cards and we favoured general public GPU. For instance, a Nvidia Tesla P100 GPU costs around 8 k€. On Figure VI.4c, a time decreasing linear trend was obtained in logarithmic scale when increasing SP, which denoted a good indicator.

Another important parameter for SVD with Java GPU application, was the amount of memory available, both on GPU (device) and on motherboard (host). Plassman stated that SVD needed up to $8\,n^2 + 12\,n$ work storage (*86*), for a matrix with n columns. This value had to be multiplied by 4 bytes for both floating and integer numbers to be stored in memory. Additionally, there was a 1.5-3 times transient overhead during low-rank approximation. Following this rule, the largest tractable matrix was 6657 × 6656 = 4.4e7 points on a GPU with 2 GB of memory.

In this subsection we have seen influence of hardware on SVD computation time under Java. As stated above, the time difference between CPU and GPU Java application was intriguing, especially when comparing the low-end graphic card 8400 GS to the middle-range processor Core i5 4670K. Additionally, the CPU and GPU application were mono-threaded and multi-threaded, respectively. This was typically an algorithm problem and we explored it using Matlab.

B.3.b. Influence of algorithm under Matlab

Algorithms are the mathematical operations involved and their informatics implementation to obtain the relevant function. Plassman compared the available SVD algorithms and their impact on ill-conditioned matrices (*86*). The simplest computation method is to use eigendecomposition, but its lack of precision was demonstrated by Läuchli (*87*). The classic complete SVD uses a three steps process:

1. reduction to bidiagonal form,
2. computation of SVD on bidiagonal matrix,
3. obtention of singular vectors.

Step 1 involves Householder reflections and step 2 can either use QR iteration in Golub-Kahan-Reinsch (GKR) algorithm (*88*, *89*), divide and conquer method (*43*, *90*) or Multiple Relatively Robust Representations (MRRR) (*91*). An alternative SVD algorithm, combining steps 1 and 2, is to use Jacobi rotations and convergence criteria (*92*).

To explore algorithm influence we focussed on two computers, one from 2008 with a Core 2 Quad Q8200 and a GTX 260 under Linux, and the other one from 2013 with a Core i5 4670K and a GTX 660 under Windows. Both were in the same price segment and reflected middle-range equipment available at those dates. The used Matlab versions were detailed in Subsection B.2.b.ii. According to Figure VI.3a, we fixed hardware level and observed software level influence.

B.3.b.i. Matlab R2010a

Under Matlab R2010a, DP computation times on CPU were already smaller than those with Java CPU application for a matrix of 1025 × 1024 = 1.0e6 points (plain dark blue line on Figures VI.5a and VI.5b): 78.2 s instead of 149 s for Core 2 Quad Q8200, and 11.6 s instead of 67 s for Core i5 4670K. It should be noted that computation times can be divided by two when using single precision (plain red line), and by two again with non-complex data (not shown). Java CPU application used complex numbers and accordingly to the above results, it presumably used DP. While Java used only mono-threading, Matlab computation started with a multi-threaded step and kept on with a mono-threaded one. This already denoted a different algorithm between the two programming languages.

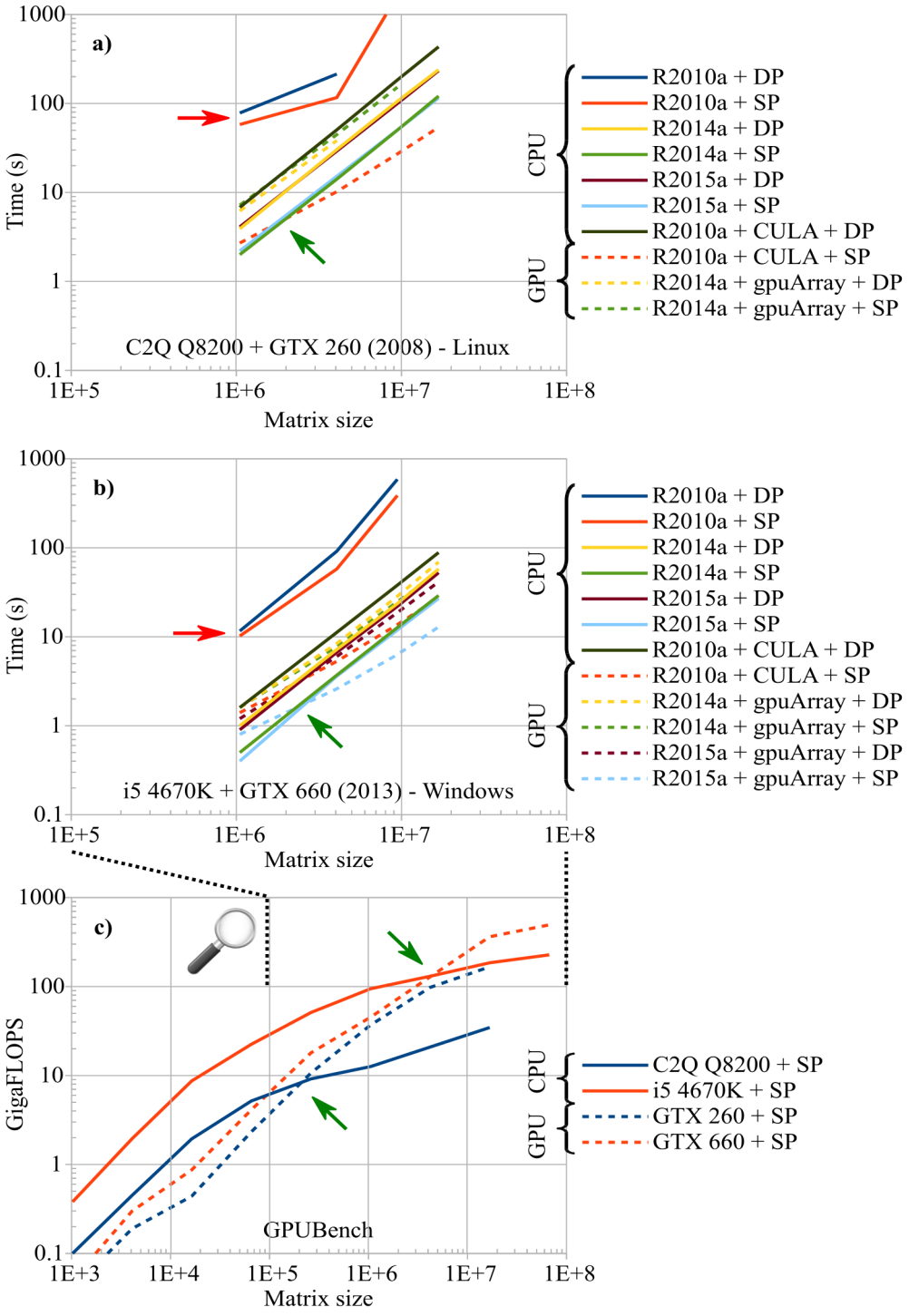


Figure VI.5: a) and b) Computation times for SVD under **Matlab**, in logarithmic scale. Comparison of CPU and GPU times against matrix size for different Matlab versions. a) C2Q Q8200 + GTX 260 (2008). b) i5 4670K + GTX 660 (2013). c) SP performance against matrix size measured with GPUBench v1p7 under R2014a; the horizontal scale is larger than on a) and b). CPU and GPU times are drawn with plain and dashed lines, respectively. Red and green arrows indicate bad algorithm and CPU-GPU crossing, respectively.

Under this Matlab version it was also possible to use GPU with CULA, which was the library implemented under Java GPU application. Results were similar between Java GPU and Matlab R2010a + CULA + SP GPU applications (dashed red line). However, as we used the free version of CULA, DP computation was not allowed on GPU and R2010a + CULA + DP (plain black line) felt back on CPU with LAPACK library in multi-threading mode. As a consequence, strong improvement was observed against R2010a + DP (plain dark blue line). At this point, an order of magnitude on computation times has already been gained for CPU DP under Matlab.

B.3.b.ii. Matlab R2014a

Further improvement on CPU was obtained with R2014a + DP (plain yellow line on Figures VI.5a and VI.5b) being almost two times faster than R2010a + CULA + DP (plain black line), and 7-33 times faster than R2010a + DP (plain dark blue line), depending on matrix size. This was explained by the divide and conquer approach preferred for SVD starting from R2010b. Gu *et al* claimed that this algorithm was 9 times faster on bidiagonal matrices (*93*), in agreement with our observations. The extra gain is due to the four cores simultaneously used on the CPU. Again SP (plain green line) was two times faster than DP. Small matrices, up to $1025 \times 1024 = 1.0e6$ points for Core 2 Quad Q8200, and up to $3073 \times 3072 = 9.4e6$ points for Core i5 4670K, were even computed faster on CPU with R2014a + SP than on GPU with R2010a + CULA + SP, as indicated by the green arrow. Unfortunately, CULA free was not compatible with R2014a, but it was nevertheless possible to use GPU thanks to the gpuArray Matlab function. Surprisingly, worst results were obtained, with a GPU time longer than its corresponding CPU time. Moreover R2014a + gpuArray + DP times (dashed yellow line) were shorter or equal to R2014a + gpuArray + SP times (dashed green line), what is in contradiction with DP/SP ratio on GPU (1/8 and 1/24 for GTX 260 and GTX 660, respectively). This revealed that part of the computation was done in DP, despite SP was called. When checking CPU and GPU activity during SVD, it was observed that GPU was only used at the beginning and at the end of the processing. This denoted that SVD using gpuArray under Matlab R2014a was not an optimised algorithm and that this version should be avoided.

