Laboratory of Organic Chemistry
Department of Chemistry
Faculty of Science
University of Helsinki, Finland

Automated Analysis of Quantitative NMR Spectra

Valtteri Mäkelä

Academic Dissertation

To be presented, with the permission of the Faculty of Science of the University of Helsinki, for public examination in Auditorium A129, Department of Chemistry, A.I. Virtasen aukio 1, on 8th of January 2016, at 12 noon.

Helsinki 2015

Supervisors

Dr. Sami Heikkinen
Department of Chemistry
University of Helsinki
Finland

Prof. Ilkka Kilpeläinen
Department of Chemistry
University of Helsinki
Finland

Reviewers

Prof. Perttu Permi
Biological and Environmental Science/Chemistry
Nanoscience Center
University of Jyväskylä
Finland

Prof. Juha Vaara
Department of Physics and Chemistry
University of Oulu
Finland

Opponent

Prof. James Keeler
Department of Chemistry
University of Cambridge
United Kingdom

ISBN 978-951-51-1834-9 (paperback)
ISBN 978-951-51-1835-6 (PDF)

Keuruskopio 2015

Abstract

NMR spectroscopy is an invaluable tool for structure elucidation in chemistry and molecular biology, which is able to provide unique information not easily obtained by other analytical methods. However, performing quantitative NMR experiments and mixture analysis is considerably less common due to constraints in sensitivity/resolution and the fact that NMR observes individual nuclei, not molecules. The advances in instrument design in the last 25 years have substantially increased the sensitivity of NMR spectrometers, diminishing the main weakness of NMR, while increases in field strength and ever more intricate experiments have improved the resolving power and expanded the attainable information. The minimal need for sample preparation and its non-specific nature make quantitative NMR suitable for many applications ranging from quality control to metabolome characterization. Furthermore, the development of automated sample changers and fully automated acquisition have made high-throughput NMR acquisition a more feasible and attractive, yet expensive, possibility.

This work discusses the fundamental principles and limitations of quantitative liquid state NMR spectroscopy, and tries to put together a summary of its various aspects scattered across literature. Many of these more subtle features can be neglected in simple routine spectroscopy, but become important when extracting quantitative data and/or when trying to acquire and process vast amounts of spectra consistently.

The original research presented in this thesis provides improved methods for data acquisition of quantitative ^{13}C detected NMR spectra in the form of modified INEPT based experiments (Q-INEPT-CT and Q-INEPT-2D), while software tools for automated processing and analysis of NMR spectra are also presented (ImatraNMR and SimpeleNMR). The application of these tools is demonstrated in the analysis of complex hydrocarbon mixtures (base oils), plant extracts and blood plasma samples.

The increased capability of NMR spectroscopy, the rising interest in metabolomics and for example the recent introduction of benchtop NMR spectrometers are likely to expand the future use of quantitative NMR in the analysis of complex mixtures. For this reason, the further development of robust, accurate and feasible analysis methods and tools is essential.

Preface

This book ended up being a great deal longer than I had anticipated, which lead to some anxiety and a drastically increased consumption of cola-based soft drinks in the spring and summer of 2015. I also missed much of the daylight and normal working hours during this period, but luckily the Finnish weather failed as well. In the end I think this resulted in a thicker, but more readable book for the "average" chemist, which might serve as an introduction to quantitative NMR or clarify some of the fundamental aspects involved. I was also able to include a lot of figures, which I wish I would have been able to show when explaining NMR in various contexts throughout the years.

I was introduced to NMR spectroscopy in the summer of 2004, and as a first acknowledgment I want to thank professor Ilkka Kilpeläinen for doing so. I was quite rapidly pushed to the deep end of the pool with the task of programming the pulse sequence for a quantitative HSQC experiment, but it turned out to be a valuable experience in the subsequent years, when I actually learned more about the inner workings and the theory behind NMR. I'm also grateful to Ilkka for providing me the freedom of being able to work on a diverse set of topics which interested me, as my excursion into NMR also led me to learn some electronics, signal processing and sparked an interest in data-analysis.

Secondly, I want to thank Dr. Sami Heikkinen, as he is behind most of my knowledge in NMR. As acknowledged in many other theses and thank you speeches, Sami has always had time to explain, provide references and discuss the various aspects of the complex field. I have also appreciated the several discussions on NMR, chemistry, the state of the IT support and existential music while enjoying beverages bought with cellular phones.

This book was reviewed by professors Perttu Permi and Juha Vaara, for whom I'm grateful for the provided feedback and quick responses. I'd like to especially thank JV for his meticulous work under stressful circumstances. During the finalizing of this book, Dr. Harri Koskela also provided a lot of insights, copies of old articles and expert NMR knowledge for which I am very grateful.

From the Chemistry Department I'd like to mention Jussi Helminen, who has shared the workspace with me for several years. This has lead to many

discussions regarding organic chemistry, computational methods and science in general, while many pizzas and epic youtube videos have been consumed. Alistair, Gudrun, Jari, Maarit, Mari, Outi, Pirkko, Risto and Sampo have provided a great introduction on how the university and the science community itself works, and I'd like to also thank for all of the discussions and collaborations. Arno, Ashley, Kashmira, Maiju, Mikko, Outi, Raisa, Taru, Tiina, Tom, Tuomas, Uula and other peers from the trenches, thanks for all the fun and beer.

There are also many important people from the Viikki campus, where I originally came just to study a minor subject in biochemistry. Lauri Vaahtera and Martta Viljanen are hopefully lifelong friends, and I also appreciate the collaboration in science, concepts in life decisions and baked goods. They were also present in the original formation of the ska-band Kerjääjät, which introduced me to playing the trombone. I really can't think of my life without it anymore, so cheers for the good times and bad jokes to Antti, Kärppä, Markkukapseli and Otso! Anni "Banni" Joensuu has proofread this book and supported the making of it in several ways, but I am also grateful for all of the awesome moments and spontaneous journeys. A lot of other important friends also resulted from the activities of Helix and Symbioosi. Iiro, Nalle, Tuomas, Samu: you can be my wingmen anytime. Aino, Arttu, Elli, Minna, Tea, Salla, Suvi, Vuokko: thanks all for being present and understanding. Anna, Ansq, Henna, Krista, Lina, Marja, Jussi, Jukka and all who I forgot because of the deadline: you rule. There is just so many of you.

The rest of the acknowledgments are harder to group but as I have a chance to thank people in a book with an ISBN number, I'll use it. I'm grateful to my parents for all the support during my life and for the encouragement to pursue science and my other interests. Minna and Kalle for the Eurovision song contest. Hede, Jaakko, Janos, Kasse, Murre, Raevaara and Ville for all the experiences and still being present and relatively ok. Ilari for explaining quadtrees, thank you and good luck. Jimmy and Tuukka for the blues. Janne for teaching. Hiljamaija for all the music, and for being a part of my life. Anu, Hanna, Jarkko, Jannika, Kaisamaija & Peetu: we'll get some beers.

Salla for being in a hammock with me in July. And hammocks in general. They rule.

Valtteri Mäkelä

List of original publications

I Valtteri Mäkelä, Ilkka Kilpeläinen, Sami Heikkinen, Quantitative ^{13}C NMR spectroscopy using refocused constant-time INEPT, Q-INEPT-CT. *Journal of Magnetic Resonance*, **2010**, 204, 124-130. http://dx.doi.org/10.1016/j.jmr.2010.02.015

II Valtteri Mäkelä, Outi Heikkilä, Ilkka Kilpeläinen, Sami Heikkinen, ImatraNMR: Novel software for batch integration and analysis of quantitative NMR spectra. *Journal of Magnetic Resonance*, **2011**, 211, 186-194. http://dx.doi.org/10.1016/j.jmr.2011.05.012

III Valtteri Mäkelä, Pirkko Karhunen, Sakari Siren, Sami Heikkinen, Ilkka Kilpeläinen, Automating the NMR analysis of base oils: Finding napthene signals. *Fuel*, **2013**, 111, 543-554. http://dx.doi.org/10.1016/j.fuel.2013.04.020

IV Valtteri Mäkelä, Lauri Vaahtera, Jussi Helminen, Harri Koskela, Mikael Brosché, Ilkka Kilpeläinen, Sami Heikkinen, Automated Processing and Statistical Analysis of NMR spectra obtained from *Arabidopsis thaliana* Exctracts. *Manuscript*.

Author contributions

I The author performed all of the experimental work, including the acquisition of the spectra, pulse sequence programming and data-analysis of the results. SH was responsible for the principal pulse sequence design and carried out product operator calculations, while the author performed the modulation optimization. IK and SH also assisted in operating the spectrometer and in the analysis of the spectra. IK supervised the work. The author drafted the manuscript which was revised and accepted by all authors.

II The author designed and programmed the presented software tools and performed the demonstrated data-analysis. OH and SH performed the experimental work and data-acquisition of blood plasma samples, while the author acquired all base oil spectra. SH and IK supervised the work. The manuscript was drafted by the author and edited by SH and IK, with all authors approving the final version.

III All experimental work related to NMR was performed by the author, including the pulse-sequence programming of Q-INEPT-2D. The author was also mainly responsible for the experiment design, spectrum processing and data-analysis, including the formulation of presented average structural parameters. SH was responsible for the theoretical aspects of pulse sequence design and assisted in the acquisition, processing and interpretation of the spectra. PK and SS designed and executed the synthesis of the model compounds. SH and IK supervised the work. The author drafted the manuscript, which was edited by SH and IK and approved by all authors.

IV The author acquired all NMR spectra, executed the data-analysis and developed the used software tools. LV carried out picking of the *Arabidopsis* plant lines and preparation of the samples, and was also responsible for the biochemical aspects of the experiments and results. JH synthesized the model compounds and provided related insights. HK performed the spectrum predictions on the suspected compounds. The work was supervised by SH and IK, while MB supervised the work of LV. The author drafted the manuscript with parts contributed by LV and JH. The final manuscript was edited and approved by all authors.

Contents

List of Acronyms

Colophon

This document was typeset using the LaTeX typesetting system, XeLaTeX engine, and the memoir class created by Federico Maggi and Peter Wilson. Adobe Illustrator in conjunction with Python/Matplotlib were used for most figures. The body text is set at 10pt with Adobe Garamond Pro. Additionally the Doves press font is used, revived from the bottom of the river Thames by Robert Green. Unbelievable.

Introduction 1

> "I can't explain that attraction in terms of anything else that's familiar to you. For example, if we said the magnets attract like if rubber bands, I would be cheating you. Because they're not connected by rubber bands. I'd soon be in trouble. And secondly, if you were curious enough, you'd ask me why rubber bands tend to pull back together again, and I would end up explaining that in terms of electrical forces, which are the very things that I'm trying to use the rubber bands to explain. So I have cheated very badly, you see. So I am not going to be able to give you an answer to why magnets attract each other except to tell you that they do."
>
> --Richard Feynman

Nuclear Magnetic Resonance (NMR) spectroscopy is perhaps the most useful analysis technique available for structure determination of organic compounds, and it is widely used in traditional organic chemistry, as well as in structure elucidation of natural products, proteins and other biomolecules. During the last 20-30 years, significant improvements in instruments and analysis techniques have diminished much of the traditional weakness of NMR spectroscopy: sensitivity. This has broadened the application of NMR and made quantitative and multidimensional NMR more feasible

and common in practice, while reducing the sample amount and time requirements.

This thesis explores liquid state NMR spectroscopy and the principles and methods involved when quantifying compounds in simple or complex mixtures. Not much emphasis is placed on structure elucidation or elaborate multidimensional NMR experiments, instead, the fundamental aspects of quantitative NMR and sensitivity are discussed in more detail. In addition, aspects of automatic acquisition, processing and data-analysis of NMR spectra are examined, as they are central when using NMR spectroscopy to analyze large sample sets.

1.1 NMR phenomena

NMR is based on observing the magnetic properties of atomic nuclei, and using those nuclei to probe the surrounding electron structure and other nuclei, which is of course the primary interest for chemists. At the heart of NMR phenomena is the fact that nuclei possess the quantum mechanical property of nuclear spin angular momentum, or *spin*, inherited from quarks making up the protons and neutrons in the nuclei [1]. The nuclear spin is described by the spin quantum number I, and depending on this number the spin is quantified to a limited number of states [1, 2]:

$$N_{states} = 2I + 1 \tag{1.1}$$

where I can take positive integer or half-integer values ($\frac{1}{2}, 1, \frac{3}{2}...$). The value of I is determined by the lowest energy (ground state) spin configuration of the protons and neutrons in the nuclei, which is not trivially resolved and can be regarded as an experimental quantity [1]. Only few simple rules can be deduced, such as that nuclei with even mass number have integer spin while odd mass number leads to half-integer spin, and nuclei with even number of protons and neutrons have zero spin [1]. The most often observed nuclei in NMR such as ^{1}H have a spin quantum number of $\frac{1}{2}$, and can thus have only two states:

$$2 * \frac{1}{2} + 1 = 2 \tag{1.2}$$

which can be described as the spin being in $+\frac{1}{2}$ or $-\frac{1}{2}$ state. Nuclei with non-zero spin possess magnetic moment and can be thought as small magnets. If an external magnetic field is affecting the nuclei, the states have an energy difference ΔE, which is given by [2, 3]:

$$\Delta E = \frac{h\gamma B_0}{2\pi} = \hbar\gamma B_0 \tag{1.3}$$

where h is the Planck constant, B_0 is the applied magnetic field and γ the *gyromagnetic ratio* of the nucleus. The gyromagnetic ratio essentially describes how "strong" a magnet the nucleus is, and is fixed for particular type of nuclei. As the other terms are also constants, the energy difference between the spin states for certain type of nuclei depends purely on the applied external magnetic field.

1.1.1 Spin populations and polarization

The states of the spin can also be described more intuitively to be aligned with or opposing the external magnetic field B_0. The sign of the gyromagnetic ratio describes the direction of the magnetic moment relative to the spin, and as perhaps expected, the state in which the magnetic moment is aligned with the external magnetic field has a smaller energy level than the opposing state. As a result, a slight excess of spins tend to align with the magnetic field in thermodynamical equilibrium, following the Boltzmann distribution [2]:

$$\frac{N_\beta}{N_\alpha} = e^{-\Delta E/k_b T} \tag{1.4}$$

where N_α and N_β are the populations, ΔE is the energy difference between states, k_b is the Boltzmann constant and T is the thermodynamic temperature. In the above example, γ is positive and α is the low energy state while β is the high energy state. By substituting the energy difference from Eq. 1.3 to the above formula, one can calculate the fraction of spins in the low energy level (α), which is surprisingly small even with high magnetic fields produced by modern superconducting electromagnets. For example with $\sim$14.1 Tesla (or "600 MHz") magnet and ^{1}H nuclei:

$$\frac{N_\beta}{N_\alpha} \cong 0.99990402 \tag{1.5}$$

or in other terms, for 1 000 000 ^{1}H nuclei in the β state, there exist only 1 000 096 nuclei in the lower energy α state. To describe this excess, *polarization* (P) can be defined [4]:

$$P = \frac{N_\alpha - N_\beta}{N_\alpha + N_\beta} = \frac{N_\alpha - N_\beta}{N_{total}} \tag{1.6}$$

which describes the fraction of excess nuclei in the α state. In the above scenario, polarization is 0.0048% or 48 *Parts Per Million* (ppm), quite a small excess indeed.

1.1.2 Larmor frequency

The magnetic moments of the nuclei also precess in the external magnetic field in a specific frequency ν_L given by [2]:

$$\nu_L = \frac{\gamma B_0}{2\pi} \text{ or } \nu_L = \bar{\gamma} B_0 \tag{1.7}$$

where B_0 is the magnetic field in Tesla and γ is the gyromagnetic ratio in $rad\ s^{-1}T^{-1}$ (or alternatively $\bar{\gamma}$ in $MHz\ T^{-1}$). For example, $\bar{\gamma}$ of ^{1}H nucleus is ∼42.576 MHz/T [3], so the magnetic moment of ^{1}H nuclei precesses at the frequency of ∼42.6 MHz in a magnetic field of 1 Tesla. This frequency is called the *Larmor frequency* or the *resonance frequency*, and its relevance becomes evident if we calculate the energy of a photon which matches the Larmor frequency. The energy of a photon is determined solely by its frequency:

$$E = h\nu \tag{1.8}$$

and substituting the frequency ν with the Larmor frequency (Eq. 1.7), we obtain:

$$E = h\nu_L = h\frac{\gamma B_0}{2\pi} \tag{1.9}$$

yielding the same equation given earlier for the energy difference between spin states (Eq. 1.3). The energy of a photon matching the Larmor frequency corresponds exactly to the energy difference of the spin states of the nucleus. With strong magnetic fields achievable by modern superconducting electromagnets, these photons fall in the *Radio Frequency* (RF) range of the electromagnetic spectrum. This essentially means that the spin states can be manipulated by radio waves, as the nuclei resonate in the Larmor frequency and can absorb or emit photons as they change state. However, a more appropriate picture of what is actually happening in the NMR spectrometer can be understood by precession in magnetic fields and induction [5], as discussed in the next sections.

1.1.3 Magnetization and RF pulses

In thermodynamic equilibrium, a slight excess of spins are in the low-energy (α) state aligned with the external magnetic field, with the fraction of excess spins described by polarization (Eq. 1.6). The spins are precessing at the Larmor frequency and the individual magnetic moment vectors combine forming *net magnetization* or *bulk magnetization* vector M_0, as

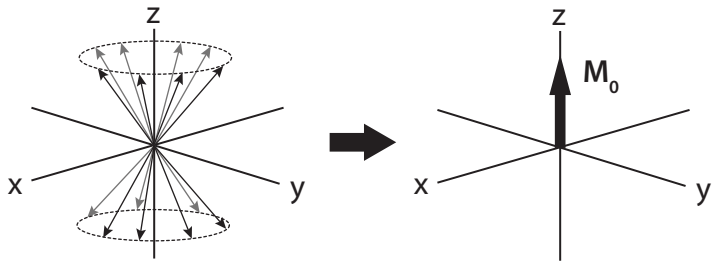

FIGURE 1.1: The individual magnetic moments of nuclei precess in the external magnetic field, which is aligned with the z-axis. The slight excess of vectors aligned with the field lead to net magnetization M_0 parallel with the external field. The figure is modeled after work of Timothy D.W. Claridge [6].

seen in Figure 1.1. The strong external magnetic field causes the sample to become essentially a (very) weak magnet.

Applying RF radiation with Larmor frequency to the sample can cause some of the spins to change state, thus affecting the polarization and the resulting net magnetization vector. For example, if the population difference is reversed, the net magnetization is flipped from the +z-axis to the -z-axis. However, a more accurate description of the effect of RF radiation is that it rotates the net magnetization vector around the axis from which it is applied. This can be understood as a similar precession of spins in a magnetic field as the precession in the static external magnetic field, only this time the magnetic field is a component of the RF wave.

The magnetic field of the RF waves is oscillating with the electric field, but if the frequency of the RF wave matches the Larmor frequency exactly, the field is static from the reference point of the precessing spins. The magnetic field causes the spins to precess analogously to the main static field, and the combined effect is rotation of the net magnetization vector. In many cases, describing phenomena in this kind of *rotating frame* of reference with the frequency of the transmitter is very practical in NMR, and the manipulation of magnetization by radio frequency pulses forms the very core of modern NMR spectroscopy.

The effect of RF radiation described by rotations in magnetization vectors is called the *vector model* or the Bloch vector model [6], and it is very useful for describing simple NMR phenomena. For more complex experiments this model is not enough and tools such as *product operator formalism* and *density matrix* are needed [7, 8, 9], while the operations still resemble vector movements of this simple model.

1.1.4 Single-pulse experiment

The most basic elements in NMR experiments are short RF pulses which rotate the magnetization by a certain amount, as seen on Figure 1.2. For example, a RF pulse which rotates the magnetization by 90° unsurprisingly is called a 90° pulse. If magnetization at thermal equilibrium is rotated by 90°, there is no net magnetization in the z-axis any more, instead the spins now precess coherently forming a rotating magnetization vector in the transverse xy-plane. From the point of view of individual spins, the precession ensues with Larmor frequency as usual, but now the net magnetization is also precessing with the Larmor frequency.

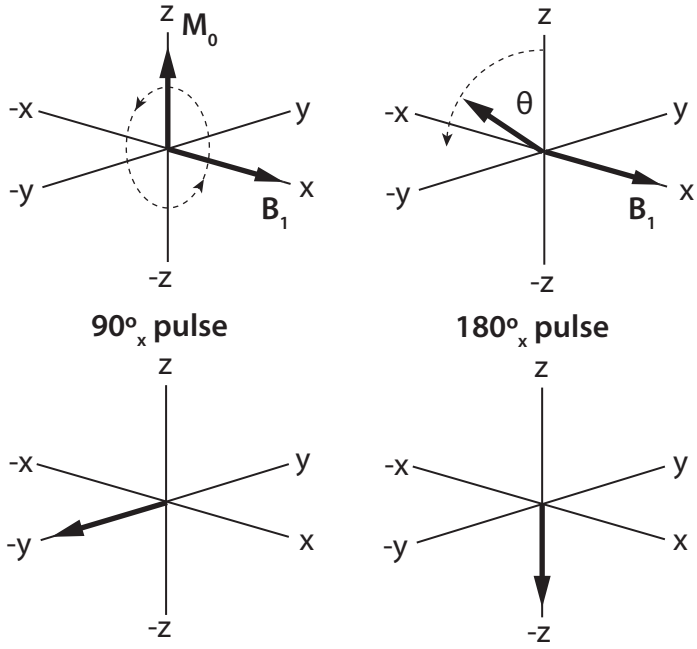

FIGURE 1.2: RF radiation causes the net magnetization vector M_0 to rotate, as the spins precess in the magnetic field component of the RF radiation (B_1) similarly as they precess around the static magnetic field (top). The rotation depends on the length and power of the RF radiation, and the simplest pulses rotate the magnetization for 90° or 180° (bottom). The direction of the precession around magnetic field vector depends on the sign of γ [1], but in NMR literature pulse with phase x simply denotes a pulse with positive rotation around x-axis (right-hand rule). Here the correct B_1 field for negative γ is shown for simplicity, in which case the rotation and magnetization vectors are parallel, while the situation is opposite nuclei with positive γ [9]. The figure is modeled after the work of Timothy D.W. Claridge [6].

The oscillating magnetic field induces a RF signal in the detector coil, which can be detected by a sensitive receiver. The observed signal decays as the rotating spins lose coherence and dephase with each other, and the net magnetization vector returns back to the +z-axis as the thermal equilibrium

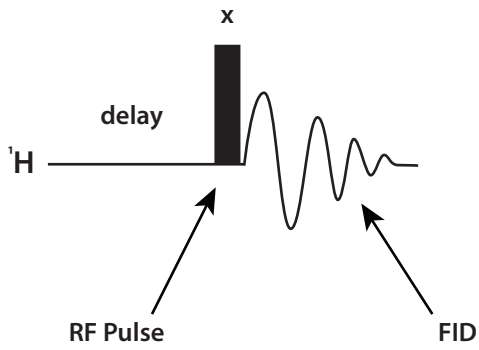

Figure 1.3: The *pulse sequence* of basic single-pulse experiment. The net magnetization vector is rotated by 90° around x-axis by RF pulse (black bar), and the resulting RF signal is recorded.

is re-established (through *relaxation* processes, section 3.1.1). This process (and the resulting data) is called *Free Induction Decay* (FID). The single pulse and observation forms the most simple modern NMR experiment, and is illustrated in Figure 1.3.