B.3.b.iii. Matlab R2015a

In order to check if a new version of Matlab could further improve computation times, we used Matlab R2015a. A slight decrease was observed on CPU from 57.8 to 52.5 s and from 29.1 to 27.0 s for R2014a + DP (plain yellow line on Figures VI.5a and VI.5b), R2015a + DP (plain brown line), R2014a + SP (plain green line) and R2015a + SP (plain light blue line), respectively, for a matrix of 4097 × 4096 = 1.7e7 points. Matlab R2015a was not compatible with GTX 260 GPU, due to its compute capability of 1.3. A much stronger improvement was obtained for the above matrix size with GTX 660 GPU: from 69.1 to 41.4 s and from 52.7 to 12.9 s for R2014a + gpuArray + DP (dashed yellow line), R2015a + gpuArray + DP (dashed brown line), R2014a + gpuArray + SP (dashed green line) and R2015a + gpuArray + SP (dashed light blue line), respectively. The latter configuration outperformed R2010a + CULA + SP (27.5 s, dashed red line) owing to a more pronounced GPU utilisation during SVD. Despite SP was faster than DP, the DP/SP = 1/24 ratio was not respected. However, SVD algorithm was strongly optimised in R2015a + gpuArray against R2014a + gpuArray. While Core i5 4670K CPU remained more efficient for matrices up to 1025 × 1024 = 1.0e6 points, GTX 660 GPU outperformed it in SP mode for larger matrices. The obtained computation times under Matlab R2015a are thus very good, both on CPU and GPU and were better than under Java.

B.3.b.iv. Matlab GPUBench

The cross in computation time between CPU and GPU was further investigated with GPUBench v1p7 (*94*). This code compared CPU and GPU performance against matrix size for matrix-vector left division, which is a linear equations system solver. Such computation gives much less peak SP and DP float performance than reported in Table VI.2, and is rather compute-bond than memory-bound. This benchmark involves lots of matrix-vector operations as SVD does. For both 2008 and 2013 computers, a crossing was visible between CPU and GPU in SP mode (Figure VI.5c). This was explained by the time needed for data goings and comings between processor and graphic card and between graphic card core and its memory (Figure VI.3a). This is a hardware limitation. Interestingly, the cross appeared in the same matrix size range (1e5 to 1e7) than the one observed for SVD (Figures VI.5a and VI.5b). However, its position strongly depends on the algorithm used and on the relative float performance of CPU and GPU. Low-end GPU are thus not recommended as better results are

obtained with CPU. In DP mode, despite the half computing power of a Core i5 4670K against SP, even a GTX 660 never overpassed it (not shown).

Similarly to Java, matrix size under Matlab is limited by host and device memory. Nevertheless, memory consumption under Matlab is improved over Java as no overhead is present during low-rank approximation, pushing away maximum matrix size. Additionally, host memory amount is considerably reduced, down to be almost identical to device memory one. A GPU with 2 GB of memory is thus limited to matrices of 7169 × 7168 = 5.1e7 points.

In this subsection, we highlighted that the divide and conquer algorithm decrease SVD computation time by a factor of nine. SP gives an additional factor of two in computation time on CPU, being faster than GPU for matrices smaller than 1025 × 1024 = 1.0e6 points (Matlab R2014a *vs.* R2010a). Despite the strong improvement for SVD on CPU, middle-range GPU remains relevant in SP mode for matrices above this size, up to the GPU memory limit (see previous paragraph). For legacy hardware dating from 2008, the best compromise is to use Matlab R2010a and CULA free R14 with SP. For hardware dating from 2013, the best choice is to use the most recent Matlab version with SP and gpuArray function. CPU computation should especially be avoided on Matlab R2010a as evidenced by the red arrows on Figures VI.5a and VI.5b. Matlab R2014a is not recommended neither for GPU. Next step was to focus on the libraries used and their call to hardware instructions, which we explored under Python.

B.3.c. Influence of libraries and hardware instructions under Python

According to Figure VI.3a, software level is divided in algorithms and libraries. After changing the algorithms, *i.e.*, the involved mathematical functions, we were interested in the underlying libraries, *i.e.*, the link between software level and hardware level. A library is a collection of functions that consists of pre-written optimised code. A single library can be called by multiple software or by other libraries. Usually, SVD first calls LAPACK (Linear Algebra PACKage) (*95*) which itself calls BLAS (Basic Linear Algebra Subroutines) (*96*). While LAPACK is a high-level library, BLAS is a low-level one, optimised by CPU hardware specialists (*97*). On GPU, CULA (*8*) is a unified BLAS/LAPACK package based on nvidia CUDA technology (*51*).

B.3.c.i. ATLAS, OpenBLAS and MKL libraries

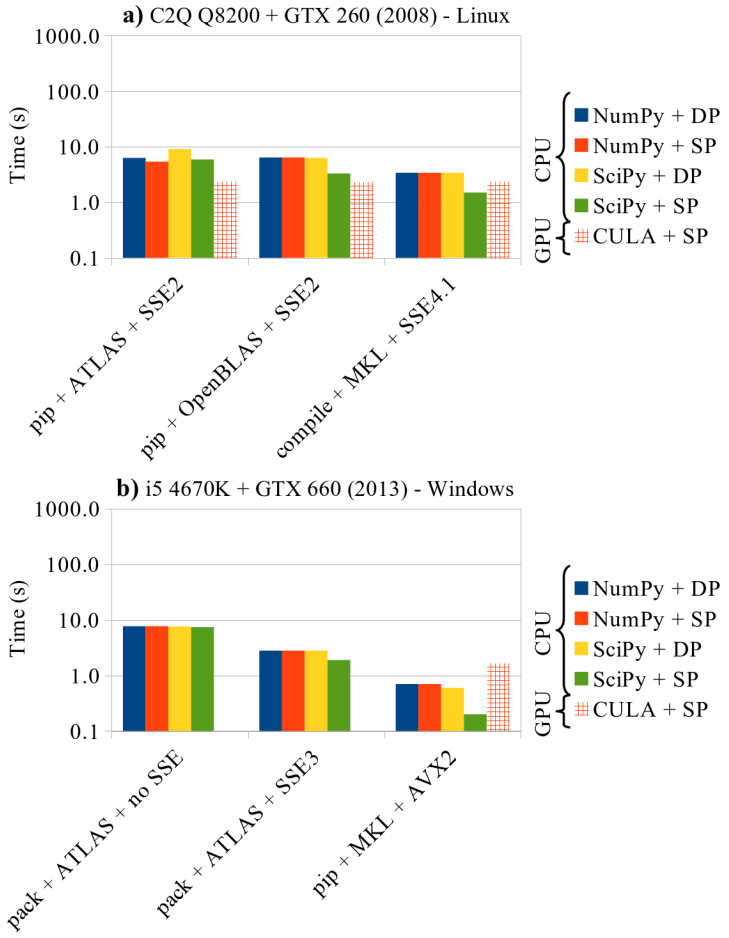


Figure VI.6: Computation times for SVD under **Python** with influence of libraries and hardware optimisations for a matrix size of 1025 × 1024 = 1.0e6 points, in logarithmic scale. a) C2Q Q8200 + GTX 260 (2008) under Linux. b) i5 4670K + GTX 660 (2013) under Windows.

Two libraries are available for SVD on CPU under Python: NumPy and SciPy. Those packages provide algorithms which are linked to low-level libraries. Under Linux Fedora 22, ATLAS (*76*) was the default[3]. It was possible to replace it either with OpenBLAS (*77*) or with MKL (*98*). Results for a matrix size of 1025 × 1024 = 1.0e6 points are presented on Figure VI.6a for our reference computer with a Core 2 Quad Q8200 and a GTX 260 (2008). First, we noticed that decreased computation times were obtained when moving from ATLAS (left column) to OpenBLAS (middle column) and MKL (right column). While OpenBLAS improved only Scipy results, MKL was almost twice faster than OpenBLAS for both Numpy and Scipy. Secondly, with ATLAS (left column), SciPy computation times (yellow and green

[3] NumPy library can be verified using 'numpy.show_config()'.

lines) were longer than NumPy ones (blue and red lines), both for DP and SP. This behaviour was surprising as SciPy was intended to do some scientific calculation. It may be improved in a newer ATLAS version. Third, for all libraries tested, no performance increase was visible with NumPy when changing from DP (blue line) to SP (red line), which may indicate a bug of NumPy. On the contrary, SciPy + SP computation times (green line) were almost half DP ones (yellow line), as expected from DP/SP computing power ratio. Finally, for this small matrix of $1025 \times 1024 = 1.0e6$ points, and no matter if ATLAS, OpenBLAS or MKL library was installed, CULA + SP (hatched red line) was slower on GPU than MKL + SciPy + SP on CPU.