1.1.5 Chemical shift

Observing the resonance frequency of nuclei is not very interesting to chemists if all nuclei of the same type have the same frequency. Fortunately (perhaps unfortunately for physicists), it turns out that this is not the case: the surrounding electrons also have spin and magnetic moment (both intrinsic and orbital), and react to the external magnetic field, altering the magnetic field experienced by the nuclei. The surrounding electrons create an opposing magnetic field, and *shield* the nucleus from the external magnetic field. As the Larmor frequency depends on the magnetic field experienced by the nuclei (Equation 1.7), the shielding changes the observed Larmor frequency. This change of frequency is called *chemical shift* as the electron structure surrounding the nuclei is a direct result of the chemical structure of the molecule.

The chemical shift is the principal reason why NMR spectroscopy can distinguish between the same type of nuclei and can provide direct evidence of the chemical structure at hand. The electrons react more strongly to

larger magnetic fields, so chemical shift is relative to the applied magnetic field. Because of this, the chemical shift (σ) is not usually expressed in absolute difference of resonance frequency in Hz, instead it is related to the resonance frequency of a reference compound. In this way, results obtained with different field strengths can be compared easily [1]:

$$\sigma = \frac{\nu_{sample} - \nu_{reference}}{\nu_{reference}} \tag{1.10}$$

The result is a unitless value, and as chemical shifts are small compared to the resonance frequency, they are usually expressed in ppm.

1.1.6 J-coupling or spin-spin coupling

In addition to electron shielding, nuclei can react to other nearby NMR active nuclei. The most important mechanism especially in ^{1}H NMR is *J-coupling*, where the spin state of neighboring ^{1}H nuclei alter the energy level and thus resonance frequency of the ^{1}H nuclei in question. This is meditated via chemical bonds by electrons and can be significant through 1-3 bonds. In practice it causes the signals to split in certain patterns, complicating the spectrum, as seen in Figure 1.4. Coupling can provide a lot of information about the chemical bonds and can be also used as a means of transferring the magnetization through the bonds (*polarization transfer*), a technique which most 2D NMR experiments rely on, and which is discussed in more detail in section 2.2.7.

The effects of coupling can be suppressed by changing the spin state of the coupled nuclei very rapidly, which can be achieved by irradiating the nuclei constantly (*decoupling*). For example in ^{13}C NMR, the carbon nuclei are usually surrounded with many NMR active ^{1}H nuclei. This causes complex coupling patterns, as in addition to the directly bonded hydrogens, the ^{1}H nuclei bonded to neighboring carbons are also coupled significantly. By radiating the sample with the Larmor frequency of the ^{1}H nuclei, ^{13}C nuclei remain intact while the ^{1}H spins rapidly change state. From the perspective of the carbon, the effect of the coupled ^{1}H nuclei averages to zero, and the coupling is suppressed. The result is a clean decoupled ^{13}C spectrum with all signals observed as singlets. The other ^{13}C nuclei are not usually a problem, as the natural abundance of ^{13}C is $\sim$1 %, so the probability of adjacent ^{13}C nuclei is low. In a good spectrum, these

satellite signals can be observed (with a low intensity of ∼1:200 compared to the main signal).

1.1.7 Nuclear Overhauser Effect

The magnetic moment of spins can also directly affect each other (*dipole-dipole coupling*), leading to another significant mechanism of interaction between spins. While this coupling is averaged to zero due to the molecular motion in liquids, the effect of this coupling can be seen during relaxation, when magnetization is transferred via this mechanism, a process called *Nuclear Overhauser Effect* (NOE). This can be exploited to find evidence of proximity of nuclei despite no direct chemical bonds being present, which is especially useful for determining three dimensional structure and stereochemistry of molecules. The NOE mechanism is discussed with more detail in section 2.2.6.

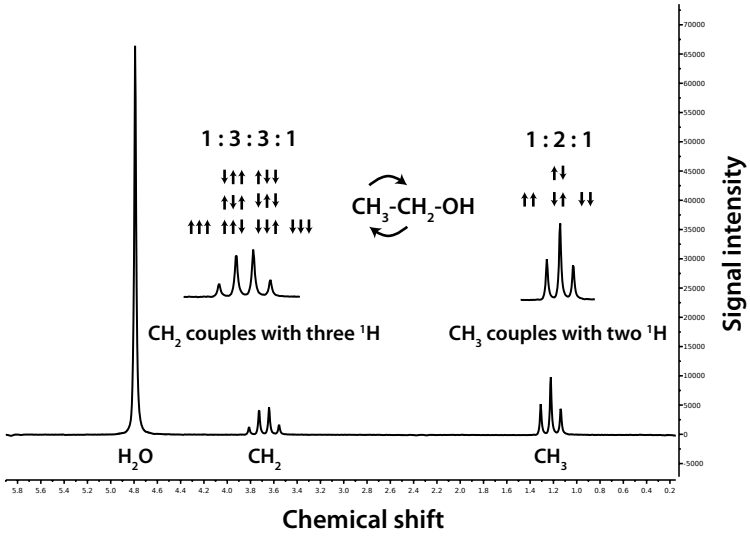

FIGURE 1.4: The basic phenomena shaping the NMR spectrum. Chemical shift (the location of the signals) and J-coupling (the splitting of signals due neighboring nuclei). The signals are split depending on how many different permutations of spin states the neighboring nuclei can possess, as illustrated. The sample is 38% (vol.) ethanol:water mixture, a neat solution of traditional Finnish alcoholic beverage *Jaloviina*, measured by 80 MHz PicoSpin benchtop NMR spectrometer.

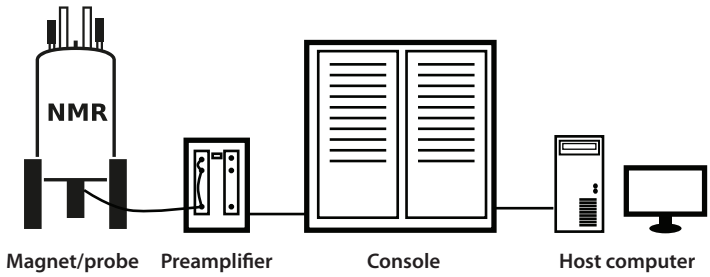

Figure 1.5: The basic components of a modern (high-field) NMR instrument, as seen by the user.

1.2 NMR spectrometers

Modern NMR spectrometers are complex instruments based on high-field superconducting magnets and sophisticated RF electronics. The precision needed to build and maintain these instruments have made them expensive compared to many other analytical instruments, but the fact that they are found in increasing numbers in laboratories around the world is a testament to the usefulness of NMR spectroscopy.

Lately also cheap "benchtop" NMR instruments based on permanent magnets are appearing in teaching, routine analysis, chemometry and quality control applications [10, 11, 12]. These instruments operate in a much lower field (42-80 MHz), but require very little maintenance and can cost approximately one tenth of a high field (>500 MHz) instrument [10]. Even open independent designs implementing all central components with *Field-Programmable Gate Array* (FPGA) chips and other off-the-shelf parts have been reported [13, 14].

The basic design of a modern high-field NMR instrument is given in Figure 1.5 along with a rough schema of the internal components in Figure 1.6. The same basic principles apply to practically all modern NMR instruments, however the components may be contained in single case (as in benchtop spectrometers) or distributed differently.

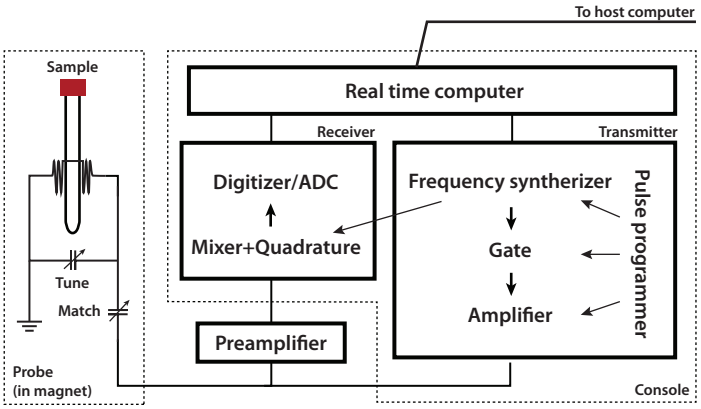

FIGURE 1.6: The basic internal design of a modern FT-NMR instrument [15, 6]. Only a single channel and no locking electronics is shown.

1.2.1 Overview of the acquisition process

The basic acquisition process of a NMR spectrum involves the following phases:

1. The sample material is diluted to a suitable solvent, which is usually chosen from "NMR solvents", common organic solvents with ^{1}H hydrogen nuclei exchanged to ^{2}H (deuterium) nuclei. This is done to diminish the solvent signal in ^{1}H NMR, which would otherwise dominate easily. The solution is transferred to a NMR sample tube, a thin glass tube of usually 3 or 5 mm in thickness.

2. The sample tube is placed in a large and highly homogeneous magnetic field, usually created with a superconducting electromagnet. The field is further homogenized ("*shimmed*") by altering the currents flowing in a set of smaller surrounding coils. *Locking* electronics keep the field stable from outside disturbance and spontaneous drift by observing the abundant deuterium nuclei present in the solvent and counteracting any field changes.

3. The sample is surrounded by coil, which is connected to the transmitter electronics. One or few short radio frequency pulses (in the

order of microseconds) are passed to the coil, depending on the NMR experiment at hand. The exact sequence is controlled by an embedded computer (*pulse programmer*), which is executing a *pulse sequence* and controlling the involved amplifiers, gates, attenuators and frequency generators.

4. The pulse(s) cause a controlled disruption of the spin states, which begin to return to thermodynamic equilibrium. The oscillating magnetization is picked up by the same coil used for transmission.

5. The weak signal is amplified greatly and fed through a mixer, which generates two distinct signals with 90 degrees phase shift by using the original transmitter signal as a reference, a technique known as *quadrature detection* (Section 1.3.6). This helps further processing and makes it possible to detect the phase of the signal and thus both positive and negative frequencies (relative to the transmitter frequency).

6. The observed signal (FID) is digitized and stored in computer memory. The data is then processed using *Fourier Transform* (FT) to convert the time domain data into a frequency domain spectrum.

There are two principal challenges in building NMR spectrometers. First, the chemical shifts and couplings are in the order of kHz and Hz, small compared to the Larmor frequencies of several hundred MHz achieved with modern superconducting electromagnets. Measuring signals with a Larmor frequency of $\sim$500 MHz in 1 Hz precision is analogous to measuring a distance of 500 km with millimetre precision, so in order to resolve such small deviations, the utilized magnetic field must have extremely good homogeneity. At the same time, NMR magnets are required to be very strong to achieve useful polarization and signal dispersion.

Secondly, even in a strong magnetic field, the low polarization in thermal equilibrium results in weak signals, requiring a very careful design of electronics, signal paths and amplification. These aspects are discussed more thoroughly in Chapter 2.

1.3 NMR Acquisition Process

The details of the NMR acquisition process might not be very important in routine spectroscopy, but in order to dive in to the principles of sensitivity and quantitative NMR in the next chapters, further treatment of these aspects is essential. In the next few sections, some of the basic principles and details behind acquiring and processing spectra with modern NMR instruments are covered.

1.3.1 Tuning

Tuning involves matching the impedance of the probe coil, electronics and cable so that the RF energy is transmitted optimally to and from the sample. This traditionally involves adjusting two capacitors within the probe (*tune* and *match*, Figure 1.6) to minimize the reflected RF energy, shown in a display [16]. Good tuning minimizes signal losses in both directions improving both transmitter and receiver efficiency [17].

The tuning depends mostly on the dielectric properties of the sample, so for example large variations in solvent polarity or salt concentration severely affect tuning [6, 16, 18]. Temperature is also a factor, which can be a important when measuring temperature series. Optimizing tuning is not strictly required for each sample if it is within reasonable limits: slightly sub-optimal tuning will somewhat reduce sensitivity and increase the pulse width, but these drawbacks are usually acceptable. Consequently large sample sets containing similar samples in identical solvent can be usually measured without adjustment after the first sample, as the optimal tuning is roughly the same.

1.3.2 Shimming

Shimming is the process of optimizing the magnetic field homogeneity, and it is generally required for each sample to obtain high resolution spectrum. The magnetic field is adjusted by many small room temperature electromagnets, usually over 30 in modern instruments, called *shim coils* or just *shims* [19]. In addition to these, there exists a small number of superconducting electromagnets (*cryoshims*), which are usually adjusted only during installation of the magnet to perform coarse optimization. Each of the shims affect the magnetic field differently, ranging from coarse low-order shims to more complex patterns in high-order shims. By altering

the current flowing through each of the coils, complex adjustments to the field can be done.

Shimming is conventionally performed by optimizing the lock signal intensity: most NMR solvents are symmetrical molecules with one deuterium signal, and a more homogenic magnetic field will make the signal narrower and more intense, as all of the spins have almost uniform Larmor frequency regardless of spatial location. This method works quite well if the shims are initially close to optimal, and only small changes to the low-order shims are required. If the field requires complex adjustments, the changes in single shim does not necessary alter the lock signal much. Furthermore, high-order shims also interact with the low-order ones and local maxima are easily formed.

To obtain more detailed information, a test spectrum can be measured, and the shape of the signals can be examined to have more clues about which shims need adjusting. Still, the ultimate information needed would be the strength of the magnetic field in different locations of the sample. This information can be obtained with clever application of *Pulsed Field Gradient* (PFG) and RF pulses in PFG shimming (Section 4.4.3) or even with special probes which map the field inside the magnet with a mechanically moved small coil and sample during magnet installation.

1.3.3 Locking

Locking electronics compensate for external disturbances and the inherent instability and field drift of the magnet. This is usually done by continuously pulsing and observing the frequency of the strong deuterium signal from the solvent, and compensating the drift of the main field by the lowest order Z_0 shim. Some cheaper benchtop spectrometers lack lock electronics, and instead align the spectra using "software lock" [20].

1.3.4 Pulse calibration

In order to execute most NMR experiments, precise RF pulses must be carried out. The RF power dictates the strength of the RF magnetic field component, and subsequently how rapidly the net magnetization is rotated by the RF radiation. Thus in order to induce the desired rotations, the power of the RF radiation reaching the sample must be known precisely.

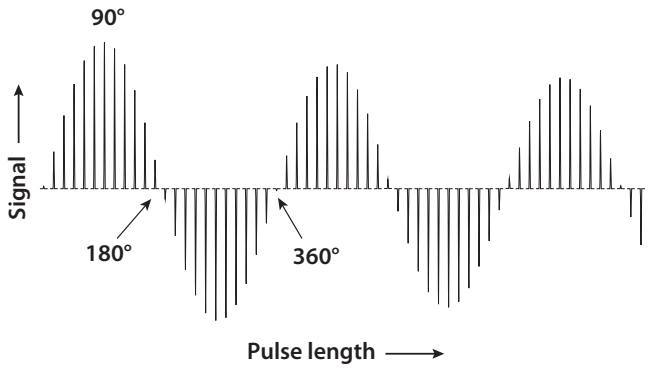

FIGURE 1.7: Traditional pulse calibration scheme. Single signal is observed with increasingly longer pulse lengths leading to sinusoidal signal intensity oscillation, from which correct pulse length can be determined. Note that the signal intensity decreases for longer pulses due dephasing caused by RF inhomogeneity, as various parts of the sample are affected by slightly differing RF power. Figure based on [21].

Specific rotation could be achieved with either altering the length or power of the RF pulse, but in most cases the power is kept constant and the length is altered, as timing is easier to adjust. The standard pulse length calibration consist of acquiring a sequence of 1D spectra with different pulse widths, and observing the intensity and phase of the observed signals. The applied RF pulse rotates the magnetization around the axis from which it is applied (determined by the phase of the signal), and increasing pulse length can be observed as a sine type of oscillation of the signal intensity (Figure 1.7).

Usually the length of a 90° pulse is cited as a measure of RF power. By determining the pulse length corresponding to maxima/minima or node points of the pulse calibration series, the length of a 90° pulse can be calculated. Commonly 360° or 180° pulse nodes are employed, as it is easier to determine the node point where the signal changes sign, as opposed to smooth maxima or minima, especially when signal-to-noise is a problem.

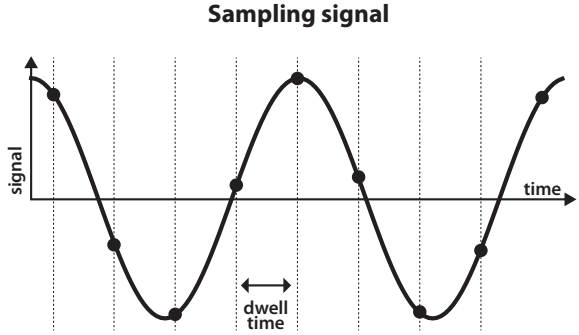

Figure 1.8: The electrical signal from the receiving coil is sampled in discrete time points and turned into numerical values.

1.3.5 Sampling the FID

The FID signal is sampled in discrete time points separated by a delay (*dwell time*), and digitized with *Analog-to-Digital Converter* (ADC) to numerical values, which can be stored in computer memory (Figure 1.8). The sampling frequency is governed by *Sampling Theorem* and *Nyquist frequency*, which dictate that the sampling rate must be at least twice the highest frequency desired to be stored correctly in the discrete signal [22, 23]. If the highest desired frequency is f_{max}, sampling rate must be at least:

$$f_{sampling} = 2\, f_{max} \tag{1.11}$$

and the dwell time (DW) or time between each sample is then:

$$DW = \frac{1}{2\, f_{max}} \tag{1.12}$$

For example 20 kHz signals can be detected with sampling frequency of 40 kHz, for which the dwell time is $1/40$ kHz $= 25\ \mu s$. The limit of human hearing is about 20 kHz, so it is unsurprising that the CD audio is stored with a sampling rate of 44.1 kHz [24].

The resonance frequencies encountered in modern NMR instruments are in the order of tens or hundreds of MHz, and sampling these kind of frequencies would require very high sampling rates reaching to GHz range, which would be quite impractical even with modern electronics. Chemists are interested mainly on the frequency differences between the resonances, which are much smaller, as chemical shifts are few ppm of the principal resonance frequency. Modern NMR spectrometers exploit this and subtract the transmitter frequency from the observed signal, producing a signal in which frequencies are relative to the transmitter frequency [1]. For example, from an initial 600 000 100 Hz signal produced by ^{1}H nuclei in a 600 MHz magnet, the subtraction of the reference 600 MHz transmitter signal results in a signal of only 100 Hz, which is much easier to sample and store. Modern spectrometers nevertheless use higher sampling rates than required by performing *oversampling*, discussed further in Section 2.3.2.

1.3.6 Quadrature detection

When subtracting the transmitter frequency from the original signal, two signals are actually produced, a *cosine* component and a *sine* component, generated by two signals 90° out of phase. This is called *quadrature detection*, and it is analogous to measuring the x and y component of the rotating magnetization vector [6]. The process can be carried out before ADC or fully digitally after sampling, which avoids the imperfections of analog electronics [25, 26].

The benefit of quadrature detection is to enable the detection of the *phase* of the signals, which can be used to separate positive and negative frequency offsets [1]. Without these two components, signals of 600 000 100 Hz and 599 999 900 Hz would be indistinguishable with a 600 MHz transmitter frequency, as both have a frequency difference of 100 Hz compared to the transmitter. This allows placing the transmitter frequency in the middle of the spectrum, which is beneficial as it reduces the frequency offsets and needed spectrum window, as discussed further in Section 3.1.2. As both positive and negative frequencies are distinguished with quadrature detection, the ensuing *spectrum width* or *window* in Hz is twice the Nyquist frequency and thus equal to the sampling frequency.

1.4 NMR Processing

After the FID is recorded and stored in the computer, one very important problem remains: as the FID contains a composite signal from all of the spins, how can individual resonance frequencies be separated? Or in other words, how can the *time domain* data containing a signal varying with time to be converted into a *frequency domain* spectrum? It turns out that this can be achieved by *Fourier transform* (FT), or when analysing signal sampled at discrete time points as in NMR, *Discrete Fourier Transform* (DFT) [24]. The DFT forms the core of NMR data processing, but it also includes all of the other optimizations and clean-ups performed on the digitized data.

1.4.1 The discrete Fourier transform

In Fourier transform methods, the signal is decomposed into sinusoid waves of different frequencies. In DFT, frequency components (Figure 1.9) are discovered by multiplying the sample data points by a set of sinusoid functions representing waves of different frequencies, and adding up the results for each wave. If the wave doesn't match the signal, the oscillating wave will yield both positive and negative values cancelling each other, with a total value close to zero. However, if the sinusoid matches the signal, the product will be positive even when both the signal and reference sinusoid are negative, and the sum will increase, indicating that the signal contains the sinusoid component in question.

This basic principle can be illustrated by a single wave, described by the function $f(x)$. If N sample values ranging from $x_0 \ldots x_{N-1}$ describe a signal, each value is multiplied by the function value at point n, and the results are added together:

$$X = \sum_{n=0}^{N-1} x_n f(n) \tag{1.13}$$

The resulting coefficient X represents the amount of the frequency component described by $f(x)$ in the samples, and repeating this process for different frequencies yields multiple coefficients, which can be arranged to form a spectrum. The actual DFT of N complex points ($x_0 \ldots x_{N-1}$) is defined as [22, 27]:

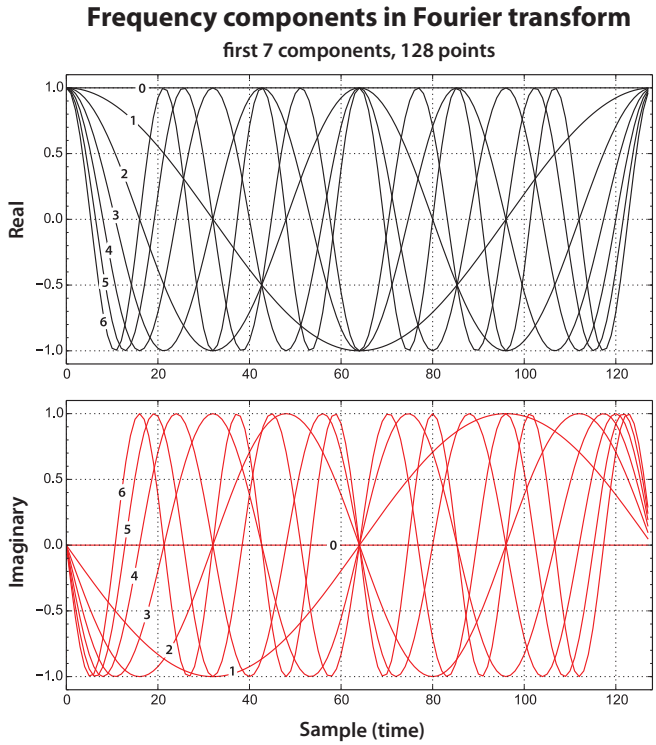

FIGURE 1.9: Plot of the first 7 frequency components or basis functions which are used to decompose the signal in discrete Fourier transform [22]. Note that the zeroth frequency component is a constant value function, and that the first point coefficient is exactly 1 for all real frequency components.

$$X_k = \sum_{n=0}^{N-1} x_n e^{-i2\pi k \frac{n}{N}} \quad k = 0, 1, \ldots, N-1 \tag{1.14}$$

where the values of k, ranging from $k = 0, 1, \ldots, N-1$, produce sinusoid functions of discrete frequencies. It is not immediately obvious that the

exponential term represents sine and cosine waves, but this is indeed true for complex valued exponentials, as described by Euler's formula [24]:

$$e^{ix} = \cos x + i \sin x \tag{1.15}$$

Using this relationship the exponential part can then be rewritten as:

$$e^{-i2\pi k \frac{n}{N}} = \cos(-2\pi k \frac{n}{N}) + i \sin(-2\pi k \frac{n}{N}) \tag{1.16}$$

It can be noted that the exponential produces complex numbers, and actually the sample values x_n and resulting coefficients X_k are also complex numbers in *complex* DFT. This is mathematically convenient and again related to the *phase* of the signal: the cosine and sine components of the signal are represented by the complex and imaginary parts of the complex numbers, so a sinusoid signal with any phase can be represented. In fact, the cosine and sine components detected in quadrature detection (Section 1.3.6) correspond to the real and imaginary parts of the sample values x_n in the complex DFT (Figure 1.9).