B.3.c.ii. SSE and AVX hardware instructions

Despite MKL seemed very promising, it was not the only factor changing in the above experiment as the implemented hardware optimisations changed from SSE2[4] to SSE4.1[5]. SSE stands for Streaming SIMD Extensions and its number reflects the version used. SIMD are embedded capabilities on CPU. Since 2008, a new family of instructions is available, named Advanced Vector Extensions (AVX). Even if the processor support them, the library does not necessarily call them. A history of SIMD development is available in reference (*59*). Under windows, NumPy and SciPy superpack provided options to selectively use no SSE or SSE3[6]. Results are presented on Figure VI.6b for our reference computer with a Core i5 4670K and a GTX 660 (2013). When moving from no SSE (left column) to SSE3 (middle column), computation times were divided by three with an additional gain for SP. Moreover, when using both MKL and AVX2, a huge performance was obtained, outperforming GPU computation with CULA + SP (hatched red line).

As underlined here, the time needed to perform SVD was impressively reduced by a factor of 38 on the same CPU under Python, by optimising the used libraries and their hardware calls. Indeed, decomposition of a matrix of size $1025 \times 1024 = 1.0e6$ points was done with SciPy in 7.6 s without optimisations and in 0.2 s using MKL library and AVX2 instructions.

---

[4] SSE2 instruction can be checked with ‘objdump -d /lib64/python/site-packages/numpy/core/*.so | grep -i ADDPD’.

[5] NumPy and SciPy are compiled with ‘-xHost’ option enabling the highest SIMD instruction set available, which is SSE4.1 on a Core 2 Quad Q8200.

[6] No SSE option is ‘numpy-1.10.1-win32-superpack-python2.7.exe /arch nosse’.

B.3.d. Comparison of Java, Matlab and Python

B.3.d.i. Computation times

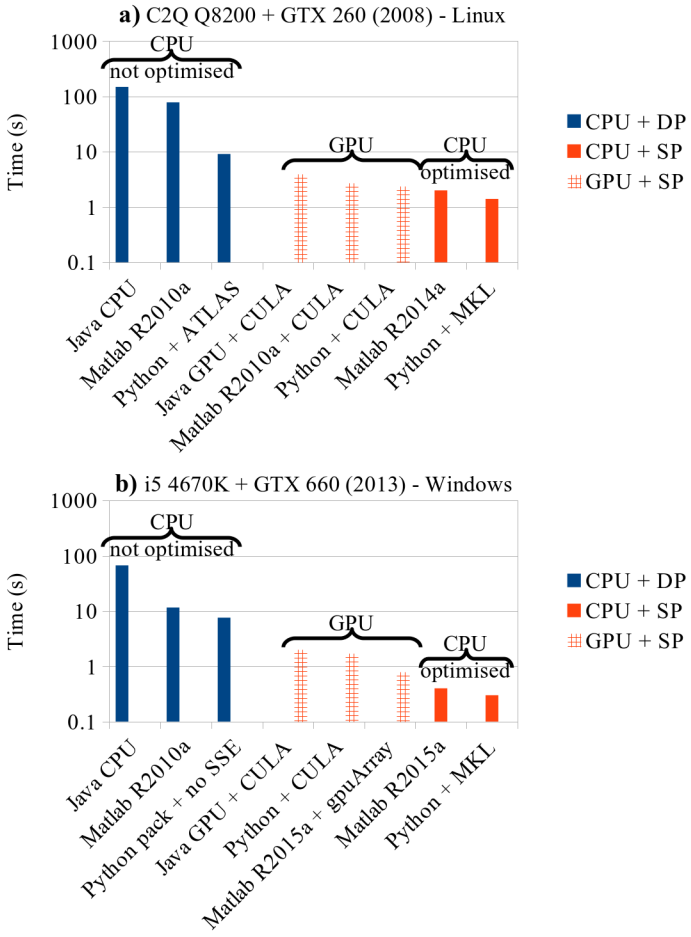


Figure VI.7: Comparison of **Java, Matlab and Python** SVD computation times for a matrix size of 1025 × 1024 = 1.0e6 points, in logarithmic scale. a) C2Q Q8200 + GTX 260 (2008) under Linux. b) i5 4670K + GTX 660 (2013) under Windows.

In previous subsections, Java, Matlab and Python software were used for their specific testing capabilities. But how do they compare to each other? In order to answer this question, Figure VI.7 shows SVD computation times for a matrix of 1025 × 1024 = 1.0e6 points, which is the maximum size for Java 32 bits CPU application. Similar conclusions were raised for our two reference computers (2008 and 2013). The measured computation times were grouped into three categories: unoptimised CPU (in plain blue), GPU (in hatched red) and optimised CPU (in plain red), from the slowest to the fastest. The first group consisted of Java CPU, Matlab R2010a and Python with default configuration. The second group contained Java GPU, Matlab R2010a with CULA or Matlab R2015a with gpuArray, depending on GPU

generation, and of Python with CULA. The third group referred to a recent version of Matlab and to compiled Python, both with MKL library and all available SIMD instructions activated[7]. For this small matrix of 1025 × 1024 = 1.0e6 points, the CPU outperformed the GPU, due to data transfer delays limiting GPU efficiency. However, for larger matrices, computation was faster on GPU.

Comparing Java CPU and Python with MKL, there was a gain of 100 on the same CPU. This was explained as follows:

- a factor of 9 using the divide and conquer algorithm
- a factor of 3 using hardware instructions such as SSE3 or AVX2
- a factor of 2 using MKL library
- a factor of 2 using single precision instead of double precision

B.3.d.ii. Maximum matrix size

This major time improvement raise the question of the absolute maximum matrix size that could be computed using SVD and low-rank approximation. As underlined in Subsection B.3.a.iii, the limiting parameter is memory, both on GPU device side and on CPU host side. The crucial point is to use 64 bits applications and a GPU with as much memory as possible. Nevertheless, this will depend on the way memory is allocated and released during SVD process. Our late investigations, on a GTX 1070 with 8 GB of memory and 7040 SP GFLOPS, gave the following maxima on our ^{87}Sr FID. The computation times were the sum of SVD and low-rank steps. To maximize the latter, the operation was performed with all singular values, that is to say without any denoising.

- Java GPU: 9217 × 9216 = 8.5e7 complex points in 150 + 32 = 182 s.
- Matlab R2018a + gpuArray + SP: 12289 × 12288 = 1.5e8 complex points in 188 + 4 = 192 s.
- Python + CULA + SP: 15219 × 15218 = 2.3e8 complex points in 594 + 6 = 600 s.

Under Python, it was thus possible to apply SVD on the full ^{87}Sr FID, without any truncation. For comparison, a Nvidia P100 GPU, with 4670 DP GFLOPS and 16 GB of memory, completed the full SVD of a 20000 × 20000 = 4.0e8 real matrix in 90 s, with a highly optimised CPU-GPU algorithm (*54*). This result was really impressive as the authors

[7] 'version('-blas')' under Matlab gives MKL 11.0.5 for R2014a and MKL 11.1.1 for R2015a.

obtained a faster computation on a much larger matrix with less computing power and double precision. There is thus plenty of place to improve SVD denoising.

Directly comparing Java, Matlab and Python was a difficult task as they were not optimised in the same way and it was hard to check what was hidden under the hood. However, Java GPU was less performant, both in speed and in matrix size. Better results may be obtained with optimised libraries. While Matlab was faster, the memory usage was limiting. Python computation was not as fast but could handle the biggest matrix. This time advantage for Matlab was explained by a better CPU usage during SVD on GPU. However memory was more finely managed under Python. Our results suggested that the key parameter was not the software and the programming language, but rather the used libraries and the calls to hardware instructions.