The related and somewhat simpler *real* DFT using real numbers can be used in NMR processing, and has been employed especially in *Time-Proportional Phase Increment* (TPPI) based 2D acquisition [16, 24]. However the basic principles in both methods are very similar.

1.4.2 FT-NMR and The Fast Fourier transform

Initially NMR spectra were acquired by *continuous wave* spectrometers in which the magnetic field (or frequency) was slowly varied to scan through the chemical shift range. Each resonance was excited and detected individually as the transmitted frequency matched the Larmor frequency, and the results could then be plotted to form a full NMR spectrum. These spectrometers were slow and insensitive, but robust and electronically simple.

The modern pulsed *Fourier Transform Nuclear Magnetic Resonance* (FT-NMR) spectroscopy was invented in 1966 by Weston A. Anderson and Richard Ernst [28, 29], with the enormous benefit that all resonances could be observed simultaneously instead of scanning the resonance frequencies one by one. This provided an immense improvement by reducing experiment

times by two orders of magnitude, and paved the way towards more complex multipulse NMR experiments and eventually 2D NMR.

Commercial FT spectrometers appeared in the early 1970s (with the first demonstration by Bruker Physik Co. in 1969 [29]), but one of the obstacles in their initial development was computing power: the basic calculation of DFT for n samples requires n^2 calculations as n frequencies are multiplied by n sample points each. In other words, a naive DFT algorithm has a time complexity of $\mathcal{O}(n^2)$, and so for a common number of 1024 data points (at the time), about 1 million operations are required. With the computer hardware available at the late 1960s and early 1970s, this meant a transformation time of about 15-20 minutes [30], barely adequate for practical use. Further, as the time grows quadratically $\mathcal{O}(n^2)$, doubling the resolution with 2048 points meant a computation time of over an hour, and 4096 points would take nearly five hours.

The reason why FT based operations are practical for use with large datasets is the invention of *Fast Fourier Transform* (FFT) algorithm, which takes advantage of the symmetry of the sinusoid signals and the ability to calculate a combined spectrum from separate transforms of interlaced data points. This allows the splitting of the problem into simple FT calculations which can be then combined sequentially in $\log n$ steps to yield the final spectrum, forming a *divide and conquer* type algorithm [24]. The most famous version was presented by Cooley and Tukey in 1965 [31], while the method was independently discovered by many individuals starting with Gauss already in 1805 [22]. The FFT algorithm reduces the time complexity of FT to $\mathcal{O}(n \log(n))$, resulting in spectacular time savings for transforms involving a large number of data points. For example the nearly five hour calculation of 4096 points mentioned above would be reduced to just under a minute with the help of the FFT algorithm.

The FFT is important in many scientific fields due its key role in signal processing, spectral analysis and digital filtering (convolution/correlation) [22], which are also paramount in modern NMR. Myriad of high-level tasks are fundamentally based on FFT and related transforms, such as many lossy compression schemes (JPG, MP3 and MPEG) [22, 24]. The FFT can be regarded one of the most important and influential modern algorithms, and it is implemented in practically all scientific calculation packages, including NumPy [32] and SciPy [33] packages utilized by SimpeleNMR presented in **Paper III** and **Paper IV** of this thesis.

1.4.3 Apodization

Before Fourier transform, the FID can be optimized with a few techniques to make the DFT to perform better. The most significant of these is apodization, where the FID data points are multiplied by a *window function* designed to emphasize different portions of the data. For example, as NMR signals are exponentially decaying sine/cosine waves, the tail end of the recorded data likely contains much more noise than the beginning, as little or no signal is present at the end of the FID. Using a window function which emphasizes the beginning and de-emphasizes the end of the FID can significantly reduce noise as demonstrated in Figure 1.10. Applying some kind of a window function is practically always done in modern NMR spectroscopy.

Window functions can be chosen according to many different objectives, such as emphasizing resolution (with the cost of increasing noise), or reducing certain artifacts (as with the sine bell functions used in many absolute value experiments to reduce dispersive signal components). Most commonly used functions are probably exponential/*Line Broadening* (LB) and Gaussian functions. These will reduce noise, but broaden the signals, as the signal in the end of the FID is diminished.

Another important reason to perform apodization is to remove possible boundary effects: if the signal doesn't decay entirely during the FID and zerofilling is used as usual (see next section), the abruptly ending signal will manifest as truncation artifacts after the DFT (Fig. 1.11). These can be very severe if short FIDs are acquired, or for example in the indirect dimension of 2D spectra, where the low number of acquired points makes using window functions essential.

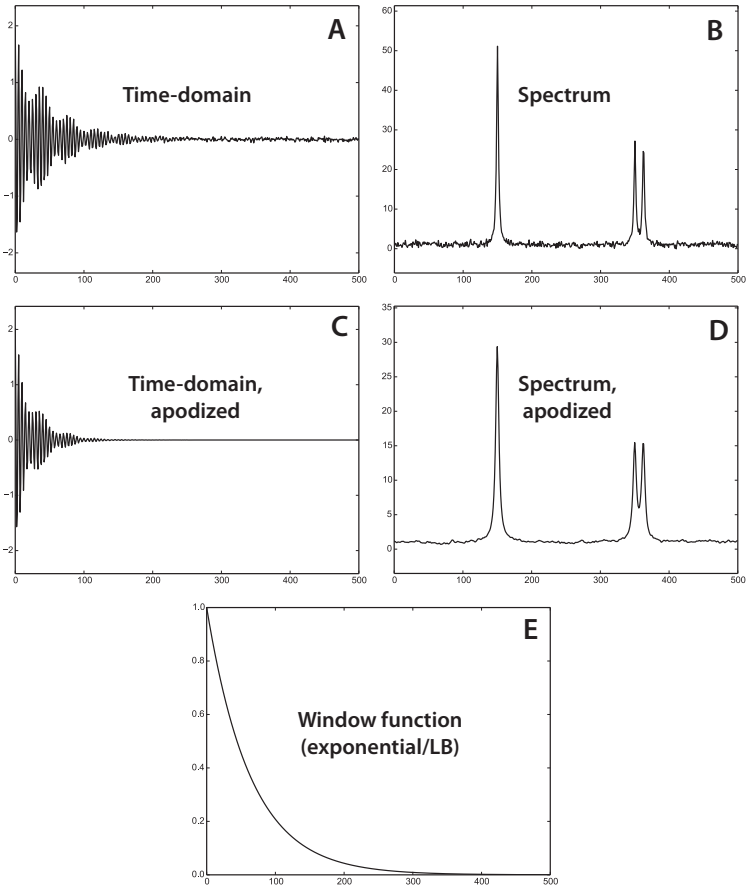

FIGURE 1.10: Apodization can be used to reduce the noise of the spectrum. FID without apodization (A) has clear high-frequency noise present in the baseline after Fourier transform (B). Multiplying the FID with window function reduces the noise from the end of the FID (C), resulting in a more clean baseline (D). The used window function of LB=1.0 Hz is shown on plot (E). Simulated data created and processed with SimpeleNMR internal testing tools.

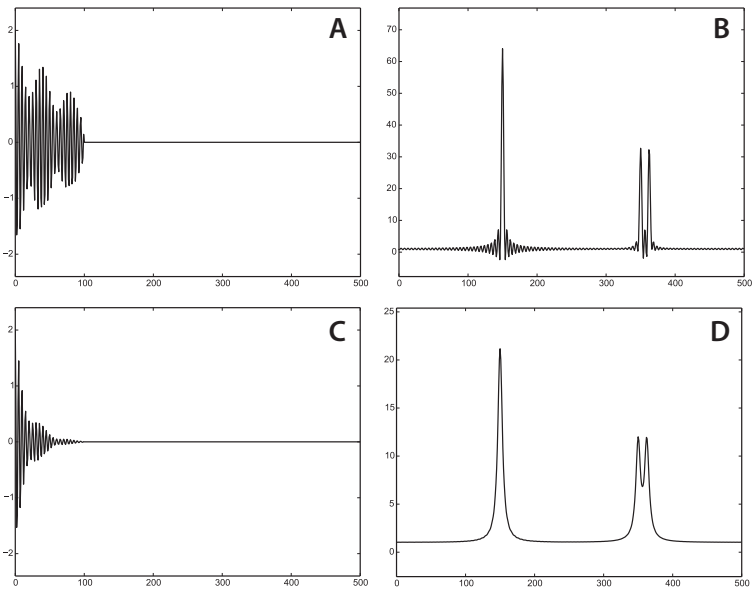

Figure 1.11: Truncation of the FID. Abruptly ending FID signal (A) will cause sinc -type wiggles in the transformed spectrum (B). Apodization of the FID (C, LB=2.5Hz) creates much better looking spectrum (D), while the signals are broadened. Simulated noise-free data created and processed with SimpeleNMR internal testing tools.

1.4.4 Zero filling and linear prediction

Fourier transform will yield exactly the same number of data points as is in the input data, which can lead to quite coarse looking signals if a low number of data points is collected. Artificially adding zeros to the end of FID yields more data points to the output, while not adding any noise (or of course, signal). This process is called *zero filling*, and can be used to produce smoother and more pleasant looking spectra. Zero filling to double the number of data points is usually considered optimal, as it improves the resolution and reduces noise by introducing the information contained in the imaginary component to the real one and vice versa, however apodization reduces this effect [34, 35].

Instead of adding zeros, adding more signal would be preferred, of course. This can be imitated by *Linear Prediction* (LP), where the continuation of signals is predicted with a linear model based on the previous data points [36, 37]. This has all the benefits of observing more signal, but it of course relies on the quality of the prediction, and has a potential to cause artefacts. LP is not very common in 1D experiments as it is easy to collect more data points in direct dimension, but in 2D and 3D spectra it becomes very beneficial as the collection of each data point takes a significant amount of time. Usually the number of data points is doubled with prediction, and as the resulting FID is apodized, the predicted part is usually de-emphasized quite heavily decreasing the chance of artifacts and using the actual observed data more efficiently.

1.4.5 Phase correction

As discussed earlier, modern NMR spectrometers use quadrature detection and both cosine and sine components of the signal are obtained. The resulting signals are then digitized and used as the real and imaginary parts of complex numbers in DFT.

In an optimal case, the real and imaginary components would match the cosine and sine components of the signal, which would then produce the desired pure absorption spectrum in real component and a dispersion type imaginary component after the DFT (Figure 1.12). Unfortunately, it is impossible to obtain this kind of data in practice due to experimental limitations in the acquisition, as discussed in detail later (Section 4.5.5). To remedy the situation, *phase correction* is performed on the data after DFT to produce pure absorption mode spectra.

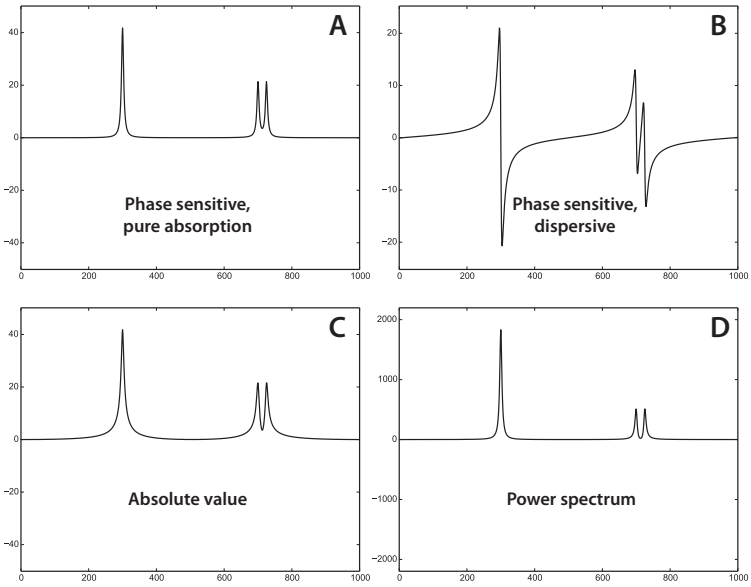

Figure 1.12: Simulated data demonstrating different handling of phasing of NMR spectra: pure absorption spectrum (A), pure dispersive spectrum (B), absolute value (C) and power spectrum (D).

1.4.5.1 Zeroth and first order phase correction

The simplest phase correction is zeroth order phase correction, which affects the phase of every signal identically. This can be thought as changing the *phase of the receiver*, and the effect is cycled for every 360° of correction. The calculation is quite simple, the new data point values are calculated from the real and imaginary components multiplied by the sin and cos functions [38]:

$$R_{i(new)} = R_i\ cos(\theta_0) - I_i\ sin(\theta_0)$$

where θ_0 is the phase correction angle, and R_i and I_i are the real and imaginary parts of i:th data point in the spectrum. To account for frequency dependent phase errors, *first order* or *linear* phase correction is used:

$$R_{i(new)} = R_i\ cos(\theta_0 + \theta_1\ \frac{i}{N_{points}}) - I_i\ sin(\theta_0 + \theta_1\ \frac{i}{N_{points}})$$

where θ_1 is the 1st order phase correction angle, i is the index of the data point and N_{points} is the total number of data points. The angle now depends on both angles θ_0 and θ_1, with θ_0 representing the constant and θ_1 the linearly frequency dependent phase correction. The correction can be formulated slightly differently, but the basic principle is the same. Zeroth and first order base corrections are enough to correct most NMR spectra which have been acquired correctly, while more complex phase errors can be introduced sometimes by for example complex off-resonance effects of pulses or non-linear phase response of filters. Phase correction is discussed further in Section 4.5.3.

1.4.5.2 Absolute value and power spectra

Phase correction can be avoided by calculating an absolute value (sometimes called also a *magnitude mode*) or a power spectrum, which discard phase information entirely. In absolute value mode, the magnitude of the rotating magnetization vector is calculated:

$$m_{av} = \sqrt{a^2 + b^2}$$

where m_{av} is the absolute value of a complex number $a+ib$. This formula can quite easily be derived from the Pythagorean theorem and the unit circle. Unfortunately, the dispersive component is also present on the resulting spectrum, which is caused by the fact that DFT treats the time-domain signal as *periodic* (and *infinite*) [24]. This means that the first and last point of the signal are actually adjacent in DFT, while there is clearly no physically meaningful relation between these data points. The large difference in the values of the data points produce frequency components, which are aliased into the spectrum. This is mostly avoided in phase-sensitive spectra by careful use of the phase correction to produce only absorptive component in the real part of the DFT result. However, in absolute value mode the phase is discarded altogether and phase correction is pointless.

The discontinuity in the time domain can easily be remedied by applying a suitable window function with a smooth roll-off to zero in both ends

of the signal (sine, sine squared, Blackman [6, 16, 24]), but this causes strong weighting of the signals by T_2 relaxation, as the start of the FID is strongly suppressed. On the other hand, the signals become much wider and more prone to overlap without windowing, and thus are very difficult to integrate fully (Fig. 1.12).

Power spectrum alleviates this problem somewhat, and is calculated similarly, by just squaring the real and imaginary parts:

$$m_{pow} = a^2 + b^2$$

The calculation yields much narrower and sharper signals, but the signal intensity relationships are not intuitive as they are squared as well (Figure 1.12). This makes spectra with large signal intensity differences quite hard to examine, and complicates the analysis of partially overlapping signals.

1.4.6 Baseline correction

In a good spectrum, in the parts where there is no signal present the data points should be close to zero, apart from some unavoidable Gaussian noise. These parts are called *the baseline*. Due to experimental errors and limitations, the baseline can have distortions, which hamper the interpretation and especially quantification of the signals. Baseline correction seeks to rectify these errors by modeling the bias in the baseline and subtracting it from the spectrum. The origin of the errors and baseline correction algorithms are discussed in detail later in Sections 4.5.4 and 4.5.5.

Signal, noise and sensitivity 2

> "Poot: How can he know where the stash at? The Narcos don't know, but he do, 'cause some nigga snitchin'!
> D'Angelo: Man, ain't nobody gotta be snitchin' for Omar, or one of his boys, to creep by and see where the stash at. "
>
> --The Wire

The sensitivity of an instrument describes the minimum level of input which produces signal with a specified power or *Signal-to-Noise Ratio* (SNR). In NMR context, sensitivity can be understood as the amount of material needed to produce a spectrum with adequate SNR in a reasonable amount of time. For exact definitions and limits, the properties of the sample, involved nuclei and NMR experiment must be described, but in practice, greater sensitivity means that the same experiments can be done with *less time or less sample material*.

This chapter deals with the origins of signal and noise in NMR, and the various strategies to improve the sensitivity of NMR experiments. The relatively low sensitivity of NMR as a technique makes these aspects especially important in quantitative NMR.

2.1 Sensitivity of NMR

The generally modest sensitivity of NMR is the result of the weak RF signal emitted from the sample, which follows from the low net magnetization and the small energy difference of the corresponding energy states inducing the polarization, as discussed in Chapter 1. This is a fundamental physical limitation, which can be only circumvented by not using thermal equilibrium as the source of the magnetization (which in special cases is indeed possible, as discussed in section 2.2.8). The main strategy of improving sensitivity is thus minimizing noise, and utilizing the available magnetization as efficiently and cleverly as possible.

The detection limit of NMR is greatly dependent on the desired information (experiments, observed nuclei), molecular size and available equipment (field strength, probe type, sample volume). With modern high-field magnets and cryoprobes, the characterization and quantification of small molecules is possible in microgram and micromolar level using ^{1}H NMR techniques [39, 40, 41, 42]. Benchtop spectrometers are more limited, but can still reach milligram and millimolar levels [10].

2.1.1 Signal-to-noise ratio (SNR)

Sensitivity is tied intimately to the SNR: the same gain in sensitivity can be achieved by doubling the signal or halving the noise, and both aspects have been significantly improved during the history of NMR. In electronics, the signal-to-noise ratio is simply defined as the *Root Mean Square* (RMS) voltage ratio squared between a signal (V_s) and the background noise (V_n) [43]:

$$SNR = \frac{V_s^2}{V_n^2} \tag{2.1}$$

As this ratio can have a very broad range of values, it is often given in decibels:

$$SNR_{db} = 10\ log_{10}\left(\frac{V_s^2}{V_n^2}\right) \tag{2.2}$$

In NMR, the amplitudes of electrical signals change rapidly with time rendering this basic definition less useful. Instead, an alternate definition

is used, in which signal intensity is related to the standard deviation of noise [4, 6]:

$$SNR = \frac{S}{2\sigma} \tag{2.3}$$

where S is the amplitude of the signal in a processed spectrum and σ is the standard deviation of the noise. This is quite intuitive, as the standard deviation describes how large deviations the background noise causes to the signal. As the noise can be assumed to be normally distributed, most often the noise affects the signal by a positive or negative change of σ or less, or a total range of 2σ, and this related to the signal intensity then gives the SNR. It also matches the visual comparison of the noise floor and the signal height, as the signal in the noise floor similarly varies mostly from $-\sigma$ to $+\sigma$, or within a range of 2σ. Indeed, the standard deviation can be estimated from the peak-to-peak noise in the baseline N_{ptp}, where N_{ptp} is estimated as 5σ [9], giving the commonly cited formula [4, 6]:

$$SNR = \frac{S}{2\sigma} \cong \frac{S}{2N_{ptp}/5} = 2.5\frac{S}{N_{ptp}} \tag{2.4}$$

Comparable SNR definitions are used in other fields where electrical signals are not directly considered, for example in image analysis [44].

2.1.2 Signal averaging and accumulation time

One easy way to increase signal is to add the results of two or more identical NMR experiments: if a single-pulse experiment results in a signal with the intensity I, adding together two FIDs from subsequent experiments result in a FID with a signal intensity of $2I$. This works because the signal in both cases is identical (small experimental errors aside), and will match exactly yielding a perfect constructive interference. The FID will also contain twice the noise, but as noise is random it adds up only partially, leading to an improved SNR.

This technique is routinely used in almost every NMR experiment, as in addition to boosting SNR, some types of experimental errors can also be canceled by altering the pulse and receiver phases (*phase cycling* [16]). For example, a four-step *CYCLically Ordered Phase Sequence phase cycle*

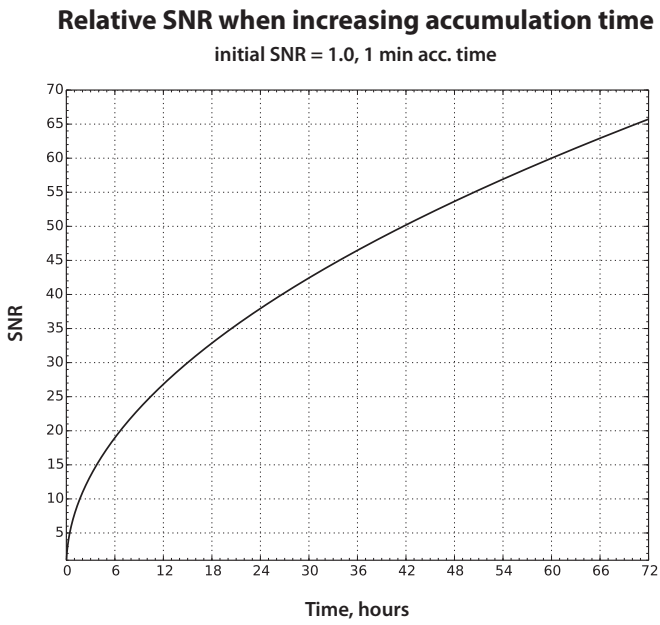

Figure 2.1: Calculated SNR results of a spectrum where 1 minute accumulation yields SNR of 1, which can be the case when obtaining a ^{13}C spectrum from a dilute sample. It is easy to see that in first hours the SNR climbs to well over 10, and "overnight" experiment of 12 hours yields a SNR of close to 30. Continuing the experiment for three days yields SNR of about $\sim$65, a significant improvement, but only about double what 12 hours could give. This scenario assumes that the repetition rate is kept constant.

(CYCLOPS) can be used to eliminate several imperfections regarding the receiver and quadrature detection [45, 46]. Each individual repetition of the experiment which is added or averaged together is commonly referred as a *transient*.