An additional advantage of Python is that it is free of charge and rather easy to program. In order to compute SVD in a minimum amount of time, we recommand to install a Python distribution with MKL library included, such as Anaconda (*99*), and to add the following libraries: SciPy (*80*), scikit-cuda (*74*), PyCUDA (*75*), and nmrglue (*100*) for NMR data. In addition, CULA (*8*) and CUDA toolkit (*51*) packages are necessary. We provide in file svd_auto.py of (*71*) an optimised SVD function using either the CPU or the GPU, depending on the matrix size. Our tests suggested a minimum value of 4096 columns or rows to switch from the CPU to the GPU. This default value will depend on the hardware used and can be checked by running directly the program. The code is designed to be as simple as possible, with only one necessary parameter, namely the matrix two-dimensional array. Automatic thresholding is applied using Malinowski's significant level indicator (*40*). This SVD function is also suitable to be used in PCA and related data mining techniques. In addition, we provide a second program (file denoise_nmr.py of (*71*)), in charge of importing and exporting Bruker NMR data and to prepare the matrix transferred to SVD program. Again, the only requested parameter is the data directory.

B.4. Conclusion

This article separated in two parts focussed on SVD, which is used both for spectra denoising and as part of PCA data mining. In the first part, we gave theoretical background and found the minimum experimental signal-to-noise ratio needed to have a correct denoised spectrum. We highlighted the overestimation of denoised Gaussian peaks. In this second part,

we focussed on the computation time needed for SVD treatment. While our first attempts under Java CPU were extremely slow even with a recent processor, their counterparts with graphic cards were extraordinary fast. This unexpected difference led us to check if different Matlab versions could improve this situation. The divide and conquer algorithm was very helpful. Additional tests were undergone under Python to check the influence of software libraries and of SIMD hardware instructions call. Combining these optimisations, computation times on processor were even better than on graphic cards, being 100 times faster than our first tests under Java CPU, for a matrix of 1025 × 1024 = 1.0e6 points. Despite this approach is generalisable to any intensive computation, specific time gain will depend on the involved mathematical operations. The take home message is thus to update software and to use optimised libraries and especially Intel MKL if available. This choice should be preferred against hardware updates.

However, for matrix above 4097 × 4096 = 1.7e7 points and middle range hardware, GPU gave better results, up to GPU memory limit. We thus provided Python programs to apply SVD either on CPU or on GPU, and to denoise NMR FID. Further improvement could be obtained with mixed CPU/GPU optimised code, *i.e.*, hybrid computing (*101*). However, such an approach is not suitable for non-computer-scientists people. Using clMAGMA library (*102*), combining divide and conquer on both CPU and GPU could be a good alternative (*103*). In this case, it would be possible not only to use Nvidia GPU with CUDA but also AMD GPU with OpenCL.

This study has given strong background and optimisations for experiments involving SVD, either for denoising or for PCA. It may thus help scientists who want to use efficiently this technique, which is expected to be widely used in the forthcoming years.

B.5. Acknowledgements

The French Région Ile de France – SESAME program is acknowledged for financial support (700 MHz spectrometer). Sylvie Masse and Cedric Lorthioir are thanked for fruitful discussions.

B.6. Supplementary material

Programs source codes are available online in (*71*).

C. Additional computation tests

C.1. Influence of algorithm precision

In our SVD denoising program, we choose to use Single Precision (SP) to reduce SVD computation time (Subsection B.3.d.i). Unfortunately, this has an impact on cumulative errors. By decomposing and reconstructing a matrix, without low-rank approximation, and without any computation error, all the elements from the reference matrix (y_i^{ref}) and the processed matrix (y_i^{proc}) should be identical, in the limit of $2^{-23} \approx 1\times10^{-7}$ for SP and $2^{-52} \approx 2\times10^{-16}$ for Double Precision (DP) (*104*). We checked the relative error (*rel*) according to Equation VI.1 on matrices with increasing integers converted to real or complex floats at the desired precision.

$$rel = \frac{max\left|y_i^{proc} - y_i^{ref}\right|}{max\left|y_i^{ref}\right|} \qquad \text{VI.1}$$

This measurement highlights the error introduced by computation, and compares it to the highest value of the matrix. A relative error $rel = 4\times10^{-5}$ was measured on CPU for a matrix of $1025 \times 1024 = 1.0\times10^6$ SP real points, whereas a value 5 times higher, rel = 2e-4, was observed on GPU for the same matrix. This highlighted the differences between MKL and CULA algorithms, for CPU and GPU, respectively. The situation got worse for larger matrices of $8193 \times 8192 = 6.7\times10^7$ SP complex points with $rel = 1\times10^{-2}$ on both CPU and GPU. For comparison, $rel = 4\times10^{-12}$ was obtained with DP on CPU for the same matrix. Despite an error of 1 % should not be a problem for many applications, high dynamic range spectra, *i.e.*, with both intense and tiny peaks, may require DP and thus larger memory amount and longer computation times for relevant results. It should be noted that the above measurement is a worst case matrix with a very high dynamic range and no data structure, except a linear increase. Further tests are needed on real or simulated spectra to better characterise the impact of computation errors on quantification.

C.2. Heterogeneous computing

In order to optimise SVD computation, it would be desirable to activate CPU and GPU simultaneously, so-called heterogeneous computing (*101*). CULA library requires CUDA compatible devices (*51*), *i.e.*, only GPU from Nvidia manufacturer, but neither GPU from AMD manufacturer nor CPU. OpenCL (*52*) is a relevant alternative, as it can be executed on

all brands of CPU and GPU (Figure VI.8). However, programming directly in OpenCL or in C++ is outside of our competence domain. Moreover, each library has its own syntax for SVD, which has to be checked into documentation, internet forums, or even program sources. For instance, SVD can either use Golub-Kahan-Reinsch (GKR) algorithm (*88*, *89*) or divide and conquer method (*43*, *90*), denoted as *xgesvd* or *xgesdd* instructions, respectively, where *x* can be replaced by *s*, *c*, *d*, or *z*, stating for real with single precision, complex with single precision, real with double precision, or complex with double precision, respectively (*69*). GKR is based on QR decomposition, where Q is an orthogonal matrix and R is an upper triangular matrix (*5*). Furthermore, OpenCL driver installation problems can occur. Under Linux CentOS, it was possible to activate the three available OpenCL devices of a notebook (CPU, Intel integrated GPU, and AMD discrete GPU) only by retrograding from version 7.4 to 7.2 and by blocking updates of Xorg and Gnome packages.

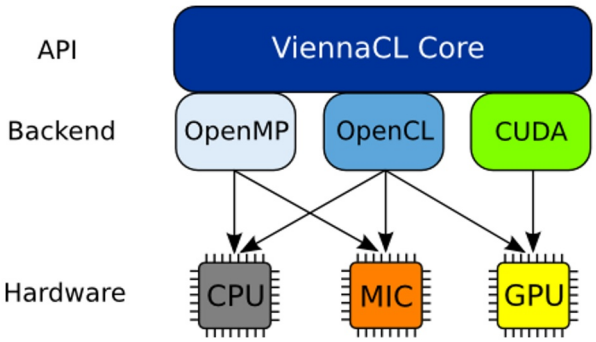


Figure VI.8: Software to hardware access, using different backends (*105*).
API: Application Programming Interface; CPU: Central Processing Unit; MIC: Many Integrated Cores; GPU: Graphics Processing Unit.

Table VI.3: Comparison of SVD libraries under C++ and Python.

Program	**Version**	**C++ SVD**	**Python SVD**	**CPU**	**CUDA**	**OpenCL**	**Simul-taneous**	**Python speed**
Armadillo (*106*)	9.100.5	yes	no	yes	no	no	no	not available
Arrayfire (*107*)	3.5.1	yes	yes	yes	yes	yes	no	slow
CLmagma (*102*)	1.3.0	yes	no	yes	no	yes	yes	not available
CULA (*8*)	R18	yes	yes	no	yes	no	no	fast
Magma (*108*)	2.4.0	yes	yes	yes	yes	no	yes	not working
MKL (*7*)	2018.0	yes	yes	yes	no	no	no	fast
ViennaCL (*109*)	1.7.1	yes	yes	yes	yes	yes	yes	not tested

Our choice was to focus on high-level OpenCL libraries under Python, to improve hardware modularity and operating system independence. Some available libraries are

presented in Table VI.3. Arrayfire was chosen for its ability to access CPU, OpenCL and CUDA backends (*107*). Despite promising, computation was too slow to be usable and it was not simultaneous on the different devices. ViennaCL (*109*) may be an interesting alternative, but it has not yet been tested.

In Subchapter C, the effect of single precision was checked, and hybrid computing was tested. For now, MKL and CULA are still the faster choices for SVD denoising with our program.