The exact improvement in the SNR can be calculated by finding out how the noise adds up. In most cases, the noise can be assumed to be random and normally distributed, and the sum of the noise can be then treated as a

sum of two independent and normally distributed random variables with identical means and standard deviations. For such variables, the sum is also normally distributed, and the properties of the new normal distribution can be calculated easily. If the standard deviation of the noise is σ and variance thus σ^2, the new variance is simply the sum of variances, from which the standard deviation is obtained by taking the square root [47]:

$$\sigma_{new} = \sqrt{\sigma^2 + \sigma^2} = \sqrt{2\sigma^2} = \sqrt{2}\sigma \tag{2.5}$$

Combining this with the SNR definition (Eq. 2.3), the new SNR resulting from adding two identical experiments together can be calculated:

$$SNR_{new} = \frac{S_{new}}{2\sigma_{new}} = \frac{2S}{2\sqrt{2}\sigma} = \sqrt{2}\frac{S}{2\sigma} = \sqrt{2}\ SNR \tag{2.6}$$

So, doubling the amount of transients will yield $\sqrt{2} \approx 1.414$ times the SNR compared to the original. Similarly, quadrupling the amount of transients will double the SNR:

$$SNR_{new} = \sqrt{4}\frac{S}{2\sigma} = 2\ SNR \tag{2.7}$$

Four repetitions of the experiment of course takes four times as long, and for every doubling of the SNR, the total measurement time is quadrupled. More exactly, the measurement time increases quadratically compared to SNR, and SNR is proportional to the square root of total accumulation time t:

$$SNR\ \propto\ \sqrt{t} \tag{2.8}$$

This is a very important practical notion; increasing the experiment time improves SNR less and less for each added unit of time, and acquiring more transients becomes a futile strategy quite rapidly (Figure 2.1). With this fundamental relationship of SNR and measurement time, sensitivity can be expressed as *SNR normalized to time* [4, 9]:

$$Sensitivity = \frac{S}{2\sigma\sqrt{t}} \tag{2.9}$$

2.2 Increasing signal

The most straightforward way to increase SNR is to boost the received electrical signal without introducing additional noise. The origin of a NMR signal is the precessing net magnetization, and for spin ½ nuclei, the net magnetization is proportional to the gyromagnetic ratio γ and the number of spins forming the magnetization [2, 4]. The number of contributing spins can be expressed as a product of the total amount of nuclei N and the polarization P, thus the net magnetization M is proportional to:

$$M \propto \gamma NP \tag{2.10}$$

Enlarging any of the above factors will increase signal, however, the relationship to SNR is not as simple in many cases. For example, the gyromagnetic ratio also impacts polarization and the Larmor frequency, which again are related to the induced voltage and noise. In the next sections, the relationship of the above and other factors affecting SNR are examined in detail.

2.2.1 Concentration

Based on Equation 2.10, increasing the number of spins contributing to the signal can be simply achieved by increasing the amount of material in the sample, provided that there is more material available and it can be dissolved. As increasing concentration doesn't increase noise, the number of observed nuclei/molecules (N) is also directly proportional to SNR:

$$SNR \propto N \tag{2.11}$$

Consequently, doubling the amount of material doubles the SNR, instead of a factor of $\sqrt{2}$ as for accumulation time, and *vice versa*. This is a very important practical relationship between accumulation time and concentration: doubling the concentration reduces the experiment time to one fourth of the original, while retaining the same SNR. Likewise, halving the concentration quadruples the measurement time, if the same SNR is desired. For this reason the sample concentration can be very crucial, and increasing the sample amount can have drastic effects on required experiment times.

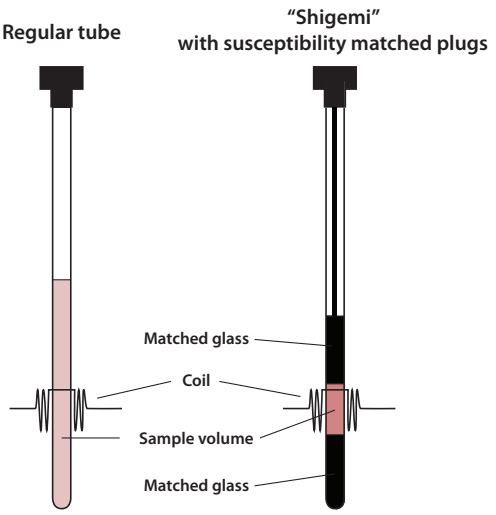

FIGURE 2.2: The geometry of a sample with susceptibility matched plugs compared to a regular one. Using the plugs all of the sample can be concentrated in the active volume of the coil, while boundary effects are minimized by matching the magnetic susceptibility of the solvent and glass.

There are no major immediate drawbacks of using a more concentrated sample, if the sample is stable and soluble in high concentrations. NMR is also a non-destructive method, and thus material can be in many cases recovered easily. Only when the sample material starts significantly affecting the consistency, viscosity and dielectric properties of the sample, downsides such as changes in relaxation times and chemical shifts or problems with tuning can appear. Even neat solutions can still be measured, provided that the instrument can handle the strong signal. Understandably a deuterium lock signal can't be established without any deuterated solvent in the sample, which can broaden signals in long acquisitions or even prevent high-resolution measurements depending on the magnet stability.

2.2.1.1 Susceptibility matched plugs or "Shigemi" tubes

When using standard NMR tubes and probes, a large part of the sample is not actually inside the active coil volume and is thus not contributing to the signal. Magnetic field inhomogeneities arise from magnetic susceptibility differences in the sample/glass/air boundaries, so the sample tube is filled with enough liquid to provide long symmetrical sample column in order to avoid these regions from contributing and broadening the signals [48]. For example with standard 5 mm NMR tubes and probe, the coil usually covers only about ∼10-20 mm of the sample height of perhaps ∼50-70 mm.

If the sample material is limited, the available amount can be exploited more completely by using less solvent to prepare the sample, thus increasing concentration. However, for above reasons shorter sample columns are harder to shim correctly and the inhomogeneity ultimately leads to a wider line shape and inferior resolution. A more elegant approach is to use susceptibility plugs (or *Shigemi* tubes, as originally made available by Shigemi Inc.[49, 50]): by filling the top and bottom parts of the tube with glass matching the magnetic susceptibility of the solvent, all of the available material can be concentrated to the active coil area. This increases signal by a factor of ∼2.5-3 compared to a "regular" sample in which the liquid column usually extends the length of the coil in both directions, with the added cost of a more expensive sample tube.

2.2.2 Filling factor and coil geometry

Unsurprisingly, the signal strength also depends on how large proportion of the coil volume is filled with sample liquid. This ratio of the sample volume to the active volume of the receiver coil is commonly referred as *filling factor* [40, 51].

The probes utilized in NMR are usually built with two saddle (Helmholtz) coils utilized for different nuclei, for example ^{1}H and ^{13}C . The inner coil can quite closely match the radius of the sample tube, while the outer coil needs to have a larger diameter. As a result, a smaller proportion of the coil volume is filled with sample for the outer coil, and the inner coil with a larger filling factor exhibits better sensitivity. Typically, the outer coil has only one half to one third of the sensitivity of the primary inner coil [16].

The sensitivity difference of the inner and outer coils is not accurately explained by the filling factor alone, as the coil diameter plays also a significant role. The diameter is important because it affects the relationship between the electric current and the magnetic field in the coil: when an identical current is applied through similar coils with unequal diameters, the coil with the smaller radius induces a stronger magnetic field inside itself. Conversely, the same oscillating magnetic field created by the net magnetization induces a larger current in the smaller coil [17, 51]. The relationship of induced current (I_{coil}) compared to the oscillating magnetic field (B_1) can be called the sensitivity of the coil, and this ratio is also proportional to the SNR [51]:

$$SNR \propto \frac{B_1}{I_{coil}} \tag{2.12}$$

The B_1/I_{coil} ratio depends on the exact geometry of the coil, but it is inversely proportional to the diameter of the coil in both saddle Helmholtz and solenoid coils [51]. Based on this the optimal SNR is achieved by building the smallest coil possible in which all sample material still fits and can be kept in solution [40]. For limited sample amounts, *microcoil probes* or *nanoprobes* can give a 5-10 fold improvement in sensitivity when compared to traditional 5 mm probes for the same sample mass [40]. However, it should be noted that mass sensitivity S_m and concentration sensitivity S_c are different:

$$S_m = \frac{SNR}{m} \qquad S_c = \frac{SNR}{c} \tag{2.13}$$

Microcoil probes are merely useful for boosting mass sensitivity, that is, increasing the signal for a limited amount of material. If only the concentration is limited (due to solubility for example) but sample material is available in abundance, the situation is reversed and the best SNR can be achieved by large sample volumes and coils. Nevertheless, microcoils can be used to reduce the sample mass requirements of readily soluble compounds to nanogram level [40].

2.2.3 Field strength

Another quite obvious way to increase the observed signal is to use a stronger magnetic field, as it increases the polarization of the spins. The relation-

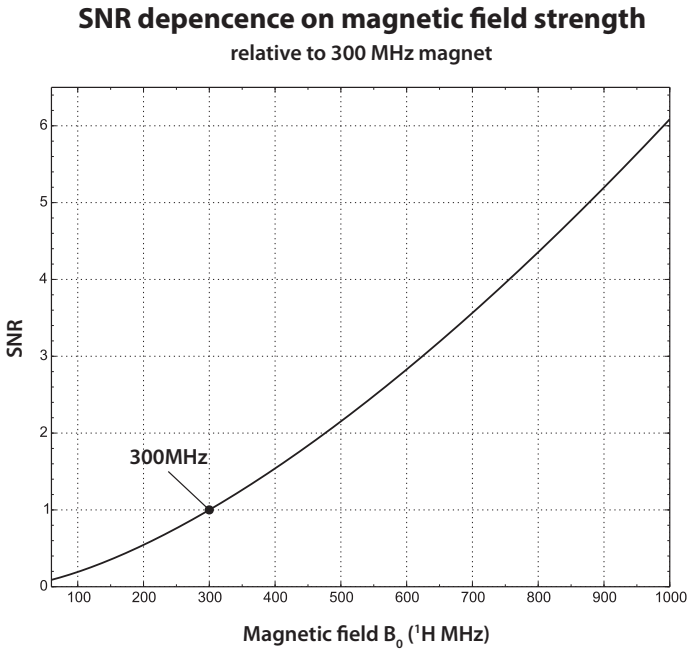

Figure 2.3: Signal-to-noise ratio grows more than linearly as a function of magnetic field strength B_0, making high field instruments almost always desirable. However, the price of NMR spectrometers tends to grow drastically when approaching the state of the art machines, and generally other methods of improving sensitivity are more feasible and cost-effective. At the time of writing, the strongest high-resolution magnets could deliver 1020 MHz, employing hybrid design with high and low temperature superconductors [52].

ship of polarization and the magnetic field can be found by substituting the spin populations in Eq. 1.6 with the Boltzmann distribution. With clever trigonometry, the exponential functions can be further simplified to a single $tanh()$ function [4]:

$$P = \frac{N_\alpha - N_\beta}{N_\alpha + N_\beta} = \frac{exp\left(\frac{\hbar\gamma B_0}{2k_bT}\right) - exp\left(-\frac{\hbar\gamma B_0}{2k_bT}\right)}{exp\left(\frac{\hbar\gamma B_0}{2k_bT}\right) + exp\left(-\frac{\hbar\gamma B_0}{2k_bT}\right)} = \tanh\left(\frac{\hbar\gamma B_0}{2k_bT}\right) \tag{2.14}$$

As the tanh() function behaves linearly near zero, and the spontaneous thermal polarization is very small in feasible magnetic fields, it can be considered proportional to the magnetic field B_0:

$$P \propto B_0 \tag{2.15}$$

This could be thought to indicate that the SNR is also directly proportional to the magnetic field similarly to concentration, but the relationship is more complex. In addition to the improved polarization, the higher Larmor frequency in a larger magnetic field induces greater voltage in the coil increasing the signal, but at the same time also the resistance is increased due the "skin effect", the tendency of high frequency alternating currents to flow in the skin of the conductor [4, 39]. The increased resistance increases thermal noise (Johnson noise [43]), and the combined result is that the SNR increases with the magnetic field slightly more than linearly [4, 6] (Figure 2.3):

$$SNR \propto B_0^{3/2} \tag{2.16}$$

For example, when using a 600 MHz spectrometer instead of a 300 MHz (with similarly capable electronics), one can expect an SNR increase of:

$$\left(\frac{600\,MHz}{300\,MHz}\right)^{3/2} = \left(\frac{14.16\,T}{7.08\,T}\right)^{3/2} \approx 2.83 \tag{2.17}$$

Note that as the B_0 field is proportional to the Larmor frequency, it can be substituted in the calculation.

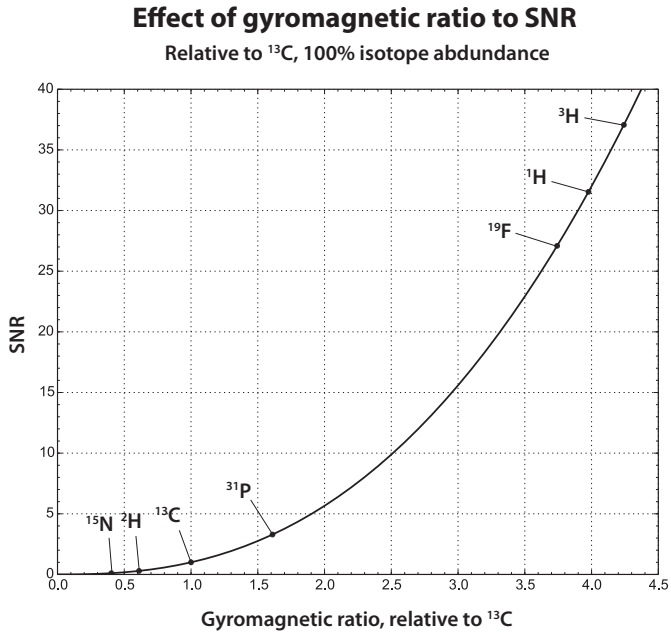

Figure 2.4: The gyromagnetic ratio has a powerful influence on the SNR, and from the graph it is clear why ^{1}H nuclei and ^{1}H detected experiments are preferred in NMR when maximizing sensitivity. The relationships shown do not take into account the natural abundance, so without isotopic labeling the sensitivity of ^{13}C is reduced to about 1.1%, while ^{1}H , ^{19}F and ^{31}P stay where they are. The absolutely best nuclei would be ^{3}H, but it is a radioactive isotope with practically zero natural abundance. Values are based on reference [53].

2.2.4 Gyromagnetic ratio

The gyromagnetic ratio has even greater influence on the SNR (Figure 2.4), because in addition to the Larmor frequency and polarization (as with greater magnetic field), the magnetic moment of the spins is also affected. Again, the noise also depends on the Larmor frequency, and the net result is [4]:

$$SNR \propto \gamma^{5/2} \tag{2.18}$$

The dependency can be split into two terms when the excited and observed nuclei have different gyromagnetic ratios: the gyromagnetic ratio of the excited nuclei (γ_e) influences the polarization, while the gyromagnetic ratio of the detected nuclei (γ_d) determines the Larmor frequency and the magnetic moment, yielding [6, 16]:

$$SNR \propto \gamma_e \gamma_d^{3/2} \tag{2.19}$$

This separation is useful when magnetization is transferred between nuclei types, as discussed in the next section. Unfortunately the gyromagnetic ratio itself is an intrinsic property of the nuclei, so the only way to influence it is changing the type of the observed nuclei, and this choice is obviously restricted to what is available in the sample material. Even when the use of some exotic nuclei could be beneficial, the equipment at hand might not be able to handle the appropriate nuclei or observing the nuclei might not be able to produce the desired structural information.

2.2.5 Magnetization transfer

While the gyromagnetic ratio is fixed, by clever manipulation of the spins, magnetization can be transferred between nuclei, accomplishing some of the benefits of higher gyromagnetic nuclei even when the properties of the low-γ nuclei are of interest. This combined with the high gyromagnetic ratio of ^{1}H has lead to the prevalence of *inverse detected* multidimensional experiments, where relevant information regarding other types of nuclei is encoded in indirect dimensions and ultimately detected as modulations of the ^{1}H signal. Magnetization can be transferred by several mechanisms, but two of most important ones are NOE and polarization transfer through J-couplings, discussed in the next two sections.

2.2.6 The NOE mechanism

NOE was already briefly introduced in the first chapter, and as discussed, it is based on dipolar through-space interactions between nuclei, essentially the magnetic moments interacting directly with each other. The effect was first reported on metals by Albert Overhauser already in 1953 [54], but its

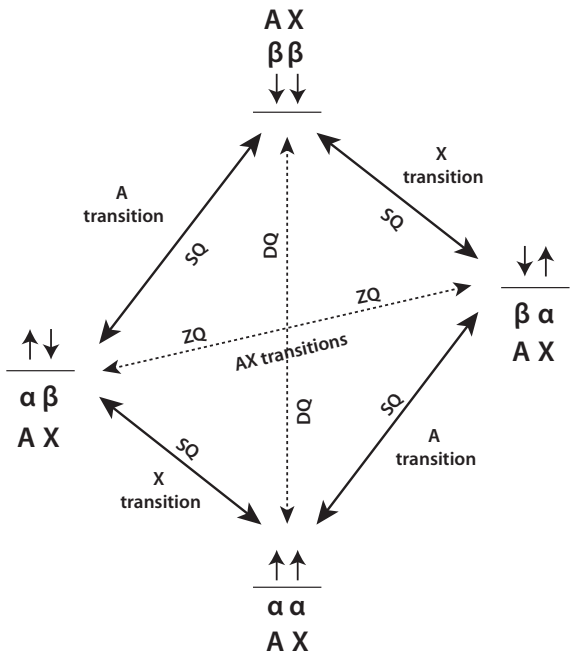

Figure 2.5: The spin transitions involved in the NOE mechanism [1, 6]. In addition to the single quantum transitions (marked as SQ), the double quantum and zero quantum transitions (DQ and ZQ) involving two spins exist for spins in close proximity. These transitions involve the flipping of two spins, and provide a mechanism in which the transitions involving A spins can affect X spins and eventually significantly alter the spin populations.

significance and application in liquid NMR was devised much later [55]. It has since become a powerful tool to investigate stereochemistry, conformations and non-covalent interactions of molecules (especially protein and peptide NMR) [15].

To illustrate the mechanism, two spins A and X can be considered, where A is the higher and X the lower gyromagnetic nucleus. When the A spin population is disturbed (by continuous irradiation of A nuclei, for example),

the A spins start to return to the thermal equilibrium through relaxation (discussed in more detail in section 3.1.1). In order to re-establish the initial state, some of the A spins must flip from the higher-energy state (β) to the lower one (α). This can happen with transitions where only the A spins flip (A transitions, Figure 2.5), but if the X spins are in close proximity, the magnetic moments of the spins can interact and there are available transitions where both spins flip at the same time (AX transitions).

The A or X transitions are *Single Quantum* (SQ) transitions, and affect only the spin populations of A or X nuclei. The AX *Double Quantum* (DQ) and *Zero Quantum* (ZQ) transitions however involve flipping of both spins, and relaxation using these pathways will also affect the populations of the other spin. For example, an AX double quantum transition from β to α includes also X spin transition from β to α state, so A spins relaxing through this transition boosts the α state population of X spins in the process. This can be observed as additional X magnetization, and is subsequently called *positive NOE*. Similarly the relaxation of A spins through the zero quantum pathway will involve flipping X spins to β state, observed as a reduction of magnetization (or making the signal even negative), correspondingly referred as *negative NOE*.

In practice, all of the transitions are taking place, but the probabilities of different transitions can vary drastically. The spin transition is triggered by the fluctuating magnetic fields of the surroundings, which are influenced by the speed of the surrounding molecules (*rotational correlation time* τ_c [1]). For example, the movement of small molecules in non-viscous solutions is fast, and the double quantum transitions (with a frequency of the sum of the Larmor frequencies of the involved nuclei) match better the fluctuating magnetic fields than the single quantum ones. The result is that small molecules tend to favor double quantum transitions and produce positive NOE, while large molecules produce negative NOE, and with a suitable intermediate size neither mechanism dominates, yielding negligible NOE.

NOE can be observed in 1D experiments by selective pulses or continuous irradiation, but most commonly 2D *Nuclear Overhauser Effect SpectroscopY* (NOESY) experiment [15, 55, 56] is used. One important and practical effect of NOE is observed when decoupling nuclei. For example, decoupling ^{1}H nuclei during regular 1D carbon acquisition produces NOE transfer from the relaxing ^{1}H nuclei to ^{13}C, yielding a signal enhancement dependent on the number of nearby ^{1}H. This is usually prefer-

able, but must be avoided when acquiring quantitative spectra (Section 3.1.3). The maximum enhancement achievable by NOE is [1, 6, 57]:

$$I_{NOE} = I_{initial}\left(1 + \frac{1}{2}\frac{\gamma_A}{\gamma_X}\right) \tag{2.20}$$

Where γ_A is the higher and γ_X the lower gyromagnetic ratio. With opposing signed γ, NOE becomes negative, for example when observing ^{15}N with ^{1}H decoupling (the γ_{15N} is negative). In these cases the effect reduces the signal or can overpower it completely.

2.2.7 Polarization transfer through J-coupling

The J-coupling between two nuclei can also be utilized to transfer magnetization. This form of polarization transfer is perhaps one of the most important techniques employed in modern liquid NMR, and it is used extensively in multidimensional experiments. The principal technique for achieving this is the *Insensitive Nuclei Enhanced by Polarization Transfer* (INEPT) method, originally introduced by Morris and Freeman in 1979 [58]. In the INEPT pulse sequence, J-coupling is used to perform a *selective inversion* of populations in the energy states associated with the coupling on one spin (Figure 2.6). This affects also the populations of the corresponding transitions on the coupled spin, and can then be observed as enhanced but antiphase magnetization (Figure 2.7).

The same selective inversion could be performed with other means such as a selective 180° pulse, but the power of INEPT lies in its ability to create this inversion in all of the coupled spins simultaneously. Because evolution under the J-coupling is used to create the desired antiphase magnetization, the length of delay Δ is chosen to be such that all magnetization has evolved to a antiphase state. This is true when the delay Δ length is exactly half of the period determined the by the J-coupling constant frequency:

$$\Delta = \frac{1}{2J_{IS}} \tag{2.21}$$

Of course, the J-couplings depend on chemical structure, and in practice the value is just chosen to match the relevant couplings as closely as possible. The inevitable mismatch of Δ delay in practice results in a suboptimal conversion to antiphase magnetization, and subsequently to a somewhat reduced polarization transfer.

INEPT sequence was initially used as an enhancement of low gyromagnetic nuclei, such as ^{13}C and ^{15}N using attached ^{1}H nuclei. In order to create in-phase magnetization for nicer looking spectra and enable the usage of decoupling, an additional delay was quickly introduced after the sequence to refocus the antiphase magnetization [59, 60]. Also a 2D experiment combining two INEPT blocks to transfer the magnetization from ^{1}H to ^{15}N and back was introduced very quickly by Bodenhausen and Ruben [61]. This experiment, known as the *Heteronuclear Single Quantum Coherence* (HSQC) is one of the principal heteronuclear experiments,

and illustrates the idea of *indirect detection* to provide maximum enhancement: the magnetization is transformed to ^{13}C or ^{15}N to take advantage of the larger initial thermal equilibrium polarization of ^{1}H , but in addition it is again transferred back to ^{1}H to take advantage of the higher Larmor frequency and magnetic moment during detection.

When transferring magnetization to nuclei with lower gyromagnetic nuclei, in perfect transfer, the polarization of the higher gyromagnetic ratio is transferred fully, leading to an intensity boost of [6, 57]:

$$I_{INEPT} = I_{Initial}\frac{\gamma_I}{\gamma_S} \tag{2.22}$$

The increased signal also boosts SNR in the same proportion, as shown earlier (Eq. 2.19). For example, optimal INEPT transfer from ^{1}H to ^{13}C boosts the SNR by a factor of:

$$\frac{\gamma_{1H}}{\gamma_{13C}} \approx 4 \tag{2.23}$$

However in HSQC, the magnetization is also transferred back to ^{1}H for observation, so using Equation 2.19 and substituting the above ratio yields:

$$\frac{SNR_{HSQC}}{SNR_{13C}} = \frac{\gamma_{1H}\,{\gamma_{1H}}^{3/2}}{\gamma_{13C}\,{\gamma_{13C}}^{3/2}} \approx \frac{4 * 4^{3/2}}{1 * 1^{3/2}} = 32 \tag{2.24}$$

Using polarization transfer, a HSQC experiment achieves up to $\sim$32 times the SNR compared to a regular 1D ^{13}C experiment without NOE from ^{1}H decoupling, a truly spectacular achievement. The polarization transfer and HSQC experiments are discussed in more detail in Section 3.4.

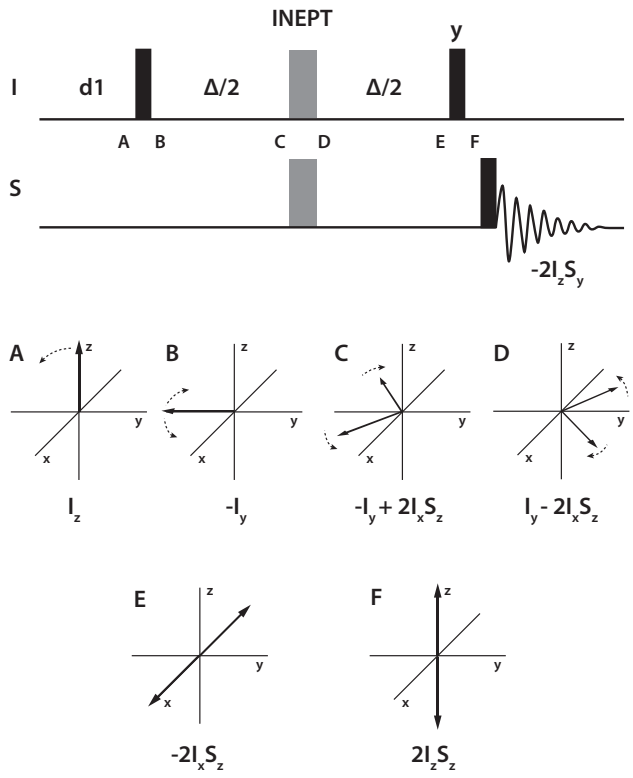

FIGURE 2.6: The basic INEPT sequence, and related vector and product operator presentations [57] of the magnetization for two spins I (high-γ) and S (low-γ), corresponding to for example CH carbon. The J-coupling is used to evolve the $-I_y$ magnetization to antiphase magnetization $-2I_xS_z$ (stages B-E). The delay Δ is chosen depending on the coupling constant J_{IS} so that the magnetization is exactly antiphase at stage E (Equation 2.21). This can be then easily rotated to the z-axis, corresponding to a state where one of the coupling transitions is selectively inverted (Figure 2.7).

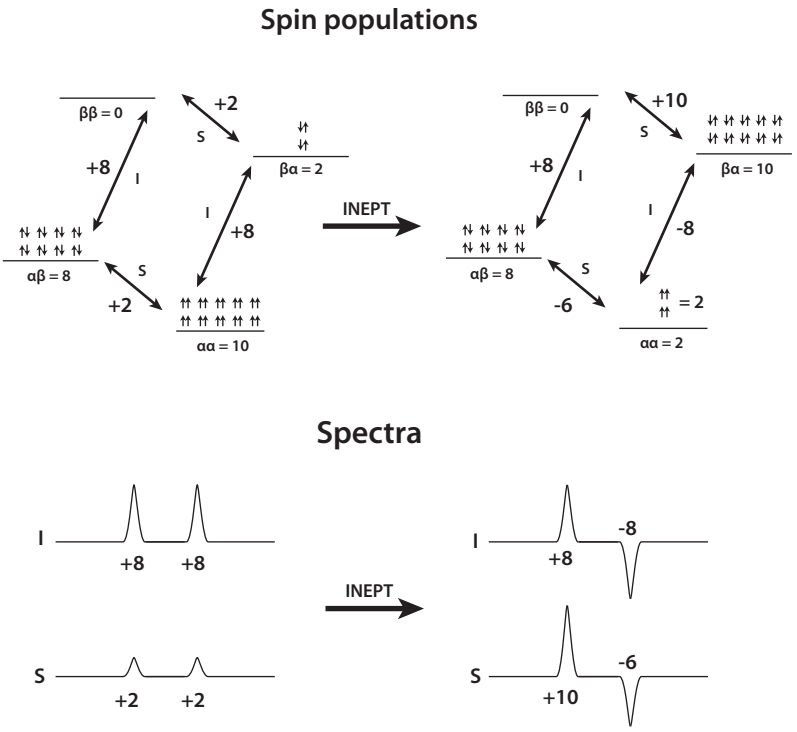

Figure 2.7: An energy level diagram corresponding to the INEPT experiment in Figure 2.6 [57]. After an INEPT pulse sequence, the selective inversion of the I transition changes the population differences between S transitions, which can be observed as enhanced antiphase magnetization instead of a regular doublet in the low-γ S nuclei. The larger population difference (8 vs 2) of the I nuclei is transferred to the S nuclei.

2.2.8 Hyperpolarization

The thermal equilibrium polarization of spins governed by the Boltzmann distribution is very small, as discussed in chapter 1.1, and thus any technique which can produce more polarization by an alternate mechanism can potentially yield an enormous improvement to the sensitivity of NMR spectroscopy. In a 1.5 T (∼64 MHz) magnetic field, the ∼5 ppm polarization of ^{1}H spins in room temperature implies a potential of 200 000 fold increase in magnetization, if 100 % polarization can be achieved instead [62]. The methods of achieving these potentially huge improvements are collectively called *hyperpolarization* techniques.

While hyperpolarization can improve the sensitivity by multiple orders of magnitude, all current methods have many requirements limiting the scope in which they can be applied, and as such have not replaced conventional FT-NMR spectroscopy. When applicable, the drastic improvements in sensitivity can enable experiments which can sound completely unrealistic in the context of regular NMR spectroscopy, such as using single-scan ^{13}C spectra to follow and quantify enzymatic reactions or even the metabolism of live yeast cells [63, 64].

In the following sections, only *Dynamic Nuclear Polarization* (DNP) and *Parahydrogen-induced polarization* (PHIP) based techniques are discussed. There are more methods for hyperpolarization, such as the optical pumping utilized to polarize ^{129}Xe especially in *Magnetic Resonance Imaging* (MRI) applications [65], but the DNP and PHIP methods are perhaps the most applicable in chemistry.

2.2.8.1 Dissolution and Overhauser DNP

Dissolution Dynamic Nuclear Polarization (D-DNP) [62] is probably the most widely utilized hyperpolarization technique, with commercial setups being produced (Oxford Instruments HyperSense). In D-DNP, *electron* polarization from an unpaired electron is transferred to NMR active nuclei in solid state. This is achieved by freezing the sample and keeping it in a very low temperature (∼1-2 K) in the presence of external magnetic field, while irradiating it with microwaves. The microwave radiation matches the Larmor frequency of the unpaired electron of stable radical compounds such as *2,2,6,6-Tetramethylpiperidine 1-oxyl* (TEMPO) [66, 67], which subsequently polarizes NMR active nuclei through the DNP mechanisms. The sample is then rapidly thawed by adding hot sol-

vent and moved quickly (under 6 seconds [62]) to a larger NMR magnet for regular observation, as T_1 relaxation starts immediately.

With this technique polarizations up to 40 % (^{13}C) can be obtained, corresponding to signal enhancements of over >10 000 times the regular equilibrium polarization, a truly spectacular result [62]. However, besides the obvious drawbacks related to the quick temperature changes, there are other severe limitations:

- T_1 relaxation starts after the removal of the microwave radiation, and both the low magnetic field during transfer and the radical agent speed the relaxation up, requiring rapid transfer in the order of seconds to the NMR spectrometer before losing the polarization [68].
- The build-up of the polarization takes a significant amount of time (tens of minutes to several hours).
- The slow and complex process means that multiple polarization steps are not feasible, rendering all traditional experiments requiring multiple samplings of the magnetization difficult, such as regular 2D experiments.

Many improvements have been recently proposed to mitigate these limitations: more rapid transfer times of the sample from polarization to measurement can be achieved by creating two fields of suitable strengths to a single bore to minimize the physical distance of the DNP polarization and NMR stages [69], and/or using ascorbate to scavenge the radicals to reduce the rate of relaxation during the transfer [70]. The build-up time can be reduced by using high gyromagnetic ratio nuclei such as ^1H and then transferring the magnetization to a lower one through cross-polarization [71]. Two-dimensional spectra can be acquired by combining ultrafast 2D techniques [72] with spatial encoding to acquire a 2D spectrum even in a single scan [73, 74]. The related technique of *Overhauser* DNP can be performed in liquid samples, but delivers much smaller improvements ($\sim$ 20x) [75]. Following on these rapid developments, it is probably safe to say that the popularity and applicability of DNP will surely continue to increase.

2.2.8.2 Parahydrogen-induced polarization

A molecular hydrogen with ^{1}H nuclei has two spin isomers, in which the spins are either aligned (*ortho*, triplet state) or opposing (*para*, singlet state). In room temperature, the ortho form dominates with a ratio of 3:1, but in low temperatures the lower energy causes the para form to dominate (99.82 % at 20 K) [76]. The transition between the states is quantum mechanically forbidden, but a paramagnetic catalyst circumvents this selection rule. By cooling hydrogen to a low temperature (e.g. with liquid nitrogen, 77 K) with the catalyst, enriched *para*-hydrogen can be formed, which is quite stable even at room temperature with the catalyst removed [77].

When *para*-hydrogen is used in a hydrogenation reaction, the spin correlation is carried to the target molecule, producing greatly amplified signals in NMR spectra. The large population of the *para* $\alpha\beta$ and $\beta\alpha$ states compared to ones in thermodynamic equilibrium yields transitions to $\alpha\alpha$ and $\beta\beta$ states, which can be observed as characteristic antiphase doublets. This technique was introduced in 1987 by two separate groups as *Parahydrogen And Synthesis Allow Dramatically Enhanced Nuclear Alignment* (PASADENA) [78] and PHIP [79], shortly followed by a version where hydrogenation is done outside of the magnet [80].

Once the *para*-hydrogen is introduced to a molecule, the polarization (or perhaps more accurately spin correlation, as PASADENA signals are antiphase) can be utilized in various ways. The magnetization can be transferred to bonded heteronuclei via similar polarization transfer experiments as in regular NMR, and many types of experiments are possible [77, 81, 82].

While PHIP methods can be very effective and approach theoretical limits of $\sim$100 % polarization [83], the application is obviously hindered by the requirement of covalent modification of the target molecules by hydrogenation. An interesting and quite recent solution to this limitation is the *Signal Amplification By Reversible Exchange* (SABRE) method, in which the hydrogenation requirement is circumvented by reversible coordination of the target molecule and *para*-hydrogen with a transition metal center, in which the polarization is transferred [82]. While the achieved enhancement is not as good as with some PHIP techniques, still 300-800 fold increases in signals have been achieved.

2.3 Reducing noise

SNR can also be improved by reducing the amount of noise. The most straightforward way to do this is to increase the quality of the electronics: using low-noise amplifiers and components, as well as properly shielded cables and connections in conjunction with good electronic and probe design all contribute to the minimization of the electronic noise. This aspect of NMR spectroscopy is by large in the realm of spectrometer vendors, and apart from using the equipment properly, often the only thing one can do to improve the electronics is to buy newer and/or better equipment. Further, through incremental improvement during many years, the electronic design of modern spectrometers is already very good and mostly limited by noise arising from the thermal movement of electrons. As such, major improvements in the basic electronics seem quite unlikely.

2.3.1 Cryoprobes

The basic relationship of temperature to RMS voltage of thermal noise is [85]:

$$V_{noise} = \sqrt{4\,k_B\,T\,R\,B} \tag{2.25}$$

where k_B is the Boltzmann constant, T is temperature, R is resistance and B is the receiver bandwidth in Hz. Clearly, lowering the temperature can decrease thermal noise very significantly, but reducing the sample temperature is not very viable in liquid state NMR. The relationship of temperature to SNR with more accurate consideration to temperatures of different parts of NMR equipment was established by Hoult and Richards in 1976 [39, 51, 86]:

$$SNR \propto \frac{1}{\sqrt{4\,k_B\,B\,[R_c(T_c+T_a)+R_s(T_s+T_a)]}} \tag{2.26}$$

where R_c and R_s are the resistance of the coil and sample while T_c, T_s and T_a are the temperatures of the coil, sample and preamplifier respectively. These results suggests that by lowering only the temperature of the principal receiving electronics, the coil and preamplifier, noise can still be reduced and SNR can be significantly improved. In addition to directly

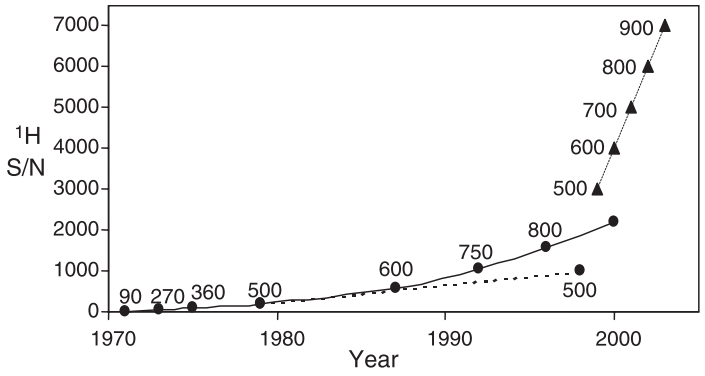

FIGURE 2.8: The improvement of signal-to-nose ratio over time due to improved instrument design, as presented by Helena Kovacs et al. [39]. The given signal-to-noise ratio is of 0.1% ethylbenzene (EB) in $CDCl_3$ for ^{1}H-observe coil configuration. The black dots mark conventional probes at the launch of a magnet operating at a particular field, while the triangles mark the launches of cryogenic probes (all data from Bruker BioSpin). The immense improvement provided by cryogenically cooled probes is easy to see. For comparison, the SNR achievable with current benchtop spectrometers is approximately 10, several orders of magnitude less than the state-of-the-art machines [84]. The dashed line indicates the increase in sensitivity during two decades for a conventional probe operating at 500 MHz. Reprinted with permission from Progress in Nuclear Magnetic Resonance Spectroscopy. Copyright 2005 Elsevier.

decreasing T_c and T_a, the resistance terms R_c and R_s are reduced in lower temperatures cutting down noise even further.

Probes taking advantage of this principle are called *cryoprobes* (or *cold probes*), and the concept was first demonstrated in practice with a custom-built probe in 1984 by Styles and Soffe [85]. This probe employed liquid helium to cool the coil and preamplifier, and achieved an SNR improvement by a factor of $\sim$8 compared to a room temperature probe while acquiring a ^{13}C spectrum at 45.9 MHz. Due to engineering challenges, commercial probes appeared much later in 1999, utilizing cooled helium or nitrogen gas circulated in a closed loop instead of liquids. The cooling reduces noise

typically by a factor of ∼4 for probes utilizing helium, producing a corresponding improvement in SNR [39].

Cryoprobes provide a drastic reduction in measurement time, and introduced a jump in sensitivity comparable to what 20 years of incremental improvements in electronics and magnet design achieved: a 500 MHz instrument equipped with a cryoprobe is comparable to state of the art ∼1 GHz machines using conventional room temperature electronics (Figure 2.8) [39]. The drawback is of course the much more complicated design and subsequently higher cost, however the offered advantage is so significant that cryoprobes are practically standard in high-field and high-sensitivity applications such as biomolecular NMR. Considering the current importance of the technique it is interesting to note that the initial cryoprobes were developed outside the major NMR instrument vendors, and even after quite usable equipment was demonstrated, there were initially little interest towards the technique from the manufacturers [87].

2.3.2 Oversampling, quantization noise and digital filters

One fundamental limit for noise in digital signal processing is *quantization noise*. This noise is the unavoidable result of converting the analog electrical signal to discrete values (Figure 2.9), which is necessary to perform the DFT and other processing in computer software [23]. To minimize quan-

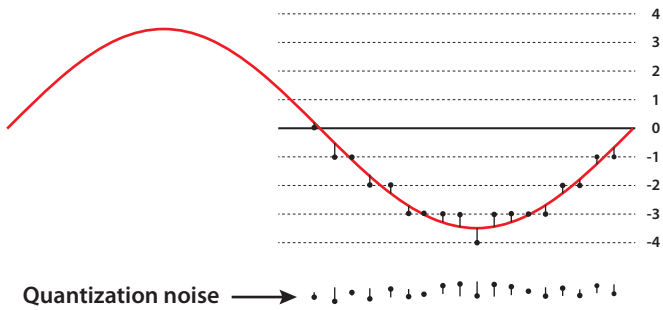

Figure 2.9: The conversion of a continuous sine signal to discrete values creates quantization errors (black lines), which are observed as noise. The effect has been emphasized here by using very coarse stepping.

tization noise, one obvious solution is to increase the digitizer resolution by using more precise numbers to describe the signal levels, with every bit doubling the possible values. This is not cost-effective after a certain point (modern NMR spectrometers use around 14-16 bit ADC boards [88, 89, 26]), and instead the dynamic range is improved by *oversampling* [23, 90].

As discussed earlier in Chapter 1, Nyquist criterion dictates that the sampling rate must be at least twice the highest frequency of interest, and faster sampling rates are considered *oversampling* [23, 24]. The amount of oversampling can be described by the oversampling factor, the factor by which the sampling rate is increased, or the relationship between the resulting spectral widths (SW). The gain in bits can be estimated by [23]:

$$gain = log_2 \frac{SW_{oversampling}}{SW} \tag{2.27}$$

For example, 16-fold oversampling increases the resolution by 4 bits. There are two main benefits to using oversampling:

1. **Reduction in quantization noise and improving the dynamic range of the data**: Quantization noise will be distributed uniformly to the spectral width, so the noise falling to the area of interest is reduced. This is directly proportional to the oversampling factor, and can be intuitively thought to result from more samples measuring the signal more precisely. The end result is improvement in noise level and dynamic range as described above [23, 90].
2. **Easier analog filtering and improved baseline**: The low-pass antialiasing filters before ADC can be opened up to have a much higher cut-off frequency, yielding a negligible attenuation in the frequencies of interest and reducing the errors introduced by the analog filters [23]. This reduces the aliasing of signals and shortens the transient response of the analog filter, yielding a much improved baseline [23, 91, 92]. The start of sampling and related aspects are discussed further in Section 4.5.5 as they are related to phase and baseline errors.

The smaller spectral area of interest can be extracted from the oversampled data by utilizing digital low-pass filters and decimation [23, 24]. Digital

filters are based on numerical calculations on the digitized data, and can easily have linear phase response and other desired features. A simple example of a digital filter is a calculation of moving average, which can also be considered a type of low-pass digital filter. The quantization noise can be overshadowed by other sources of noise [93], but the other valuable properties of oversampling and digital filtering still remain.

2.4 Combined SNR equation

All of the factors represented in this chapter can be combined to yield a more universal formula, and the following formula or some close variation is usually presented in NMR textbooks, containing the most relevant ones for practical use [6, 16]:

$$SNR \propto N \gamma_e \gamma_d^{3/2} B_0^{3/2} \sqrt{N_S} \frac{T_2}{T} \tag{2.28}$$

In the formula N is the amount of contributing nuclei/molecules, γ_e and γ_d are the gyromagnetic ratios of excited and detected nuclei, B_0 is the external magnetic field, N_S is the number of scans/repetitions, T_2 is the T_2 transverse relaxation time and T is the temperature. Most of these are already discussed in the preceding text, but there are few differences. The number of scans N_S is analogous to the accumulation time (Section 2.1.2), assuming that the delay between scans is not changed. Relaxation is discussed in more detail in Section 3.1.1, but the shorter T_2 means a faster relaxation and broader signals after DFT which decreases SNR. The single temperature term T takes both contributions of thermal noise and polarization approximately into account, but doesn't hold in low temperatures, in which liquid state NMR is not very practical anyway. Other changes and difficulties related to temperature (such as stability or viscosity and subsequently T_2 relaxation) are also easily more critical.

Quantitative NMR 3

"I asked myself about the present: how wide it was, how deep it was, how much was mine to keep."
--Kurt Vonnegut, Slaughterhouse-Five

Quantitative measurement in essence means that in addition to observing the type of species, also the *amount* is measured. In basic NMR spectroscopy, the identification of different compounds is based on observing signals with specific frequencies corresponding to distinct molecular structures, so in its simplest form, quantitative NMR means that the *intensity of each signal is directly proportional to the concentration of chemically equivalent nuclei responsible for the signal*. Obtaining quantitative *results* doesn't require this of course: correct concentrations can be calculated from various kinds of spectral data if the relationship with concentration is known. While this is just as correct, it is not as straightforward due to additional computations and calibration factors/curves being involved, but it still might be a lot easier as a whole.

Any single signal in NMR is in most cases proportional to concentration as the acquisition process is at core linear and non-linear errors are rarely introduced [94, 95]. However, a *quantitative spectrum* usually implies that all of the signals are proportional to the concentration in the same way.

In this case, if the concentration corresponding to one signal is known (usually the internal standard), the concentration corresponding to any other signal can be calculated by multiplying the concentration with the intensity ratio of the signals.

The most straightforward way to achieve this situation is to ensure consistent, *uniform response* in every stage of the experiment. In other words, if every step in the acquisition process treats every nucleus identically regardless of its molecular structure, the whole process is uniform, and the signals are comparable and proportional to the concentration automatically.

3.1 Uniform response in the single-pulse experiment

The simplest modern NMR experiment is a 1D, single-pulse experiment presented in Chapter 1, in which a single RF pulse is used to tilt the magnetization to transverse plane, and the resulting FID is recorded (Figure 1.3). The experiment can be repeated several times to improve the S/N and reduce the effects of certain experimental imperfections as discussed in Chapter 2, but ultimately, each repetition can be split into three stages, which should behave uniformly:

1. **Polarization**: during the relaxation delay, observed nuclei must form equal level of polarization regardless of the chemical structure. In other words, the thermal equilibrium must be (re-)established.
2. **Excitation**: the RF pulse must affect all relevant nuclei uniformly, regardless of the exact Larmor frequency.
3. **Detection and processing**: the detection apparatus must have uniform frequency response in the relevant frequency range, and the processing must not introduce any frequency dependent errors.

In the next sections, each stage is discussed in detail.