D. Chapter conclusion

Computation time of SVD was analysed in Subchapter B. A gain of 9 was obtained with the divide and conquer algorithm, of 3 with SSE3 or AVX2 hardware instruction call, of 2 using Intel MKL library, and of 2 using single precision. An overall SVD computation time decrease of 100 was thus available on the same CPU, including 10 years old hardware. This demonstrated that updating hardware is not the ultimate solution, but that first optimising the software layer is preferable. Our tests suggested that for matrices smaller than $4097 \times 4096 = 1.7 \times 10^7$, CPU can overpass GPU, due to GPU data transfer latencies. For larger matrices, computation was faster on GPU. The ultimate matrix size limit is given by GPU memory size. Therefore, SVD denoising of a $15000 \times 15000 = 2.3 \times 10^8$ points matrix was achieved in 600 s under Python with a Nvidia 8 GB GTX 1070 GPU. In Subchapter C, algorithm precision was checked by applying SVD and matrix reconstruction without reducing the number of singular values. For a matrix of $8193 \times 8192 = 6.7 \times 10^7$ complex points, a relative error of 1 % was found on a worst case matrix with single precision. This might be problematic for large dynamic scale matrices. In such a case, double precision has to be preferred, thus increasing computation time, whereas it is not necessary for standard matrices. Some work to use heterogeneous computing and OpenCL has been started. Unfortunately, the tested libraries did not improved computation time and so far MKL and CULA are still the best compromise. Further decrease of computation time can be achieved using sparse matrices (*45*). Alternatively, computational complexity can be lowered through uncoiled random QR denoising (*4*), which is faster and robust to the number of selected components in denoised spectra.

E. Chapter bibliography

1. J. Songsiri, presented at the EE 507 - Computational Techniques for Engineers (Faculty of Engineering, Chulalongkorn University, Bangkok, TH, 2016), available at http://jitkomut.eng.chula.ac.th/ee507/flops.pdf.
2. Y. Yang, J. Rao, *IEEE Access*. **7**, 77607–77617 (2019).
3. J. Cheng, M. D. Sacchi, *GEOPHYSICS*. **81**, V89–V101 (2016).
4. L. Chiron, M. A. van Agthoven, B. Kieffer, C. Rolando, M.-A. Delsuc, *Proc. Natl. Acad. Sci.* **111**, 1385–1390 (2014).
5. L. Marcellino, G. Navarra, in *2016 12th International Conference on Signal-Image Technology Internet-Based Systems (SITIS)* (2016), pp. 699–704.
6. G. Laurent, P.-A. Gilles, W. Woelffel, V. Barret-Vivin, E. Gouillart, C. Bonhomme, *Appl. Spectrosc. Rev.* **55**, 173–196 (2020).
7. S. Story, V. Kostin, (2013), available at www2.sscc.ru/Seminars/paper/2014/Intel_MKL.pdf.
8. J. R. Humphrey, D. K. Price, K. E. Spagnoli, A. L. Paolini, E. J. Kelmelis, in *Proc. SPIE 7705* (Orlando, FL, USA, 2010), vol. 7705, pp. 770502, 1–7.
9. C. Bonhomme, C. Gervais, D. Laurencin, *Prog. Nucl. Magn. Reson. Spectrosc.* **77**, 1–48 (2014).
10. R. S. Das, Y. K. Agrawal, *Vib. Spectrosc.* **57**, 163–176 (2011).
11. M. H. Levitt, *Spin dynamics: basics of nuclear magnetic resonance* (John Wiley & Sons Ltd, Chichester, England, 2nd ed., 2008).
12. R. Gautam *et al.*, *Curr. Sci.* **108**, 341–356 (2015).
13. G. Laurent, W. Woelffel, V. Barret-Vivin, E. Gouillart, C. Bonhomme, *Appl. Spectrosc. Rev.* **54**, 602–630 (2019).
14. M. C. D. Tayler, S. (Bas) G. J. van Meerten, A. P. M. Kentgens, P. J. M. van Bentum, *Analyst*. **140**, 6217–6221 (2015).
15. C. Cardell, I. Guerra, *TrAC Trends Anal. Chem.* **77**, 156–166 (2016).
16. M. R. K. Ali *et al.*, *J. Am. Chem. Soc.* **138**, 15434–15442 (2016).
17. J. Marchand, E. Martineau, Y. Guitton, G. Dervilly-Pinel, P. Giraudeau, *Curr. Opin. Biotechnol.* **43**, 49–55 (2017).
18. P. Gemperline, *Practical guide to chemometrics* (CRC Press, Taylor & Francis Group, Boca Raton, FL, USA, 2nd ed., 2006).
19. W. Elmi Rayaleh, PhD dissertation (Université des sciences et technologies de Lille 1, France, 2006), available at http://www.theses.fr/2006LIL10028.
20. X. Chen, M. Vorvoreanu, K. Madhavan, *IEEE Trans. Learn. Technol.* **7**, 246–259 (2014).
21. L. Wang, X. Dong, X. Cheng, S. Lin, *Multimed. Tools Appl.* **78**, 929–945 (2019).
22. S. Wold, K. Esbensen, P. Geladi, *Chemom. Intell. Lab. Syst.* **2**, 37–52 (1987).
23. W. Wu, R. Manne, *Chemom. Intell. Lab. Syst.* **51**, 145–161 (2000).
24. P. Geladi, B. R. Kowalski, *Anal. Chim. Acta*. **185**, 1–17 (1986).
25. P. S. Gromski *et al.*, *Anal. Chim. Acta*. **879**, 10–23 (2015).
26. P. Comon, *Signal Process.* **36**, 287–314 (1994).
27. W. Woelffel *et al.*, *J. Non-Cryst. Solids*. **428**, 121–131 (2015).
28. Y. B. Monakhova, R. Godelmann, T. Kuballa, S. P. Mushtakova, D. N. Rutledge, *Talanta*. **141**, 60–65 (2015).
29. C. Maione, R. M. Barbosa, *Crit. Rev. Food Sci. Nutr.*, 1–12 (2018).
30. M. Wang *et al.*, *Planta Med.*, in press, doi:10.1055/a-0585-5987.
31. D. Vimalajeewa, D. Berry, E. Robson, C. Kulatunga, (New Orleans, LA, USA, 2017), pp. 545–552.

32. F. Othman, Md. S. Chowdhury, W. Z. Wan Jaafar, E. M. M. Faresh, S. M. Shirazi, *Pol. J. Environ. Stud.* **27**, 1659–1671 (2018).
33. S. J. Orfanidis, lecture notes (Optimum signal processing, Electrical & computer engineering department, Rutgers university, USA, 2002), available at http://www.ece.rutgers.edu/~orfanidi/ece525/svd.pdf.
34. J. Shlens, (2014), available at http://arxiv.org/abs/1404.1100.
35. D. W. Tufts, R. Kumaresan, I. Kirsteins, *Proc. IEEE*. **70**, 684–685 (1982).
36. J. A. Cadzow, *IEEE Trans. Acoust. Speech Signal Process.* **36**, 49–62 (1988).
37. G. Laurent, W. Woelffel, V. Barret-Vivin, E. Gouillart, C. Bonhomme, in *Actes du C2i-2016 : 7ème Colloque Interdisciplinaire en Instrumentation* (Saint-Nazaire, France, 2016), pp. 1–8.
38. D. W. Tufts, A. A. Shah, *IEEE Trans. Signal Process.* **41**, 1716–1721 (1993).
39. G. Heinig, K. Rost, in *Algebraic methods for Toeplitz-like matrices and operators* (Birkhäuser Basel, 1984), vol. 13 of *Operator Theory: Advances and Applications*, pp. 9–135.
40. E. R. Malinowski, *Factor analysis in chemistry* (Wiley, Hoboken, NJ, USA, 3rd ed., 2002).
41. J. R. Cavallaro, F. T. Luk, *J. Parallel Distrib. Comput.* **5**, 271–290 (1988).
42. F. Vogt, M. Tacke, *Chemom. Intell. Lab. Syst.* **59**, 1–18 (2001).
43. M. Gu, S. Eisenstat, *SIAM J. Matrix Anal. Appl.* **16**, 79–92 (1995).
44. N. Halko, P. Martinsson, J. Tropp, *SIAM Rev.* **53**, 217–288 (2011).
45. L. Condat, A. Hirabayashi, *Sampl. Theory Signal Image Process.* **14**, 17–47 (2015).
46. M. Bourgoin, E. Chailloux, J.-L. Lamotte, *Int. J. Parallel Program.* **45**, 242–261 (2017).
47. M. Andrecut, *J. Comput. Biol.* **16**, 1593–1599 (2008).
48. S. Lahabar, P. J. Narayanan, (Rome, Italy, 2009), pp. 1–10.
49. V. Novakovic, S. Singer, *BIT Numer. Math.* **51**, 1009–1030 (2011).
50. P. Irofti, B. Dumitrescu, (Lisbon, Portugal, 2014), pp. 271–275.
51. J. Nickolls, I. Buck, M. Garland, K. Skadron, *Queue*. **6**, 40–53 (2008).
52. Khronos group, (Stanford, CA, USA, 2009), pp. 1–54.
53. J. Dongarra *et al.*, in *Numerical Computations with GPUs*, V. Kindratenko, Ed. (Springer International Publishing, 2014), pp. 3–28.
54. M. Gates, S. Tomov, J. Dongarra, *Parallel Comput.* **74**, 3–18 (2018).
55. G. Xie, Y. l Zhang, (Anyang, China, 2017), pp. 15–18.
56. P. P. Man, C. Bonhomme, F. Babonneau, *Solid State Nucl. Magn. Reson.* **61–62**, 28–34 (2014).
57. P. P. Man (2012), available at http://pascal-man.com/navigation/faq-java-browser/SVD-Java-application2012.shtml.
58. P. P. Man (2012), available at http://www.pascal-man.com/navigation/faq-java-browser/SVD-Java-application-GPU.shtml.
59. J. Hofmann, J. Treibig, G. Hager, G. Wellein, in *Proceedings of the 2014 Workshop on Programming Models for SIMD/Vector Processing* (ACM, New York, NY, USA, 2014), *WPMVP '14*, pp. 57–64.
60. M. Faustini *et al.*, *Chem. Mater.* **22**, 4406–4413 (2010).
61. W. J. Malfait, W. E. Halter, *J. Non-Cryst. Solids*. **354**, 4107–4114 (2008).
62. D. Laurencin *et al.*, *ChemistrySelect*. **1**, 4509–4519 (2016).
63. D. Iuga, H. Schäfer, R. Verhagen, A. P. M. Kentgens, *J. Magn. Reson.* **147**, 192–209 (2000).
64. R. W. Schurko, *Acc. Chem. Res.* **46**, 1985–1995 (2013).