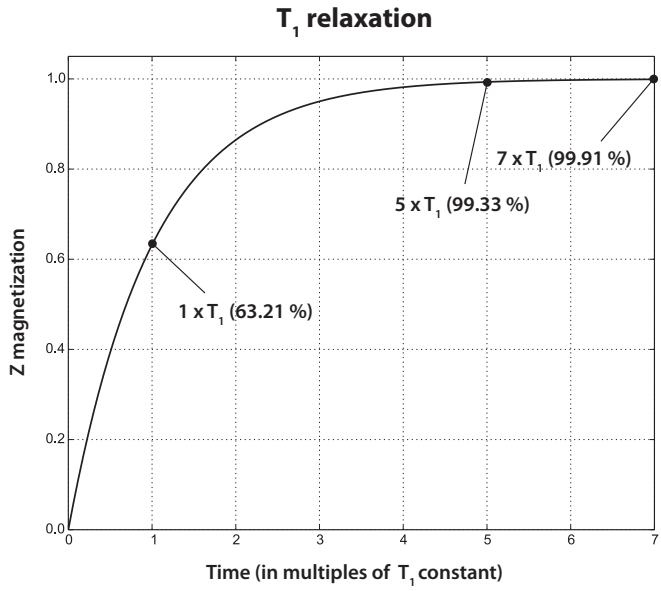

FIGURE 3.1: The recovery (or forming) of z magnetization as a function of time. After $5 * T_1$ constant >99% of the magnetization has recovered, but for very precise quantification even longer delays might be desirable.

3.1.1 Polarization and relaxation

The thermodynamic equilibrium distribution of spin states depends on the energy difference of the states, which again depends on the applied magnetic field, as discussed in Chapter 1. The chemical shifts created by electronic structure influence the magnetic field, but luckily these ppm level differences are negligible compared to the external magnetic field and the equilibrium distribution is essentially same for a given nucleus type. A much more important factor is the *speed* in which this equilibrium is established.

The spin states are constantly proceeding towards thermal equilibrium by various mechanisms, collectively called *relaxation processes*. The speed of these processes depends on multiple factors, including the strength of the applied field, the movement of the molecules (governed by temperature,

molecular structure, viscosity) and, for example, the presence of paramagnetic species. However, relaxation can be divided into two principal types, for which the time constants can be measured:

- T_1 **relaxation** or *spin-lattice relaxation*: the process of regaining the original thermal equilibrium of the spin populations, in other words, the return of net magnetization vector to the +z axis.
- T_2 **relaxation** or *spin-spin relaxation*: the process of losing coherence of the precessing spins, which form the transverse magnetization and observable signal. This can be visualized as shrinking of the transverse component of the magnetization vector, which is precessing in the x-y plane. The T_2 relaxation is always faster or as fast as T_1, as there is no coherence in thermal equilibrium.

Both processes follow exponential functions with time constants. For example in T_1 relaxation, magnetization is recovered to z-axis as described by the following equation (with zero initial magnetization present in z-axis, such as after a perfect 90° pulse) [6]:

$$M(t) = M_0 \left(1 - e^{-t/T_1}\right) \tag{3.1}$$

where $M(t)$ is the magnetization at time t, M_0 is the magnetization in thermal equilibrium, and T_1 is the time constant describing the speed of the T_1 relaxation.

As the relaxation rate varies depending on molecular structure, a sufficiently long relaxation delay has to be used in quantitative experiments to ensure that the different relaxation rates do not affect the observed magnetization. Usually a delay of $5 * T_1$ is considered adequate when using 90° excitation pulse, where T_1 is the longest observed T_1 value in the sample [94]. This corresponds to recovering 99.3% of the magnetization, while even longer delays can be used to slightly improve accuracy (Figure 3.1).

With smaller excitation flip angles a shorter relaxation delay can be used, but for quantitative experiments this approach doesn't provide significant reduction of the total experiment time when the SNR is considered. In fact, a detailed analysis Traficante concluded that optimum SNR per time

unit can be achieved by using $\sim$82.9° flip angle in conjunction with $4.5 * T_1$ recovery delay (>99% recovery) [96]. However, a shorter pulse has a wider excitation bandwidth, which can be useful, as discussed in the next section.

In order to determine a suitable relaxation delay, the T_1 values are estimated or measured (using for example the *inversion recovery* experiment [6]). Luckily, slightly too short relaxation delay will skew the results relatively little, as the process is exponential and the largest changes occur in the beginning of the relaxation period. The importance of sufficient relaxation delay also depends on the differences of T_1 values; if the relaxation rates vary very slightly, even with fast repetition rates the relative differences are small [94].

3.1.2 Pulses and off-resonance effects

The most commonly used and the simplest pulses are so called *hard* or *rectangular* pulses: short pulses of RF radiation with single frequency and phase. The effects of these pulses can be described as rotations of the net magnetization vector as discussed in Chapter 1, and ideally the rectangular pulse rotates the magnetization vector around some axis in the x-y plane independent of the exact Larmor frequency or other small differences between the spins. This kind of ideal behavior would be optimal for quantitative experiments, but unfortunately it can be achieved only if the RF frequency matches the Larmor frequency exactly when the spin is *on resonance*.

If the transmitter and Larmor frequencies differ, the magnetic component of the RF radiation does not exactly oscillate with the precessing spin, and the effective magnetic field and rotation axis is tilted out of the x-y plane [1, 6], as seen in Figure 3.2. This leads to different rotations of the magnetization depending on the Larmor frequency, commonly called *off-resonance effects*. For example, in a single-pulse experiment with 90° pulse, the magnetization is not rotated accurately to the the transverse plane (leading to signal intensity errors) or along y axis (leading to phase errors). As it is impossible to have all spins on resonance if there exists more than one chemical shift, off-resonance effects are inevitable, while they can be negligible in practice.

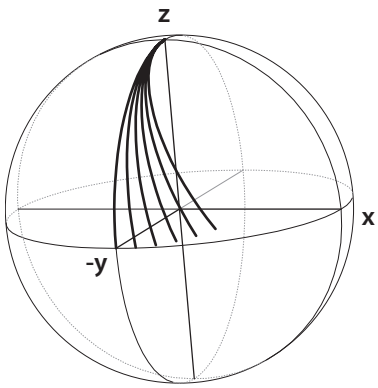

Figure 3.2: The effect of transmitter frequency offset in rectangular pulses. If the transmitter frequency matches the Larmor frequency, the magnetization rotates (or nutates) along the axis determined by the phase of the RF pulse. This axis is always in the x-y plane when the transmitter is on resonance, and a 90° pulse with x phase yields a perfect rotation around the x-axis, transforming magnetization from z to -y. The mismatch between the transmitter and Larmor frequency shifts the axis of effective magnetic field away from the x-y plane towards the z-axis, causing the magnetization to rotate with skewed trajectory and giving rise to intensity and phase errors. The figure was generated with Bruker NMR-Sim, using 25 kHz RF power (90° pulse of 10 μs) with 3 kHz frequency offset steps.

3.1.2.1 Performance of rectangular pulses and RF power

The performance of an RF pulse can be characterized with a profile desribing the flip angle and phase as a function of frequency offset. For a certain range, the properties are approximately constant regardless of the frequency offset, and this frequency window can be said to be the *bandwidth* of the pulse (Figure 3.3). For rectangular pulses, the bandwidth grows if RF power is increased, as the off-resonance effects are reduced. This improvement follows from the stronger magnetic field component associated with the more powerful RF radiation, causing the magnetization vector to rotate more quickly (just like a stronger static magnetic field causes the Larmor frequency to increase). A quicker rotation means that shorter pulse durations achieve the same rotation, and phase errors between the Larmor

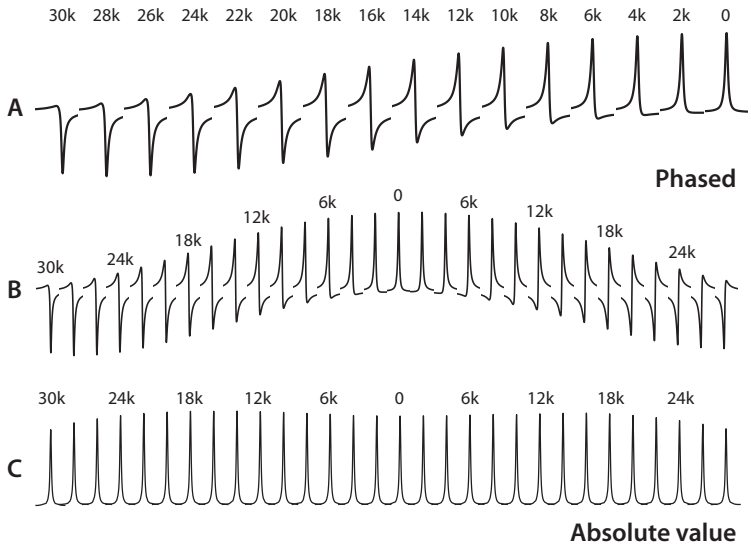

FIGURE 3.3: The response of a rectangular/hard 90° pulse in A) 0-30 kHz B) ±30 kHz frequency offset range using no linear phase correction and C) with absolute value processing. Slightly reduced transmitter power was used to emphasize the effect with 90° pulse length of 13.6 μs, corresponding to RF power of 18.3 kHz. Slight phase errors occur even with small frequency offsets of few kHz, while significant changes in signal intensity start with offsets comparable to the RF power. In single-pulse experiments phase-correction can be used to remove the phase errors, but in most multi-pulse experiments the undesired magnetization would be purged (by phase cycling / PFG elements) or in some cases result in errors and artifacts. The figures were created from real data (water signal in model compound spectrum of **Paper IV**), using SimpeleNMR to process, align and generate sequences of crops from the spectra.

frequency and the oscillating magnetic field component of the RF radiation are reduced, as they have less time to evolve.

The RF power can be expressed as *nutation frequency* [97], which is the frequency of full rotations of the magnetization vector when the transmitter is on-resonance. The rotation rate depends on the magnetic field compo-

nent B_1 of the RF radiation and the gyromagnetic ratio, identically to the Larmor frequency (Eq. 1.7):

$$\nu_{nutation} = \frac{\gamma B_1}{2\pi} = \bar{\gamma} B_1 \tag{3.2}$$

Another common way to express the magnitude of RF power is to state the pulse length of a 90° pulse. These quantities are easily converted to each other, as the RF power is the reciprocal of the duration of a 360° pulse. For example, if the length of a typical 90° pulse is 7.0 μs, the full 360° rotation takes $4 * 7.0 = 28.0\,\mu s$, and the RF power is:

$$\frac{1}{28.0 * 10^{-6}\,s} = 35.7\,kHz \tag{3.3}$$

For quantitative spectra, ensuring that all relevant signals are within the bandwidth of the utilized pulses is crucial in order to make sure that the imperfections of pulses do not affect the signal intensity significantly. The required bandwidth varies depending on the used magnetic field (a higher field produces larger absolute chemical shifts) and the typical chemical shift range themselves.

The bandwidth of a rectangular pulse is around factor ± 0.1 of the RF power [98], implying a bandwidth of ± 3570 Hz for the aforementioned 7.0 μs pulse. In a 500MHz magnet, this corresponds to about ± 7.1 ppm offset for ^{1}H nuclei, yielding 14.3 ppm total bandwidth, quite sufficient for common proton chemical shifts. In ^{13}C spectra however, the same ± 3570 Hz offset corresponds only to a 56 ppm window, clearly inadequate considering the $\sim$200 ppm range of typical chemical shifts.

Another important imperfection regarding RF pulses is the inhomogeneity of the RF field in different parts of the sample, which are caused by the imperfections of the coil and the shape of the sample. Similarly to the frequency offset, the rotation of the magnetization is affected, as different parts of the sample experience varying RF strengths leading to over- or undershooting of the desired rotation.

3.1.2.2 Composite pulses

The off-resonance effects and RF inhomogeneity form the principal imperfections of rectangular RF pulses, which can be severe problems when high field strengths, long pulses and/or large chemical shifts are involved. To address these shortcomings, *composite pulses* combining multiple subsequent hard pulses can be utilized to achieve larger bandwidths. The concept behind these pulses is that flaws of individual pulses are compensated to some extent by flaws of other pulses, and the combined effect yields more accurately the desired change of the magnetization [97]. This can be illustrated by the first 180° pulse introduced by Levitt and Freeman in 1979 consisting of ${90^\circ}_x\ {180^\circ}_{-y}\ {90^\circ}_x$ pulses (or ${90^\circ}_x\ {180^\circ}_y\ {90^\circ}_x$ in older left-hand precession convention [100]), in which the self-compensation is quite evident when looking at the vector model trajectories of the pulse for different frequency offsets (Figure 3.4).

The first composite pulses were mostly 180° inversion pulses, as the long length makes it more susceptible to off-resonance effects leading to a narrow usable bandwidth, but also 90° pulses were developed from the beginning [97, 100, 101, 102]. Composite pulses can also mitigate the effects of RF inhomogeneity, which was a severe problem in the early spectrometers and hindered the usability of multipulse experiments. The pulses were quickly utilized in other new techniques such as 2D NMR and broadband heteronuclear decoupling schemes [103, 104, 105], and advanced pulses emphasizing features such as low phase distortion, arbitrary flip angle or broadband inversion were developed [98, 106, 107, 108].

3.1.2.3 Shaped pulses

Advanced composite pulses, such as the inversion pulses developed by Shaka et al. already used computational optimization methods in development [107], and the next evolutionary step from composite pulses is the modelling of *shaped pulses*, where the phase and amplitude is varied in even smaller grains with pulses consisting of hundreds of steps in microsecond precision. Shaped pulses were initially used for selective excitation with quite long durations (Gaussian, *Band-selective, Uniform Response, Pure-phase* (BURP), *Selective exitatioN fOr Biochemical applications* (SNOB) [109, 110, 111]), but more recently short shaped pulses with wide bandwidths such as *Broadband Inversion Pulse* (BIP) and *Broadband Excitation by Optimized Pulses* (BEPOP) have been introduced, produc-

ing over 40-50 kHz of bandwidth with execution times of 500-1000 μs [112, 113, 114, 115].

An especially useful shaped pulse technique for broadband inversion is *adiabatic pulses* pioneered by Baum et al. in 1983 [116]. These pulses are based on transmitter frequency sweeps (*adiabatic sweeps*), and can produce a very broad inversion with good tolerance to RF inhomogeneity, requiring only that strong enough RF power is used to ensure that the precession induced by the B_1 field is fast compared to the sweep speed (the *adiabatic condition* [117, 118]). This causes the net magnetization to rotate and follow the *effective magnetic field* generated by the RF pulse which travels from z to -z when the transmitter frequency sweeps over a very large frequency range (Figure 3.4). Adiabatic pulses are commonly used in decoupling (for example *Wideband, Uniform Rate and Smooth Truncation* (WURST), [119]), but can be used in many pulse sequences, as long as the relatively long length of the pulse (from $\sim$100 μs to several ms) is accounted for in the sequence design, as the exact inversion time depends on chemical shift [120].

3.1.3 Detection and processing

The final stage of a single-pulse experiment is detection, where the coil, amplifiers and other electronics must also have uniform response within the observed frequency range. This is probably the least difficult problem in contemporary NMR, as modern electronics can easily handle the bandwidths involved.

One important consideration during acquisition is decoupling, for example removing ^{1}H couplings when observing ^{13}C spectra. This is very beneficial, as the complex multiplets generated by several couplings to nearby protons are removed and signals are reduced to singlet resonances. However, the irradiation of ^{1}H nuclei can affect quantification, as magnetization is transferred from relaxing nearby protons through NOE. For this reason *inverse gated decoupling* should be used when obtaining quantitative spectra, meaning that the decoupling is performed only during detection [121].

Needless to say, some of the spectrum processing decisions such as poorly chosen window function or inappropriate baseline correction can also affect quantification. These are discussed further in the Chapter 4.

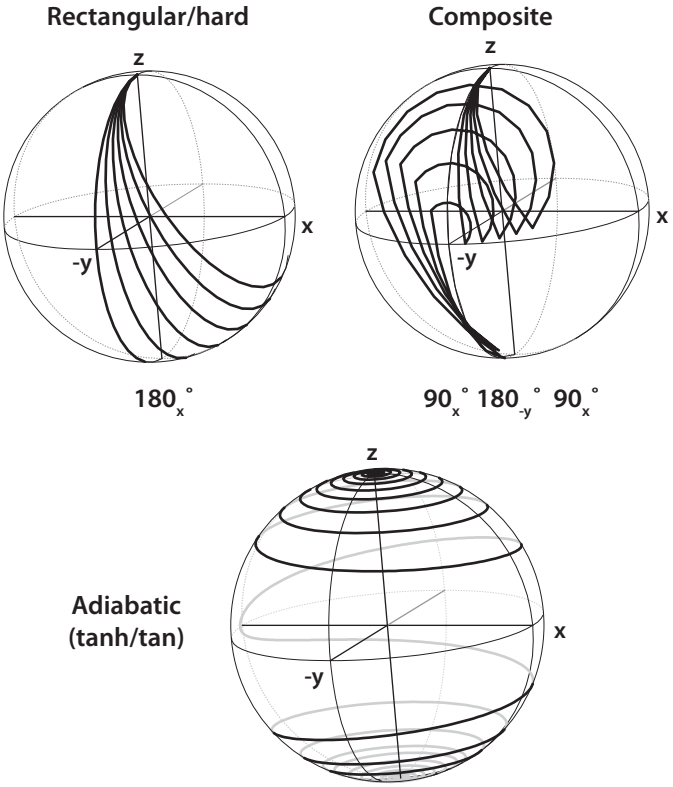

FIGURE 3.4: The magnetization trajectories of hard, composite and adiabatic inversion pulses. The rectangular and composite pulse trajectories were generated with Bruker NMR-Sim using 25 kHz RF power (90° pulse of 10 μs) with 3 kHz offset steps, and the poor performance of rectangular 180° pulse is evident already, while the simple composite is able to achieve nearly perfect inversion. In the adiabatic pulse only one trajectory is shown for clarity, but the others are very similar. The tanh/tan pulse [99] is 500 μs in length and covers a range of 1 MHz.

3.2 Standards, precision and accuracy

In order to quantify compounds from NMR spectra, a standard signal with a known concentration should be present, or the relationship between signal intensity and concentration must be otherwise known. Without this information, all the factors which affect the received absolute signal strength such as tuning, amplifier gain and equipment instabilities can't be accounted for. Even when only comparing relative concentration differences, a standard signal with a constant concentration is beneficial for normalizing the concentrations to some arbitrary scale, while also other normalization methods can be used (Section 4.6.2). There are three basic types of standards [94, 122, 123]:

- **Internal standard**: A known concentration of a chosen reference compound is added to the sample mixture. Alternatively, the reference compound is kept in a separate solution in the sample cavity, typically using a co-axial capillary insert inside the NMR tube.
- **External standard**: A sample with a known concentration is used for calibration before measuring the actual samples.
- **Electronic standard**: An artificially generated RF signal is created and transmitted using another coil or added by other means to the measured signal. The signal is calibrated using a reference a sample and can then be used as a standard for other samples.

Using an internal standard is by far the most common and straightforward method, but the chosen compound must satisfy several requirements depending on the sample material and solvent: the compound should be soluble, inert regarding the sample material, have non-overlapping chemical shift(s) and have a T_1 relaxation rate comparable to the measured compound to avoid long acquisition time. The compound should also be easily measured by volume or weight; volatile compounds such as the commonly used chemical shift reference *Tetramethylsilane* (TMS) are unreliable because they evaporate easily [41].

Using co-axial capillary avoids contact between the reference compound and sample, but the signal overlap and relaxation must be still considered. The sample preparation is also simplified and non-contaminating, as no

deuterated solvent is needed in the sample material, as long as the possible strong solvent signals can be managed or suppressed. The sensitivity is reduced because some of the sample solution is displaced by the insert, and the relationship of the sample and insert volumes is critical when determining absolute concentration.

With an external standard, there are not much requirements for the used compound, as it is measured separately from the actual samples. However, the quantification is affected by all variabilities in the measurements and instabilities of the spectrometer, such as the accuracy of the pulse length calibration and sample tube volume [122]. Changes in the dielectric properties of the sample and probe tuning also affect the efficiency of the coil (probe/coil *quality factor Q*), which can vary significantly with changes of solvent or salt concentration. This can be largely compensated by utilizing the calibrated 90° or 360° pulse length: the principle of reciprocity states that the quality factor is inversely proportional to the pulse length when the same coil is used for irradiation and detection [17, 51], so multiplying the signal with the ratio of the calibrated pulse lengths compensates the changes in Q. This kind of correction is used in *PUlse Length based CONcentration determination* (PULCON) techniques [18, 122, 124], and makes it possible to obtain accurate measurements even in different solvents.

Electronic standards were developed in MRI context and introduced to NMR in 1999 with *Electronic REference To access In vivo Concentrations* (ERETIC) [123]. Using an electronic reference is very versatile, as the signal is synthesized and its frequency and intensity can be easily chosen. However, the signal intensity is affected by coupling factors between the transmitting and receiving coil, including tuning and dielectric properties of the sample [95]. As these can change, the change of the solvent or salt concentration needs recalibration in order to maintain accurate results. Recent improvements have tried to eliminate this limitation using different coil design[125], using PFG coils [126] or injecting the signal directly into the receiver path and using pulse calibration to account for differences in coil quality [95]. The need for calibration, changes in signal routing and/or additional hardware make electronic standards quite complex and have slowed the adoption of the technique.

3.2.1 Precision and accuracy of quantitative NMR

With internal standard, precision and accuracy of <1 % can be expected in quantitative ^{1}H NMR with adequate SNR of ∼1:400 [41, 127, 128, 129, 130]. In optimal conditions *Relative Standard Deviation* (RSD) values of the integrals as low as ∼0.3 % can be reached yielding comparable results with chromatographic methods [127, 131], but in such levels everything counts: operator differences in data processing can yield 0.3-1.0% differences, and similar level of errors can be easily made in weighting or purity of the internal standard [127, 122]. The inclusion or exclusion of ^{13}C satellite signals means ∼1.1% difference, although they can be removed by decoupling [132]. Even handling of the NMR tubes can drastically affect the results in reaction monitoring [133].

Also the quantification method must be considered: a Lorentzian line shape extends to infinity, so a basic numerical integral always captures only part of the signal. Including 99 % of the signal can be achieved with a integration area of 24 times the line width in both directions, while 99.9 % requires 76 times correspondingly [134]. For a 1 Hz ^{1}H signal at 500 MHz, these translate into 0.096 and 0.304 ppm integration windows. Line shape fitting is at least in theory more accurate in this respect, but only when fitting uncomplicated signals: in complex metabolite quantification scenario, Tredwell et al. reported up to 20 % RSD in results produced by five different human operators [135].

The additional instabilities and variation between measurements cause external standards to be less accurate, but RSD precision of close to ∼1 % can be still achieved [122]. The errors can be minimized by performing the calibration before each measurement, and with recent highly stable spectrometers values under 0.6 % are claimed [136]. The electronic standards seem to suffer from more instabilities, variation of several percent can be expected even with recent improved techniques if the dielectric properties of the sample fluctuate [95, 123, 125, 126].

3.3 Quantitative multipulse experiments

In multipulse NMR experiments additional factors affecting the observed signals are introduced and some of the already introduced problems become more complex. If an experiment with uniform response is desired, at least the following aspects must be considered:

- **Pulses and off-resonance effects:** Like in the single pulse experiment, each of the pulses or pulse elements in the pulse sequence must treat spins uniformly, or the non-uniform effects must be compensated somehow. Furthermore, in multipulse experiments the errors are cumulative and the phase response is more important, so even small errors can become relevant.
- **Relaxation during pulse sequence:** T_1 and T_2 relaxation is proceeding during the execution of the pulse sequence as usual, and differences in the relaxation rates can lead to a non-uniform loss of magnetization. This must be accounted for, or the pulse sequence must be short enough so that the relaxation can be safely ignored.
- **J-couplings**: The J-couplings can evolve during the pulse sequence, and in many pulse sequences this is of course exactly what is intended. However, the coupling constants usually vary between different nuclei depending on the molecular structure, which can easily create significant differences in the evolving magnetization. These effects must be compensated for or taken into account.