65. T. Hudek, M. Sasouvanh, N. Bazan, available at https://msdn.microsoft.com/en-us/Library/Windows/Hardware/ff569918(v=vs.85).aspx.
66. G. van Rossum, Technical report, CS-R9526 (Centrum voor Wiskunde en Informatica, CWI, Amsterdam, NL, 1995), pp. 1–65, available at https://gvanrossum.github.io//Publications.html.
67. G. W. Stewart, available at ftp://math.nist.gov/pub/Jampack/Jampack/AboutJampack.html.
68. P. Javin (2013), available at http://javarevisited.blogspot.com/2013/04/what-is-maximum-heap-size-for-32-bit-64-JVM-Java-memory.html.
69. G. Laurent (2015), available at http://www.culatools.com/forums/viewtopic.php?p=2423#p2423.
70. J. D. van Beek, *J. Magn. Reson.* **187**, 19–26 (2007).
71. G. Laurent, P.-A. Gilles, W. Woelffel, V. Barret-Vivin, E. Gouillart, C. Bonhomme (2018), available at http://doi.org/10.5281/zenodo.1406172.
72. T. E. Oliphant, *A guide to NumPy* (Trelgol Publishing USA, 2006), vol. 1.
73. E. Jones, T. Oliphant, P. Peterson (2001), available at http://www.scipy.org/.
74. L. E. Givon *et al.* (2015), available at http://dx.doi.org/10.5281/zenodo.40565.
75. A. Klöckner *et al.*, *Parallel Comput.* **38**, 157–174 (2012).
76. R. C. Whaley, J. Dongarra, Technical report, UT-CS-97-366 (University of Tennessee, Knoxville, TN, USA, 1997).
77. Z. Xianyi, W. Qian, Z. Yunquan, (Singapore, Singapore, 2012), pp. 684–691.
78. J.-P. Gehrcke (2014), available at https://gehrcke.de/2014/02/building-numpy-and-scipy-with-intel-compilers-and-intel-mkl-on-a-64-bit-machine/.
79. C. Harris, J. Millman, R. Kern, R. Gommers, T. Oliphant (2015), available at https://sourceforge.net/projects/numpy/files/NumPy/1.10.1/.
80. C. Harris *et al.* (2015), available at https://sourceforge.net/projects/scipy/files/scipy/0.16.1/.
81. C. Gohlke, available at http://www.lfd.uci.edu/~gohlke/pythonlibs/.
82. V. W. Lee *et al.*, in *Proceedings of the 37th Annual International Symposium on Computer Architecture* (ACM, Saint-Malo, France, 2010), pp. 451–460.
83. R. Dolbeau, *J. Supercomput.* **74**, 1341–1377 (2018).
84. N. Whitehead, A. Fit-Florea, White paper, TB-06711-001_v10.0 (Nvidia, 2011), pp. 1–7, available at http://docs.nvidia.com/pdf/Floating_Point_on_NVIDIA_GPU.pdf.
85. G. W. Stewart, Technical report, CS-TR-2539, *Technical Reports of the Computer Science Department* (University of Maryland Institute for Advanced Computer Studies (UMIACS), College Park, MD, USA, 1990), available at http://hdl.handle.net/1903/552.
86. G. E. Plassman, Technical report, CR-2005-213500 (NASA Langley Research Center, Hampton, VA, USA, 2005), pp. 1–27, available at https://ntrs.nasa.gov/search.jsp?R=20050192421.
87. P. Läuchli, *Numer. Math.* **3**, 226–240 (1961).
88. G. Golub, W. Kahan, *J. Soc. Ind. Appl. Math. Ser. B Numer. Anal.* **2**, 205–224 (1965).
89. G. H. Golub, C. Reinsch, *Numer. Math.* **14**, 403–420 (1970).
90. J. J. M. Cuppen, *Numer. Math.* **36**, 177–195 (1980).
91. P. R. Willems, B. Lang, C. Vömel, *SIAM J. Matrix Anal. Appl.* **28**, 907–926 (2006).
92. J. Demmel, K. Veselić, *SIAM J. Matrix Anal. Appl.* **13**, 1204–1245 (1992).
93. M. Gu, J. Demmel, I. Dhillon, Technical report, UT-CS-94-257-LAPACK Working Note 88 (Department of Electrical Engineering & Computer Science, The University of Tennessee, Knoxville, USA, 1994), available at https://library.eecs.utk.edu/pub/463.

94. MathWorks parallel computing toolbox team (2012), available at http://fr.mathworks.com/matlabcentral/fileexchange/34080-gpubench.
95. E. Anderson *et al.*, *LAPACK Users' Guide* (Society for Industrial and Applied Mathematics, Philadelphia, PA, USA, 3rd ed., 1999).
96. J. J. Dongarra, J. Du Croz, S. Hammarling, I. S. Duff, *ACM Trans Math Softw*. **16**, 1–17 (1990).
97. L. Hogben, Ed., *Handbook of Linear Algebra* (Chapman and Hall/CRC, Boca Raton, 1st ed., 2006).
98. V. Nguyen (2014), available at http://blog.nguyenvq.com/blog/2014/11/10/optimized-r-and-python-standard-blas-vs-atlas-vs-openblas-vs-mkl/.
99. Anaconda (2016), available at https://www.anaconda.com/.
100. J. J. Helmus, C. P. Jaroniec, *J. Biomol. NMR*. **55**, 355–367 (2013).
101. D. Liu, R. Li, D. J. Lilja, W. Xiao, in *Proceedings of the ACM International Conference on Computing Frontiers* (ACM New York, NY, USA, Ischia, Italy, 2013), *CF '13*, p. 36: 1–10.
102. P. Du *et al.*, *Parallel Comput.* **38**, 391–407 (2012).
103. A. Haidar *et al.*, in *Parallel and distributed processing symposium, 2014 IEEE 28th international* (Phoenix, AZ, USA, 2014), pp. 491–500.
104. D. Goldberg, *ACM Comput Surv*. **23**, 5–48 (1991).
105. ViennaCL, available at http://viennacl.sourceforge.net/.
106. C. Sanderson, R. Curtin, *J. Open Source Softw.* **1**, 26, 1–2 (2016).
107. J. Malcolm *et al.*, in *Modeling and Simulation for Defense Systems and Applications VII, Proc. of SPIE*, E. J. Kelmelis, Ed. (2012), vol. 8403, pp. 84030A, 1–8.
108. S. Tomov, J. Dongarra, M. Baboulin, *Parallel Comput.* **36**, 232–240 (2010).
109. K. Rupp *et al.*, *SIAM J. Sci. Comput.* **38**, S412–S439 (2016).

Chapter VII. General conclusion

'Et maintenant, que vais-je faire ? De tout ce temps, que sera ma vie ?'

Gilbert Bécaud (1927-2001), singer

Solid-state NMR is a domain in constant evolution. Numerous improvements have been achieved among the years, either instrumental or methodological (Chapter I). Static magnetic field strength has been continuously increasing. Ultra-fast Magic Angle Spinning (MAS) is now reaching 170 kHz on development probes. Additionally, new pulse sequences have been developed, allowing homo-/heteronuclear decoupling or recoupling. All these improvements allowed to improve ^{1}H resolution and to increase sensitivity of other nuclei acquired by indirect detection. Spectra of more and more complex molecules such as large proteins were deciphered.

Despite solid-state NMR is a very powerful characterisation technique, its main drawback is its lack of sensitivity. This is especially noticeable for nuclei with a low Larmor frequency, which decreases macroscopic magnetisation. Secondly, strong anisotropies such as large quadrupolar effect broaden peaks over several hundreds of kilohertz and flood them into experimental noise. A third sensitivity case concerns low-abundant species, either due to natural abundance of the studied nuclei, to the low content of these species into the sample, or to sample amount. Since 2010, Dynamic Nuclear Polarisation (DNP) has gained a lot of interest to study sensitivity-limited materials. However, this is not the unique solution and other methodologies can be used, especially in laboratories where such expensive equipments are not available, or in combination with DNP to further tackle the detection limit. In France, two 800 MHz DNP NMR are accessible through the NMR very high fields research infrastructure (Lyon and Paris, FR 3050).

In this dissertation, we focussed on three complementary aspects of sensitivity, namely *instrumentation*, *fast acquisition* and *signal processing*. Instrumentation step was investigated in Chapter II by mean of microcoils allowing to study minute sample quantities (Magic Angle Coil Spinning, MACS). Two approaches were chosen for fast acquisition: Carr-Purcell-Meiboom-Gill echoes (CPMG, Chapter III) for one-dimensional spectra or for direct dimension of multi-dimensional (nD) experiments, and Non-Uniform Sampling (NUS, Chapter IV) for indirect dimensions of nD experiments. Signal processing was performed by spectra denoising with Singular Value Decomposition (SVD, Chapter V). Computation time was analysed on Central Processing Unit (CPU) and on Graphical Processing Unit (GPU) in Chapter VI.

We applied this methodology to materials of chemical interest. The first category of samples was related to biological materials applications with hydroxyapatite, the main inorganic component of mammal's bones and teeth, and the starting point of many kidney stones. Depending on synthesis condition, hydroxyapatite can be crystalline or partly amorphous, and substituted with various ions such as carbonates. Gelatin, the denatured form of the main protein of bone, was chosen for NUS as a sample with very narrow lines due to its mobility. The second category concerned sol-gel chemistry and in particular hybrid organic-inorganic silica. Finally, ^{91}Zr and ^{87}Sr quadrupolar nuclei were used for CPMG and computation, respectively. These nuclei were involved into bioactive glasses.

In Chapter II, microcoils were used to study microquantities of sample with a mass around $m \sim$ 100-200 µg. After reviewing the available options into the literature, we detailed the necessary parameters to prepare microcoils. Wire diameter has to be minimised as much as possible to decrease eddy currents, in the limit of 5 times the skin depth at the frequency of interest. In order to obtain high radio-frequency field homogeneity, the sample has to be confined to only the inner half length of the microcoil. We provided a LibreOffice sheet able to calculate all the necessary parameters for solenoids, with automatic warnings for incorrect values. A ^{1}H 2D spectrum was obtained on a fragment of sea urchin spine placed into a solenoid microcoil. ^{31}P acquisition was compared on Transmission Line Resonators (TLR) and on solenoids, with a time gain of 4.0 and 6.8, respectively. PARIS / PISSARRO pulse sequences allowed to gain 16 % of signal against one-pulse experiments. In order to limit microcoil degradation by radio-frequency, we implemented TEDOR low power experiment, in absence of microcoil, that allowed us to acquire a {^{1}H}-^{31}P 2D spectrum. However, a broad component on the spectrum was missing, maybe due to relaxation, and further tests are needed to improve this pulse sequence. D-HMQC would be a good alternative. Despite TLR are easy to manipulate, their RF efficiency is lower than solenoids, which are hand-made. TLR are more complicated to design and have to be printed externally. An interesting feature of TLR is a radial field, that could allow imaging experiments. A new PhD is starting on microcoils.

In Chapter III, we enhanced solid-state NMR sensitivity during acquisition step, with the help of (Q)CPMG echoes on direct dimension of spectrum. We highlighted that individual echo apodisation and global apodisation are crucial steps to enhance the spectrum. Artefacts can occur in case of wrong delays calibration, and of incorrect synchronisation between echo

and MAS rate. A gain of 10 in sensitivity and thus a gain of 100 in time was not uncommon, at least for quadrupolar nuclei with up to 600 echoes. In case of a few number of echoes, a time gain of 2-3 was rather obtained. Three different approaches were investigated: spikelets, superposition and SVD denoising. The first solution favoured overall peak shape detection, with narrow peaks regularly spaced and overpassing noise. The second solution reconstructed the overall shape, but tails of shapes were difficult to detect, whereas tiny spikelets were easily discernable. Both techniques were sensitive to relaxation rate of different species, leading to non-quantitative results. By combining partial sum and SVD denoising, we proposed a new way to reconstruct a more quantitative shape. We developed a Python program called from TopSpin and able to process data with the three different methods.

In Chapter IV, we were interested in indirect dimensions, for two or higher dimensions experiments. In conventional experiments, acquisition times are quickly increasing, and a compromise has to be chosen between resolution ($AQ > 3\ T_2^*$) and sensitivity ($AQ = 1.26\ T_2^*$). NUS reconciles these two concepts. In order to decrease artefacts, sampling scheme was carefully optimised with Poisson scheme, and reconstruction algorithms were compared, in particular using maximum entropy and compressed sensing. In solid-state NMR, sampling scheme was adapted, for instance by using hybrid sampling. We obtained a gain of ~ 4 in acquisition time on 2D experiments, without degrading the overall spectral shape. Further gains could be obtained on 3D or higher experiments. Finally, we provided some implementation details to use NUSscore, MDDnmr, and TopSpin software.

In Chapter V, SVD denoising was investigated. After reminding some definitions concerning signal-to-noise ratio and limits of detection and quantification, we focussed on matrix shape and on thresholding. Best denoising was obtained on square matrices in time domain and with automatic thresholding provided by Malinowski's indicator. Influence of Lorentzian and Gaussian peak shape was compared, and we proved that peak area was systematically overestimated by ~ 20 % in the latter case. Preprocessing had an impact on peak detection, and cosine apodisation revealed to be a better option than exponential weighting. Whatever peak width and shape, peak height had to be twice higher than maximum of noise for automatic thresholding. As quantitative results required a factor 3 between peak height and maximum of noise for standard processing, this corresponded to a time gain of 2.3. SVD denoising and signal-to-noise measurements with the different formulae have been implemented into Python software and can be called from TopSpin

software. Additional tests were undergone on 2D spectra and highlighted artefacts and amplitude distortion during the process. Further optimisation steps are thus necessary before confident use on 2D spectra.

In Chapter VI, we tried to understand why SVD computation time could vary so much between Java, Matlab and Python languages, and why a low-end GPU of 2008 was surpassing a middle-range CPU of 2013 under Java. Different layers were investigated from the scientist SVD call to the machine execution. Firstly, algorithm layer used different mathematical operations, either or not with the multithreading 'divide and conquer' approach. The algorithm was then calling a library, more or less optimised depending on program version used. Intel MKL and CULA libraries, for CPU and GPU, respectively, were the faster ones tested. Finally, the library was benefiting – or not – from hardware instructions such as SSE3 and AVX2. Single precision was also a key point to gain a factor 2 in time. All in all, a computation time gain of 100 was achieved on the same matrix and hardware. In the continuation of this article, we focussed on computation errors that may be problematic for high dynamic range spectra denoised with single precision. Heteregeneous computing was also tested.

Table VII.1 summarises the time gains achieved with the studied methods. By combining all of them, except computation time, acquisition time could be decreased by ~ 140. Moreover, these approaches could easily be combined with DNP, which gives an impressive enhancement, to further tackle the sensitivity limit of solid-state NMR.

Table VII.1: Time gain measured for each enhancement method.

Domain	**Instrumentation**	**Fast acquisition**		**Signal processing**	
Method	MACS	CPMG	NUS	SVD	CPU/GPU
Time gain	~ 5	~ 3-100	~ 4	~ 2.3	~ 100

And now, what will I do from all this time, what will be my life? The question is very relevant, as we have gained a lot of acquisition and processing time. Reducing acquisition time allows either to analyse more samples in a given time or to focus on more complex ones with further characterisations. Clearly, this is a challenge to circumvent intrinsic low sensitivity of solid-state NMR. However, such developments require a lot of time, and programming is an important need to adequately process the resulting datasets. Hence, the NMR spectroscopist job is evolving from a chemical characterisation support to an expert in programming, signal processing and hardware pitfalls.