Building multipulse and 2D experiments with uniform response can be very demanding, and using calibration or other computations to compensate for the introduced errors can be much more feasible. In the following sections, experiments aiming at uniform response are examined first, and approaches based on computations or additional post-processing of the spectral data are discussed in more detail in Section 3.6.

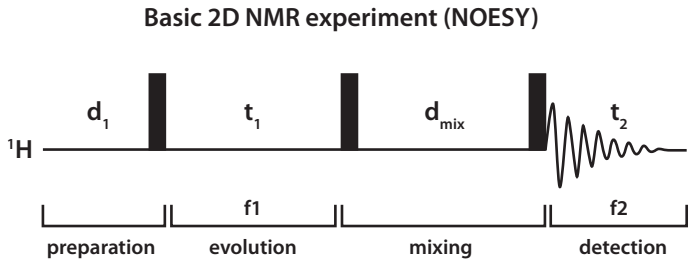

Figure 3.5: The principal parts of a 2D NMR experiment [6, 137, 138]. Here a 2D NOESY experiment is presented as an example [55, 56]

3.3.1 Two-dimensional NMR

The principle of 2D NMR was conceived by Jeener in 1971, and subsequently developed by Ernst and others [137, 138, 139]. The benefit of using 2D NMR for quantification is the increased resolution: by separating the signals into an additional dimension, overlap is reduced and integration of the signals becomes easier. Improved resolution can also be achieved by changing the type of the observed nuclei, for example, using ^{13}C instead of ^{1}H, as the former has a larger chemical shift dispersion than the latter. A low gyromagnetic ratio or natural abundance of an alternate nuclei can often prohibit its direct observation, but as discussed in Chapter 2, the use of heteronuclear 2D experiments and indirect detection can improve the situation considerably. The drawback is that the complexity of 2D NMR makes quantitative measurements much harder to perform.

In 1D NMR experiments, there is only a single time dimension: the directly observed FID signal varying with time. During this dimension, called t_1 in 1D experiments, the spins evolve with associated J-coupling and chemical shifts, and the resulting signal is sampled. The frequency components can then be visualized as a spectrum using Fourier transform, as discussed in Chapter 1. Two-dimensional NMR experiments are very similar, but introduce an additional time dimension: this dimension is not directly observed, instead, several spectra with incrementally longer time delay are acquired, and the changes in magnetization during the time period are encoded to the resulting set of 1D spectra as signal intensity

and phase changes. The frequencies of these changes can be then similarly determined using additional Fourier transform across the 1D spectra.

Any sequence of 1D experiments with a varying time element can be considered a 2D experiment, such as pulse length determination by obtaining a sequence of 1D experiments with different pulse durations. However, modern 2D NMR experiments are usually divided into four principal stages, as shown in Figure 3.5:

1. **Preparation:** The magnetization is brought to some stable and repeatable state, usually by a long enough relaxation delay followed by a sequence of RF pulses.
2. **Evolution:** The desired couplings and/or chemical shifts evolve for a variable length of time. The experiment is designed in a way that the magnetization evolves in some specific and meaningful way during this phase.
3. **Mixing:** The magnetization is transferred or mixed to other nuclei. Some experiments can omit this phase.
4. **Detection:** The magnetization is detected as usual.

The time dimensions (t_1 and t_2) are named sequentially in the pulse sequence, so in 2D experiments t_1 corresponds to the evolution period, or *indirectly detected* dimension, while t_2 is the *directly detected* dimension.

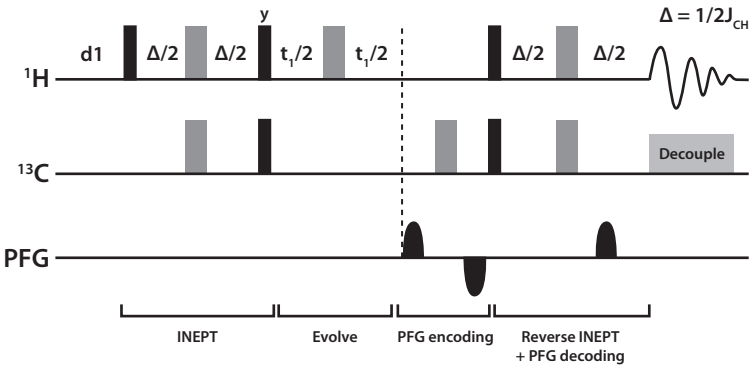

Figure 3.6: The basic ^{1}H-^{13}C HSQC experiment with PFG coherence selection [61, 16]. Magnetization is transferred to ^{13}C with the INEPT element, carbon chemical shift evolves during t_1 after which magnetization is transferred back to ^{1}H. The INEPT transfer period is tuned to approximately match $^{1}J_{CH}$ couplings present in the target molecule(s). PFG is used for coherence selection, as the ratio of ^{12}C bound protons is very large. Many variations of the experiment exist.

3.4 Quantitative HSQC and uniform polarization transfer

Quantitative HSQC [140] is not the simplest quantitative multipulse NMR experiment, but it was the first experiment which tackled the chief problem of non-uniform polarization transfer, and the incremental development of the experiment is a good example of the many factors involved in quantitative multipulse NMR experiments. In the basic HSQC experiment [61], the idea is to transfer the magnetization from abundant and high gyromagnetic ratio nuclei (such as ^{1}H or ^{31}P) to a lower one (such as ^{13}C or ^{15}N), and then back for detection, in order to increase sensitivity on observing the low-gyromagnetic nuclei. This is accomplished with two INEPT elements with the t_1 evolution period in between, as shown in Figure 3.6. The resulting 2D spectrum has correlation peaks between coupled nuclei, providing also useful structural information.

Polarization transfer using INEPT element was already introduced in section 2.2.7, and as discussed earlier, only the antiphase magnetization is

transferred. This makes the success of polarization transfer dependent on the J-coupling and the length of the Δ delay utilized to develop the antiphase magnetization. Consequently, the delay is usually optimized to some compromise value matching as closely as possible the ${}^1J_{CH}$ couplings present to produce maximum antiphase magnetization (Eq. 2.21). The J-couplings vary depending on molecular structure, and as the selected delay can't match all of the coupling constants present, polarization transfer with INEPT is inherently non-uniform.

In a regular HSQC experiment, the non-uniform transfer is not a major problem, as the coupling constant variation is small enough in practice that a significant amount of signal will get through anyway: the natural range of J_{CH} couplings is around 115-220 Hz [140], so a compromise value around $\sim$150 Hz works fairly well. Adiabatic pulses can also be used to approximately compensate for the differences of J_{CH} using *Compensation of Refocusing Inefficiency with Synchronized Inversion Sweep* (CRISIS), as they are approximately linearly correlated with chemical shift [141].

For quantitative experiments however, this is insufficient. The exact behavior of the transferred magnetization can be modeled using product operator formalism [7], and the volume of the observed 2D correlation peak V_C can be calculated to be as follows [140]:

$$V_C \propto sin^2(\pi\Delta^1 J_{CH}) \tag{3.4}$$

The first solution to neutralize this effect was presented with the *Quantitative HSQC* (Q-HSQC) experiment by Heikkinen et al in 2003 [140] with the key idea that while a single HSQC experiment has non-uniform polarization transfer, combining the results of several experiments with differently optimized polarization transfer periods can yield approximately constant *combined* transfer (Figure 3.7). The number of attached hydrogen nuclei still affects the amount of signal (for example CH_3 will get three times more than CH), but this can be easily taken into account as in regular quantitative ^{1}H spectra.

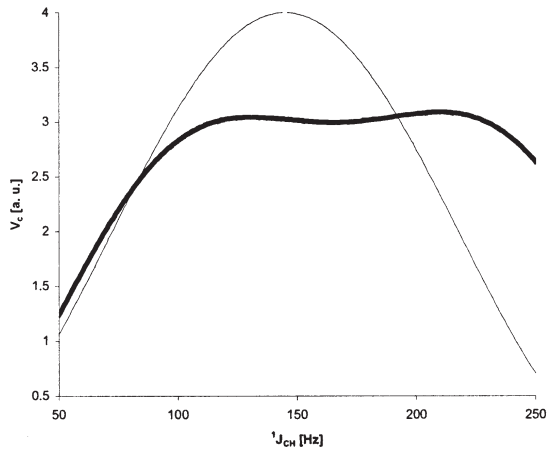

Figure 3.7: The response of regular (gray) and modulated (black) HSQC experiments, depending on the coupling ${}^1J_{CH}$. The figure shows a combined response of four HSQC experiments, with Δ optimized for ${}^1J_{CH}$ of 145 Hz (3.45 ms) and experiments with Δ values of 2.94 ms, 2.94 ms, 2.94 ms and 5.94 ms. The transfer can be considered approximately uniform between 115 and 220 Hz. Reprinted with permission from "Quantitative 2D HSQC (Q-HSQC) via Suppression of J-Dependence of Polarization Transfer in NMR Spectroscopy: Application to Wood Lignin" by Heikkinen et al. [140]. Copyright 2003 American Chemical Society.

3.4.1 Improving Q-HSQC

The modulation of INEPT provided a solution to the principal obstacle for non-uniform response in HSQC, and the Q-HSQC experiment was later refined with further improvements to more minor problems affecting quantification, as discussed in the next sections. Polarization transfer using J-coupling is also a central technique utilized by many other 2D and 1D NMR experiments, and the basic idea of modulating the transfer period was subsequently used in other experiments, as discussed later.

3.4.2 Homonuclear couplings: Q-CAHSQC

Shortly after the original Q-HSQC, Koskela et. al. published an improved experiment, *Quantitative, CPMG Adjusted HSQC* (Q-CAHSQC) [142], which used CPMG-INEPT polarization transfer [143, 144] based on the X-16 version of the *Carr-Purcell-Meiboom-Gill* (CPMG) pulse train [145]. The idea was to suppress the evolution of the ^{1}H-^{1}H couplings, which can create significant loss of signal in cases where protons have many and/or large homonuclear couplings.

By sheer luck, the CPMG could be neatly added to the pulse sequence in two sections, in which the CPMG is alternatively only performed in the ^{1}H channel in the first section (nullifying the evolution of ^{1}H-^{13}C coupling), and always in both channels in the second section. This was possible as the required delays were three times 2.94 ms and one time 5.92 ms, approximately double the first delay.

3.4.3 Off-resonance effects: Q-OCCAHSQC

Large off-resonance effects can lead to significant signal loss as discussed earlier. In HSQC, the problem is mainly in the several 180° carbon pulses, and the original Q-HSQC already utilized basic $90_x180_{-y}90_x$ composite pulses in the INEPT periods. The CPMG sequence introduced with Q-CAHSQC was demonstrated to further improve the situation over the original composite pulses, even while using rectangular pulses [142]. Additionally, composite carbon 90° and 180° pulses in the middle section were suggested, further boosting the performance.

An even more drastic improvement was introduced with the Q-OCCAHSQC experiment in 2010 [146]: by utilizing adiabatic tanh/tan inversion pulses [99] in the CPMG pulse train along PM-BEBOP pulses [115], a usable

bandwidth of about ± 20 kHz is achieved. This corresponds to approximately a 180 ppm ^{13}C window at 900 MHz field strength, enough to satisfy all but the most extreme field strengths and chemical shifts.

3.4.4 Required transients: QQ-HSQC

Another drawback with the original Q-HSQC family of experiments is the number of scans required for each 2D increment: the four different polarization transfer periods necessitate at least four transients for each increment, eight if even basic two-step phase cycle is desired.

The *Quick, Quantitative HSQC* (QQ-HSQC) experiment introduced by Peterson et. al. in 2007 [147] solved the problem by using slice-selective adiabatic sweep pulses to acquire the signals corresponding to different INEPT periods from different parts of the sample. This effectively multiplexes the four required experiments into one, reducing the minimum number of transients to one, as with a regular HSQC experiment. However, the sensitivity itself is not improved as the sample volume and thus signal is also divided, so the experiment time is reduced only when several transients are not desired to improve SNR. The technique is related to *ultrafast* 2D techniques [72, 148], in which the acquisition of multiple t_1 increments is multiplexed into one experiment.

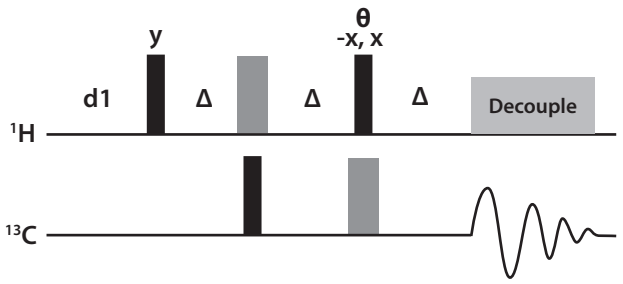

FIGURE 3.8: The basic pulse sequence of a DEPT experiment [149]. The Δ is optimized to match $\Delta = 1/(2*J_{CH})$ as usual and the pulse length of θ alters the amplitude and phase of the signals, depending on the number of attached protons.

3.5 Other polarization transfer experiments

3.5.1 Q-DEPT, Q-DEPT⁺ and Q-POMMIE

Shortly after introduction of Q-HSQC, Henderson introduced *Quantitative DEPT* (Q-DEPT) [150], a version of *Distortionless Enhancement of NMR signals by Polarization Transfer* (DEPT) [149] utilizing a similar delay modulation scheme. The DEPT experiment is an 1D experiment, with a similar idea of transferring the ^{1}H magnetization to ^{13}C to enhance sensitivity, but the detection is performed directly from ^{13}C (Figure 3.8). Additionally, varying the length of a single ^{1}H pulse produces spectra with different intensities for CH/CH_2/CH_3 carbons (Figure 3.9), which can be used to assign the type of carbons present. This provides similar information as the simpler *Attached Proton Test* (APT) experiment [151], which doesn't utilize polarization transfer. DEPT itself is an improvement over the INEPT based polarization transfer and editing reported earlier [59, 60], having only one ^{13}C 180° pulse.

The response profiles of the polarization transfer delay (Δ) and editing pulse (θ) can be derived similarly as for Q-HSQC using product operators, and a set of delays with more uniform combined response can be found. Unfortunately the response formulas Henderson used to derive timing modulations of Q-DEPT were inaccurate for CH_2 and CH_3 carbons, and the sequence worked only partially. It was also only evaluated

using a single compound (Ethylbenzene), which masked the problems.

A revised method, Q-DEPT$^+$, was published by Jiang et. al. in 2008 [152]. In both experiments, the length of the editing pulse is modulated in addition to modulating the polarization transfer delay. The delay modulation (8 steps), pulse length modulation (6 steps) and phase cycling (8 steps) produce a quite long minimum experiment time in Q-DEPT$^+$, as the minimum number of transients is 384 ($8 * 6 * 8$).

Jiang et. al. also proposed an alternate experiment, *Quantitative POMMIE* (Q-POMMIE), a similar modification to the *Phase Oscillations to MaxiMize Editing* (POMMIE) [153] experiment, a DEPT type experiment in which the phase, not the length of the ^{1}H pulse is used for the editing. This method is probably more tolerant to pulse calibration errors, as small errors in the DEPT editing pulse can result in quite large deviations from the intended response. In the proposed scheme the minimum repetitions requirement is also lower (2 step phase cycle, 6 step edit pulse cycle, 8 step delay cycle = 96 transients).

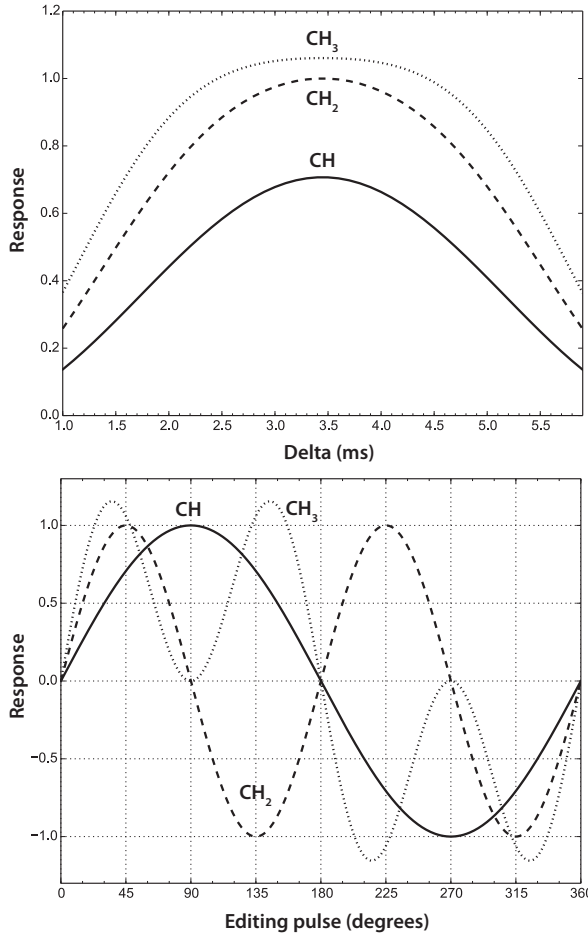

FIGURE 3.9: The calculated response of a DEPT experiment depending on the delay Δ and the flip angle of the editing pulse (θ), for different types of carbons with the same coupling constant ${}^1J_{CH} = 145$ Hz. For Δ (upper plot), the edit pulse was fixed to 45°, yielding a positive signal for all carbon types. Similarly for θ (lower plot), the Δ was fixed to 3.448 ms, the optimal value for ${}^1J_{CH} = 145$ Hz.

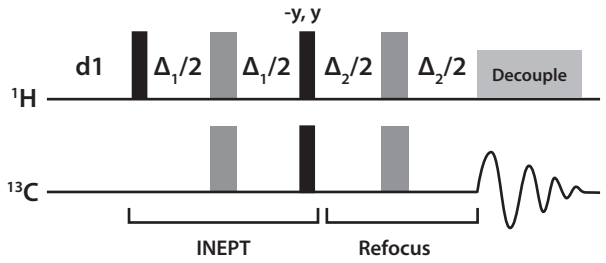

Figure 3.10: The basic pulse sequence of a refocused INEPT experiment [57, 154], consisting of regular INEPT and a refocusing element with simultaneous 180° pulses. As usual the Δ_1 is optimized to match $\Delta_1 = 1/(2 * J_{CH})$, while the length of Δ_2 controls the refocusing and provides editing, similarly to the θ pulse in DEPT. The experiment is simplified compared to the pulse sequence used for Q-INEPT-CT, but follows the same principles.

3.5.2 Q-INEPT-CT and Q-INEPT-2D

Another variation of a 1D polarization transfer experiment is refocused INEPT, which is just an INEPT block followed by a refocusing block to create in-phase magnetization and allow decoupling [154] (Figure 3.10). The two delays (Δ_1 and Δ_2) work quite similarly as the Δ delay and θ editing pulse in DEPT, making similar results possible.

The **Paper I** of this thesis introduces the *Quantitative INEPT, Constant Time* (Q-INEPT-CT) experiment, a quantitative modification of the refocused INEPT experiment with uniform polarization transfer. Again, behavior of the magnetization was modeled by product operator formalism, and a suitable combination of delays producing a near-uniform polarization transfer was created. The Q-INEPT-CT experiment modulates the delays Δ_1 and Δ_2 in pairs, with 8 delay pairs resulting in a relatively uniform response (Figure 3.11). Two alternate modulation schemes were developed, regular (with better theoretical performance) and time-restricted, which performed better in actual real-world use (Figure 3.12).

The Q-INEPT-CT experiment has a number of advantages over Q-DEPT$^+$. First, the twin modulation scheme needs only 8 modulations to produce a uniform response, lowering the minimum number of transients to 16

with a two-step phase cycle. Secondly, as the experiment doesn't rely on pulse lengths to do the editing, it should be more tolerant to pulse calibration errors and RF inhomogeneity. The pulse sequence is also constructed to have a constant length, independent of the timing modulation, which results in constant losses due relaxation for each modulation.

The experiment also addresses off-resonance effects arising from the large ^{13}C chemical shifts by utilizing wide-band 180° composite pulses for carbon [107, 108], while PFG elements are used to purge undesired magnetization in the beginning of the pulse sequence and to clean up the imperfections of 180° pulses. The experiment was tested with several molecules, and produced a quantification precision comparable to quantitative carbon experiments with similar SNR (with errors around $\pm$5%).

Also a 2D version of the experiment Q-INEPT-2D is presented in **Paper III** for the analysis of base oils. By adding a t_1 delay in the first INEPT period, the proton chemical shift modulates the transferred magnetization producing a carbon detected 2D-INEPT experiment, which is useful when ultimate ^{13}C resolution is desired [155, 156]. Additionally, States-TPPI phase incrementation was added [157] and the modulation scheme was halved to four modulations in order to reduce the minimum transients for each increment.

Recently Manu and Kumar used genetic algorithms to re-optimize the delay modulations of Q-INEPT-CT in *Genetic Algorithm Q-INEPT-CT* (GAQIC) experiment [158], claiming to reduce the experiment time to half without degrading accuracy. The improvement is accomplished by increasing the total polarization transfer, and while the claim is true, it is achieved only with optimizations producing non-equal response for CH, CH_2 and CH_3 carbons (but uniform response regarding J-couplings for each type). Quantification using this kind of experiment is of course possible, but necessitates additional calculation and knowledge about the type of the carbons. These kind of optimizations were actually also examined and considered during the development of Q-INEPT-CT, but deemed not worth the drawback of non-equal carbon response.

Uniform response regardless of carbon type was also presented, but the reported modulation schemes yield negligible ($< 1\%$) improvement in signal strength compared to original Q-INEPT-CT, even with the used elegant optimization algorithm.

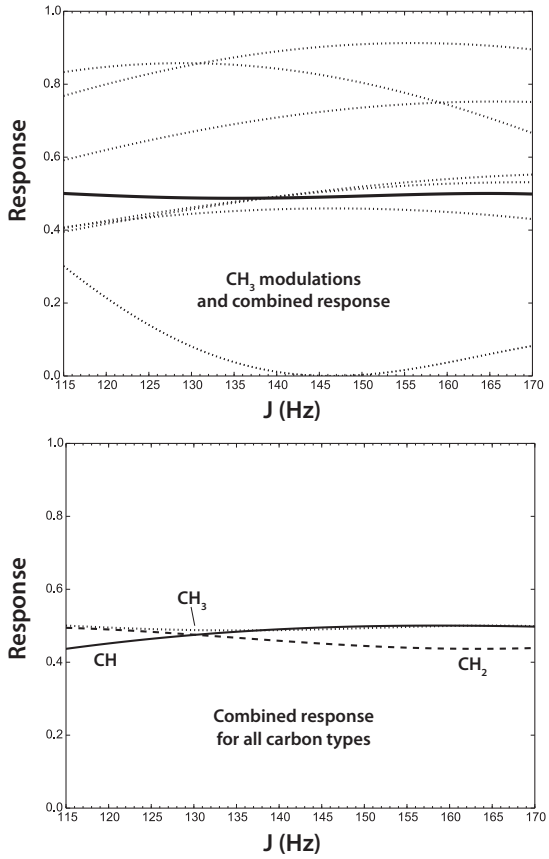

Figure 3.11: The response of individual Δ_1 and Δ_2 value modulations in Q-INEPT-CT for CH_3 carbons (top) and combined response for all carbon types (bottom). In the top figure, the individual modulations (dotted lines) and the normalized sum of the 8 modulations (solid line) for CH_3 carbons are shown. Each modulation is a pair of delays used in the Q-INEPT-CT experiment (Δ_1 and Δ_2), yielding different response profiles depending on the ${}^1J_{CH}$ and carbon type (only CH_3 shown here for clarity). In the bottom figure, the combined response of the modulations for all carbon types is shown for the proposed time-restricted modulation. It is easy to see that about half of the potential intensity is lost, but the combined response is quite uniform for ${}^1J_{CH}$ of 115-170 Hz.