List of works

Articles (2)

1. G. Laurent, W. Woelffel, V. Barret-Vivin, E. Gouillart, C. Bonhomme, ‘Denoising applied to spectroscopies – part I: concept and limits’, *Appl. Spectrosc. Rev.* **54**, 602–630 (2019), doi: 10.1080/05704928.2018.1523183.
2. G. Laurent, P.-A. Gilles, W. Woelffel, V. Barret-Vivin, E. Gouillart, C. Bonhomme, ‘Denoising applied to spectroscopies – Part II: Decreasing computation time’, *Appl. Spectrosc. Rev.* **55**, 173–196 (2020), doi:10.1080/05704928.2018.1559851.

Proceeding (1)

1. G. Laurent, W. Woelffel, V. Barret-Vivin, E. Gouillart, C. Bonhomme, ‘SVD Performances to denoise NMR and Raman spectra’, in *Actes du C2i-2016 : 7ème Colloque Interdisciplinaire en Instrumentation* (Saint-Nazaire, France, 2016), pp. 1–8, available at https://hal.archives-ouvertes.fr/hal-01277387.

Invited oral communication (1)

1. G. Laurent and C. Bonhomme, ‘Non-uniform sampling applied to solid-state NMR’, presented at the *Laboratoire des Biomolécules* (Paris, France, 2019).

Oral communications (7)

1. G. Laurent, B. Fassbender, C. Bonhomme, and D. Sakellariou, ‘Apport des microbobines pour l’étude par RMN en phase solide de faibles quantités d’échantillons’, presented at the *Institut des Matériaux de Paris Centre* (Paris, France, 2013).
2. G. Laurent, C. Bonhomme, and D. Sakellariou, ‘Microbobines et RMN en phase solide’, presented at the *Instrumenter et innover en chimie physique pour préparer l’avenir* (Paris, France, 2015), available at https://hal.archives-ouvertes.fr/hal-01139041.
3. G. Laurent, P. Man, and C. Bonhomme, ‘NMR spectra denoising using graphic card’, presented at the *École Doctorale de Physique et Chimie des Matériaux* (Paris, France, 2015).
4. G. Laurent, W. Woelffel, V. Barret-Vivin, E. Gouillart, and C. Bonhomme, ‘SVD Performances to denoise NMR and Raman spectra’, presented at the *C2i-2016 : 7ème Colloque Interdisciplinaire en Instrumentation* (Saint-Nazaire, France, 2016), available at https://hal.archives-ouvertes.fr/hal-01277387
5. G. Laurent, W. Woelffel, V. Barret-Vivin, E. Gouillart, and C. Bonhomme, ‘Spectra denoising: using graphic cards ?’, presented at the *RMN structurale dans le bassin parisien* (Thiais, France, 2016), available at https://hal.archives-ouvertes.fr/hal-01277415.
6. G. Laurent and C. Bonhomme, ‘Non-uniform sampling applied to solid-state NMR’, presented at the *RMN structurale dans le bassin parisien* (Orléans, France, 2018), available at https://hal.archives-ouvertes.fr/hal-01745319.

7. G. Laurent, W.-C. Teh, J. A. Lehmann-Horn, D. Sakellariou, and C. Bonhomme, ‘Microcoils to improve solid-state NMR sensitivity’, presented at the *XXXIème Congrès du GERM, Approches rapides en RMN* (Saint-Pierre d’Oléron, France, 2019), available at https://hal.archives-ouvertes.fr/hal-02156281.

Posters (3)

1. G. Laurent, P. Man, and C. Bonhomme, ‘Spectra denoising using graphic cards’, presented at the *XXVIème conférence du Groupement d’Études de Résonance Magnétique (GERM) - simulation et modélisation en RMN* (Sète, France, 2015), available at https://hal.archives-ouvertes.fr/hal-01168997.
2. G. Laurent and C. Bonhomme, ‘Non-uniform sampling & denoising applied to nuclear magnetic resonance’, presented at the *MacSeNet/SpaRTan spring school on sparse representations and compressed sensing* (Ilmenau, Germany, 2016, international), available at https://hal.archives-ouvertes.fr/hal-01516795.
3. G. Laurent and C. Bonhomme, ‘Improving solid-state NMR sensitivity using instrumentation, fast acquisition and post-processing’, presented at the *EUropean MAgnetic Resonance meeting (EUROMAR)* (Nantes, France, 2018, international), available at https://hal.archives-ouvertes.fr/hal-01832993.

Programs (3)

1. G. Laurent, P.-A. Gilles, ‘Post Processing of NMR data.’ (2017), available at https://github.com/gul916/NMR_post_proc.
2. G. Laurent P.-A. Gilles, W. Woelffel, V. Barret-Vivin, E. Gouillart, C. Bonhomme, ‘Denoising applied to spectroscopies: parts I and II - programs and datas’ (2018), available at http://doi.org/10.5281/zenodo.1406172.
3. G. Laurent, C. Bonhomme, ‘LibreOffice sheet to design microcoils for MACS’ (2020), available at http://doi.org/10.5281/zenodo.3598539.

[Augmentation de la sensibilité de la RMN en phase solide: instrumentation, acquisitions rapides et traitement du signal]

Résumé :

La Résonance Magnétique Nucléaire (RMN) en phase solide souffre d'une faible sensibilité intrinsèque, malgré les récentes améliorations. Les approches instrumentales, d'acquisition rapide et de traitement du signal ont été examinées afin de remédier à ce défaut le plus efficacement possible. Premièrement, les microbobines (bobine en rotation à l'angle magique, MACS) ont été placées dans un rotor et couplées inductivement à la bobine de la sonde standard. Un gain en temps de ~ 5 a été obtenu sur des microquantités avec une masse m ~ 100-200 µg. Deuxièmement, le temps d'acquisition a été diminué grâce aux échos Carr-Purcell-Meiboom-Gill (CPMG) pour l'acquisition directe. Un traitement du signal adéquat est nécessaire pour tirer le meilleur de cette technique. Nous avons produit un programme Python pour traiter les données soit en utilisant les méthodes standard de peigne ou de superposition, soit avec une méthode de débruitage. Un gain en temps de ~ 3-100 a été possible. Troisièmement, l'échantillonnage non-uniforme (NUS) a été choisi comme un moyen de diminuer le temps d'acquisition des dimensions indirectes des expériences multi-dimensionnelles. L'échantillonnage Poisson s'est avéré le meilleur choix pour limiter les artefacts, tandis que l'échantillonnage hybride s'est révélé efficace sur les spectres présentant à la fois des pics larges et fins. Un gain en temps de ~ 4 a été obtenu. Quatrièmement, les spectres ont été traités avec le débruitage par décomposition en valeurs singulières (SVD). Nous avons mis en avant une surestimation des pics gaussiens de ~ 20 %. Le seuillage automatique a été mis en place, donnant un gain en temps de ~ 2,3. Enfin, le temps de calcul a été étudié et diminué par ~ 100 en combinant l'algorithme "diviser pour mieux régner", des bibliothèques optimisées, l'appel aux instructions matérielles et la simple précision. Une comparaison entre les unités centrales (CPU) et les cartes graphiques (GPU) a été fournie.

Mots clés : [RMN en phase solide, sensibilité, microbobines MACS, échos CPMG, échantillonnage non-uniforme NUS, débruitage SVD, carte graphique]

[Increasing solid-state NMR sensitivity: instrumentation, fast acquisitions and signal processing]

Abstract:

Solid-state Nuclear Magnetic Resonance (NMR) is suffering from an intrinsic low sensitivity, despite recent improvements. Instrumentation, fast acquisition and signal processing approaches were investigated to circumvent this drawback as far as possible. Firstly, microcoils (Magic Angle Coil Spinning, MACS) were placed into rotors and inductively coupled to the standard probe coil. A time gain of ~ 5 was obtained for microquantities with a mass m ~ 100-200 µg. Secondly, acquisition time was decreased by mean of Carr-Purcell-Meiboom-Gill (CPMG) echoes for direct acquisition. Adequate processing is required to get the best enhancement from this technique. We provided a Python software to process data either using standard spikelets or superposition methods, or with a denoising method. A time gain of ~ 3-100 was possible. Thirdly, Non-Uniform Sampling (NUS) was chosen as a way to decrease acquisition time of indirect dimensions of multi-dimensional experiments. Poisson sampling revealed to be the best choice to limit artefacts, whereas hybrid sampling proved to be efficient on spectra with both broad and narrow peaks. A time gain of ~ 4 was achieved. Fourthly, spectra were processed with Singular Value Decomposition (SVD) denoising. We highlighted an overestimation of Gaussian peaks by ~ 20 %. Automatic thresholding was implemented, giving a time gain of ~ 2.3. Finally, computation time was decreased by ~ 100 by combining 'divide and conquer' algorithm, optimised libraries, hardware instruction calls and single precision. A comparison between Central Processing Units (CPU) and Graphical Processing Units (GPU) was provided.

Keywords: [Solid-state NMR, sensitivity, microcoils, CPMG echoes, non-uniform sampling NUS, SVD denoising, graphic card]