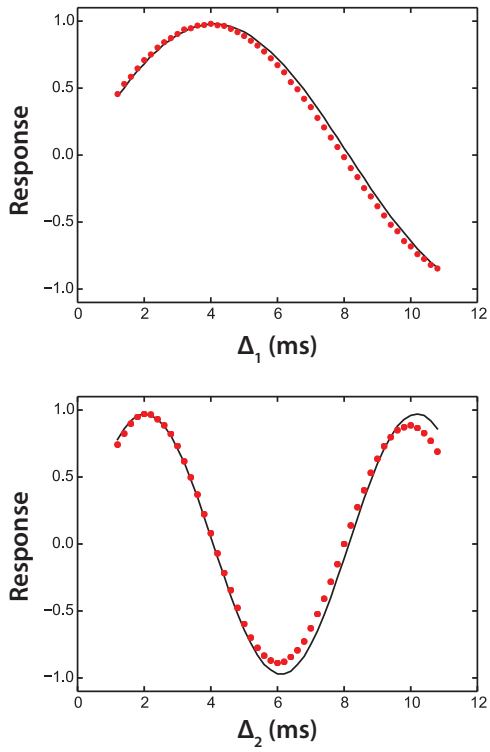

FIGURE 3.12: Calculated (black line) and observed (dots) response of a CH_2 carbon signal in ethylbenzene, measured during development of Q-INEPT-CT presented in **Paper I**. The deviations seen with longer >5 ms delays were the principal reason of creating the time-restricted modulation set, in addition to avoiding modulations relying on negative responses. The time-restricted version performed better in real-world use.

3.6 Quantitative NMR through computation

The use of quantitative NMR experiments is not required for quantitative results: if the relationship between the observed signal intensity and the concentration of the substance is known, the observed values can be corrected easily. For example, if the intensity of signal I_{real} is attenuated or boosted by a factor of a in the experiment or processing, the observed signal has the intensity of:

$$I_{obs} = a\, I_{real} \tag{3.5}$$

and the original correct intensity can be recovered simply by division:

$$I_{real} = \frac{I_{obs}}{a} \tag{3.6}$$

In NMR spectroscopy, the observed signal can be usually assumed to have linear relationship to the concentration, in which case the observed signal intensity can be described by the basic linear equation:

$$I_{obs} = a\, I_{real} + b \quad \text{or} \quad I_{real} = \frac{I_{obs} - b}{a} \tag{3.7}$$

and to calculate the correct results, coefficients a and b must be determined for each signal. It must be noted that while this linear relationship can be assumed for a single signal, the magnitude of the coefficients can depend non-linearly on some quantity. For example, the relationship of a transmitter offset and signal intensity can be very complex, but for a certain offset, the signal intensity is still linearly correlated with concentration.

3.6.1 Calibration curves

The above linear equation is the simplest example of a *calibration curve*, a mathematical model which describes the relationship between observed signal and actual concentration. The required coefficients a and b can be found by preparing a series of authentic standards of different concentrations, measuring the corresponding signals and performing a simple linear regression. Similarly, more elaborate equations can be fitted to

model more complex relationships between signal and concentration. It is also possible to calculate correction factors for specific error sources, such as for mismatch of ${}^{1}J_{CH}$ in polarization transfer or off-resonance effects [156, 142].

In practice, with many experiments and a properly corrected baseline, signal is not only linearly dependent on concentration, but zero signal is observed for zero concentration. This means that $b = 0$ in the linear equation, reducing the equation to Eq. 3.5, or in another words, the signal intensity is proportional to concentration c:

$$I_{obs} \propto c \tag{3.8}$$

Using calibration curves to derive quantitative information is relatively simple in principle, but practical application has some drawbacks:

- **Determining correction factor / coefficients:** finding the correct correction factor or coefficients for the model is usually the main problem, or that the determination of the correct factor adds too much work. For example, using authentic standards requires pure reference compounds for all species which are to be measured.
- **Complexity:** the added complexity of performing the required computations to derive the real concentrations, which can be minimized with proper tools.
- **Clean integrals required:** signals can't be corrected properly unless they can be measured independently, as each signal might require a different correction. This requires enough resolving power to obtain non-overlapping signals.
- **Non-uniform errors:** assuming that all integral measurements have equally distributed error terms, the correction also amplifies or diminishes this error. This can be taken into account and visualized, but can make the interpretation of the results less intuitive.

3.6.2 COSY and TOCSY

COrrelation SpectroscopY (COSY) and *Total Correlation SpectroscopY* (TOCSY) are simple and sensitive 2D ^{1}H-^{1}H correlation experiments which provide cross-peaks between coupled protons (COSY) or coupled spin systems (TOCSY) [15, 137, 138, 159, 160]. The advantage of utilizing these experiments in quantification over plain ^{1}H spectra is additional resolving power: overlapping signals can be resolved if the resonances are coupled to distinct chemical shifts, even if all relevant signals suffer from overlap in 1D spectrum.

The expected cross peak intensity and form can be analyzed by examining the pulse sequence and the resulting observed magnetization. In COSY, the main problem is that the cross peaks are in antiphase form, which results in a total integral volume of zero by definition. This feature can be overcome by only integrating the positive or negative signals from phase-sensitive COSY or ideally *Double Quantum Filtered COrrelation SpectroscopY* (DQF-COSY) [15, 161]. With a linear calibration curve, this approach can provide excellent results: Giraudeau et al. achieved accuracy of 2 % and standard deviation of 0.4 % [162]. Alternatively absolute value processing (Section 1.4.5.2) can be used, but this broadens signals and some self-cancellation can occur [94].

TOCSY spectra have in-phase cross peaks and as such are much more suitable for quantification. Still, dispersive components can create distortions in the line shapes, but they can be filtered with suitable elements [163]. The magnetization transfer also depends on the length and implementation of the mixing period, and the theoretical prediction of the cross peak volume is more complex, while it can be simulated [164]. However, identical mixing periods should provide equal transfers in similar measurement conditions.

While quantification is possible using these experiments, the use of a calibration curve is required to account for aforementioned factors. The use of a calibration curve also accounts for other minor inaccuracies such as T_1/T_2 relaxation rate differences. However, the added resolving power is modest, and still comes with a cost of more complex experiments, full 2D processing and integration.

3.6.3 HSQC$_0$

A drastically different and computation-based approach to dealing with the non-uniform aspects of the HSQC experiment was developed by Hu et al. in 2011: instead of trying to make every aspect of the experiment as uniform as possible, the response for each signal is characterized by performing 3 special experiments, each composed of a series of 1-3 constant time HSQC experiments [165]. The idea is that each successive HSQC causes a similar reduction in signal caused by the various factors presented in the previous sections, and after quantifying signals from each experiment, a linear equation can be fitted to the results. Analogously to a regular linear calibration curve, a signal corresponding to the theoretical 0th HSQC experiment can be calculated, in which no signal loss has occurred.

The advantage of this technique is that it takes into account all of the factors affecting the signals, even those which are not anticipated or deemed significant for a certain sample. However, the method is quite slow and complex, as it involves acquiring three HSQC experiments, integrating relevant signals and performing linear regressions. Also, enough signal should be recovered even after three subsequent HSQC experiments (with a total length of $\sim$40 ms) to obtain three data points, which can be problematic with species undergoing fast T_2 relaxation.

3.6.4 Non-uniform sampling

In a *Non-Uniform Sampling* (NUS) or *Non-Linear Sampling* (NLS) scheme, only a fraction of the data points are sampled, and the missing points are reconstructed using various mathematical methods such as maximum entropy reconstruction [166, 167, 168, 169]. The aim is to produce a nearly equivalent data set in significantly shorter time frame, and as such it has been primarily used to reduce the time requirements of multidimensional NMR experiments [170].

Recently, also the analysis of complex metabolite samples using HSQC experiments have been reported [171, 172, 173, 174]. While all NUS techniques produce some degree of sampling artifacts, comparable performance with uniformly sampled experiments have been demonstrated along with multivariate statistical analyzes [172, 173, 174]. Even so, the setup and processing of these experiments is more complicated, and optimal reconstruction algorithms are still being developed. However, as speed improvements of an order of magnitude can be achieved, it is likely that

the technique continues to be developed further and is eventually established as a routine method for acquiring 2D spectra in high-throughput settings.

3.7 Homonuclear decoupling of ^{1}H spectra ("pure shift")

The principal problem of quantifying ^{1}H spectra is the low resolving power due to signal overlap, which is caused by low chemical shift dispersion and complex scalar coupling patterns due to adjacent ^{1}H nuclei. While heteronuclear coupling patterns are routinely suppressed by decoupling, similar techniques can't quite obviously be used in a homonuclear case without destroying the magnetization of interest. Accomplishing this would be very beneficial, as collapsing the coupling patterns to singlets would reduce the chances of signal overlap dramatically, simplify quantification and improve SNR. This would render the otherwise quite optimal (simple, fast, sensitive) 1D ^{1}H spectra into much more broadly applicable experiment.

Unsurprisingly, in the past decades many different approaches have been proposed to produce homonuclearly decoupled ^{1}H spectra. Here we concentrate on the most common methods based on *J-resolved* (JRES), and the slice-selective separation based on variations of the *Zangger-Sterk* (ZS) element, and their quantitative potential [175]. A more complete review on all reported methods can be found in a recent review by Klaus Zangger [176].

3.7.1 JRES

JRES spectroscopy is one of the oldest 2D NMR experiments [177], which simply resolves the J-coupling constants in the indirect dimension. While the experiment can be used to determine J-coupling constants more precisely in crowded spectra, one striking feature of JRES is that with proper processing it can produce spectra in which ^{1}H-^{1}H couplings appear to be removed: by rotating (or shearing) the spectra by 45°, the separated coupling signals are aligned over each other in the indirect dimension, and a "decoupled" 1D spectrum can be produced with symmetrization and projection (Figure 3.13). Strong coupling can cause artifacts in regular JRES experiments, but these can be removed by modifying the pulse sequence [178].

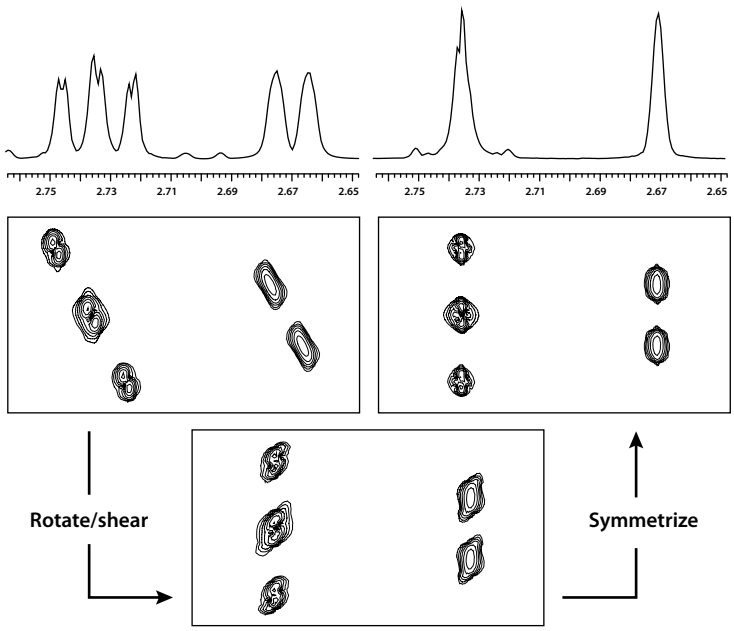

FIGURE 3.13: Typical 2D-JRES processing scheme, with skyline projections on the initial and final spectrum. The spectrum is real data acquired during work on **Paper IV** from epithionitrile model compound (1-cyano-3,4-epithiobutane).

The use of this remarkable feature is limited by the fact that the phase of the signal is modulated also in the t_1 dimension and as a result the signals are a mixture of absorption and dispersion line shapes in 2D, known as "phase-twisted" line shape [179]. Consequently, JRES spectra are normally processed in absolute value mode, which broadens the signals and partly negates the enhanced resolution achieved.

3.7.1.1 JRES processing

The undesired line shape can be alleviated by utilizing sine-type window functions creating a "pseudo-echo" shape to the signal, because time-domain signals with symmetrical decay on both sides of the midpoint minimize

the dispersive component of the signal after DFT [179]. While simple and very usable, this apodization causes a significant reduction in sensitivity, as the first points of FIDs with the largest SNR are severely affected. Quantification is also harshly affected, as signals with faster signal decay are attenuated proportionally more, and the intensity of the signal is weighted strongly by T_2 relaxation rate.

The dispersive component can also be eliminated by more sophisticated processing such as *A Linear Predictive Estimation of Signal Time REversal* (ALPESTRE), in which data points corresponding to experimentally unattainable negative time values in t_1 are constructed using linear prediction [180, 181]. In addition to creating a good line shape, a recent variant of this approach can yield SNR improvements by a factor of about 6 compared to traditional processing, as the t_1 echo can be constructed so that the signal amplitude matches the sine window more closely [182]. A similar echo can be generated in the direct t_2 dimension by more complex modeling, making sine apodization unnecessary and subsequently yielding quantitative spectra with quasi-lorenzian line shapes [183]. Even extracting the full amplitudes, line widths and frequencies has been done with the *Filter Diagonalization Method* (FDM) [184, 185], while it is not suitable for noisy spectra.

Other approaches to improve the line shape involve modifying the pulse sequence and removing the magnetization causing the dispersive components, but this changes additional signals in the 2D spectrum and the traditional rotate/project processing doesn't work. The cross shaped coupling patterns can then be analyzed by employing a pattern recognition algorithm exploiting this symmetry [186]. A Zangger-Sterk block can be used to selectively invert coupled nuclei to build a proper phase-sensitive 2D experiment [187], but this reduces the sensitivity significantly as usual with ZS based methods, with few percent of the signal retained. Methods based on ZS are discussed further in Section 3.7.2.

All of the above advanced methods involve more complex processing along with some limitations, and can't match the sensitivity, simplicity and predictability of a regular JRES experiment and processing. The complex implementation is a major challenge for the wide adoption of these methods, unless the functionality is integrated to the acquisition and processing software by the vendor. For example, even while processing tools along source code are freely available for ALPESTRE, it is designed to work with Bruker

spectra under TopSpin, and uses a mix of Python, GNU Octave and C programs to do the processing [188]. The complex setup can be overwhelming to a typical user, and for spectra acquired using other vendors instruments, re-implementation of at least a part of the software is required. As such, traditional JRES processing seems to remain the most frequently used technique even for quantification.

3.7.1.2 JRES quantification

As already mentioned, JRES spectra utilizing a sine-type apodization function weights the signals strongly according to the t_2 relaxation rate. One striking example of this was encountered during the processing of spectra for **Paper IV**, in which the broad water signal almost vanished while it was the strongest signal in the regular quantitative ^{1}H spectra. Absolute quantification is still possible using calibration curves as discussed earlier, and an accuracy of <1% RSD can be obtained, comparable to regular quantitative ^{1}H spectra [162]. Further, relative concentrations can be compared between samples without calibration, which is especially useful in biological samples, where the variation between samples and grouping is usually more important than the absolute concentrations. Indeed, the use of JRES in the analysis of biological samples such as urine, blood plasma and plant extracts begun already in the 1990s and was established as a routine technique in the 2000s [189, 190, 191, 192, 193, 194].

Another important source of error is the post-processing (rotation, symmetrization and projection), which has been evaluated quite recently [195, 196]. Parsons et al. used both measured and simulated spectra to examine sine and *Sine-bell and Exponential* (SEM) window functions along with the errors introduced by post-processing. Symmetrization reduced the signal intensity to less than half, but with difference of less than 2% between singlet and triplet signals. Most of the errors were caused by the long dispersive tails accidentally overlapping with nearby signals during symmetrization, with errors up to $\sim$33 % (sine window) and $\sim$20% (SEM) being possible with signals close to each other ($\sim$2 *Full-Width-Half-Maximum* (FWHM)). However with a larger separation of >6 FWHM the error drops to less than 5% and approaches negligible at a distance of $>$25 FWHM. Skyline projection and SEM apodization seemed to deliver the best results. The above errors along with sample preparation from biological source and other factors can result in quite large median RSD values of >10% for the integrals [194, 197]. However, biological variations can

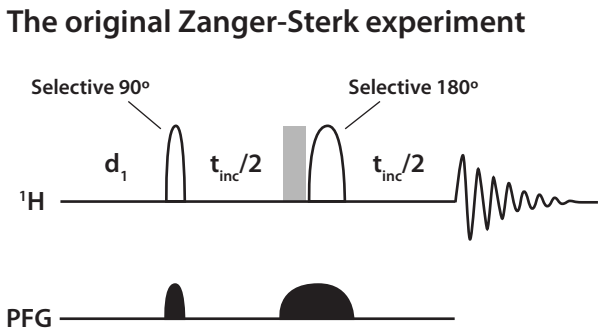

Figure 3.14: The Zanger-Sterk slice-selective experiment for homonuclear decoupling [175]. The two selective pulses which are executed simultaneously with PFG pulses affect spins with different chemical shifts depending on the spatial location in the sample, while the regular 180° pulse inverts all spins. The net result is a refocusing of the homonuclear J-coupling for the spins affected by the selective pulses at the start of acquisition. By performing repeated measurements with incremented t_{inc} delays an artificial FID with decoupled data can be constructed.

still easily be much higher, and the signals remain highly correlated to quantitative ^{1}H results, yielding a viable tool for metabolite analysis.

While not as accurate as quantitative ^{1}H, the definite strength of JRES is its robustness provided by simple acquisition, processing and data-analysis of the projected spectra, while still improving resolving power. Because of this it was chosen as the base for automation in **Paper IV**, which is discussed further in the next chapter.

3.7.2 Slice-selective methods

Using slice-selective methods to achieve homonuclear decoupling was pioneered by Zangger and Sterk in 1997, with the ingenious idea to apply a field gradient in combination with a selective pulse to invert ^{1}H resonances during spin-echo depending on spatial location in the sample [175] (Figure 3.14). As the inversion doesn't affect other passive ^{1}H spins, and by observing only the inverted spins, the J-coupling is refocused after the spin echo. The couplings still exist and start to evolve, so a full decoupled ^{1}H

spectrum is achieved by constructing the FID from several experiments with incremental delays, combining the data points just after the echo where the couplings have not evolved significantly.

While the experiment impressively achieves a phase-sensitive and decoupled ^{1}H spectrum, much of the sensitivity is lost, because for each resonance frequency only a slice of the sample contributes to the signal, typically rendering the sensitivity to a few percent of a comparable regular experiment [198]. Total measurement time is also increased as several experiments must be recorded to construct the FID. Because of the weak sensitivity and complex processing, the technique initially did not receive much use.

Over a decade later, an improved version was presented by Aguillar et al., in which the use of selective pulses was optimized and the FID was constructed in twice as long chunks, making the experiment more feasible with an up to 16-fold reduction in measurement time [199]. Recently the decoupling was also demonstrated during acquisition ("real-time"), making the collection of separate experiments to construct the FID unnecessary [198]. This speeds up the experiment significantly, but the interruption of the acquisition of course leads to some discontinuity artifacts as some relaxation happens during the decoupling element, and the slice-selection still incurs a hefty loss in sensitivity.

The ZS based decoupling has been incorporated to existing experiments such as TOCSY and *Diffusion-Ordered SpectroscopY* (DOSY) to simplify and improve their performance. Especially DOSY benefits from a reduced overlap in processing, as separating the exponential decays of overlapping signals is quite difficult problem [199, 200]. Related methods also include pure shift HSQC, in which the inversion of decoupled protons is based on *BIlinear Rotation Decoupling* (BIRD) elements flipping spins not attached to the ^{13}C nuclei [201, 202, 203]. This version of homonuclear decoupling remarkably doesn't involve sensitivity loss, as the experiment already observes only protons attached to ^{13}C nuclei. Because of this, it can also circumvent strong coupling unlike basic ZS, but geminal non-equivalent protons are not decoupled.

3.7.2.1 Quantification

The slice selective experiments are very impressive, yielding a potentially quantitative decoupled ^{1}H experiment with resolution equivalent to cou-

pled spectra acquired in a field of several GHz [199]. The main obstacle for their quantitative use is their low sensitivity, limiting the use to relatively concentrated samples. Furthermore, the gains from decoupling would be greatest with complex mixtures, which are quite likely also the ones with limited concentration and/or interesting low concentration components. In this context, acquiring quantitative HSQC to improve resolving power can be more desirable, and slice-selective methods seem to not have gained widespread adoption in quantitative use.

Recent developments of anti z-COSY based *Pure Shift Yielded by CHirp Excitation* (PSYCHE) with improved sensitivity [204, 205] and combining hyperpolarization to slice-selective techniques [206] might pave a way for even better pure shift experiments, but the feasibility of these methods in real-world scenarios remains to be seen.

Automation 4

> "The major cause [of the software crisis] is that the machines have become several orders of magnitude more powerful! To put it quite bluntly: as long as there were no machines, programming was no problem at all; when we had a few weak computers, programming became a mild problem, and now we have gigantic computers, programming has become an equally gigantic problem. In this sense the electronic industry has not solved a single problem, it has only created them, it has created the problem of using its products."
>
> --Edsger W. Dijkstra

Automation began with the industrial revolution in the 18th century, which gave birth to mechanical automation and energy production, ultimately leading to a major reduction in manual human labour and forming the basis of modern society. This process has evolved now for well over two hundred years forming complex global production chains and intricate end products, such as the modern smartphone. In modern context, automation is progressively more about automating "thought" as well as mechanical work and logistics: the automation of information gathering, processing and decision making. The rest of this chapter considers mostly this aspect of automation in the context of NMR spectroscopy.

4.1 Computers and software

The chief information processing tools are of course general purpose computers, in which new functionality can be produced just by writing new software. This has significant practical advantages, one of which is the cheapness of duplication: software is essentially information, so after building it the distribution costs are the same as with any data. This cost has diminished as processing power, storage and internet access have become a commodity and as such very cheap, making software distribution virtually free.

On the other hand, writing good and *correct* software turns out to be extremely hard and expensive: in "Code Complete" Steve McConnell proposes based on several sources that the industry average is about one to 25 defects per 1000 lines of delivered code, and achieving 0.1 defects per 1000 lines is rare and requires extensive testing, code review and well thought development methods [207]. This can be compared to the fact that a modern pacemaker contains 80-100 000 lines of code, while operating systems contain several million lines of code [208, 209]. Of course all defects are not important and might not have any practical consequences during the lifespan of the software.

4.1.1 Scientific software

In commercial software the large and dominating software development costs are divided by the users, so the number of potential users and their willingness to support it essentially define the available resources for creating the software. Special and purpose-built software thus tends to be very expensive, while major pieces of software such as modern operating systems can be cheap or even free, as the potential user base is enormous.

This is one of the core principles of software business and an essentially unavoidable problem with many scientific software packages (or any specialist software for that matter): the small audience of the product means that the resources to develop the software are quite limited, while the complexity of the required software might still be quite high and require specialist domain knowledge. As a result, scientific software can be very expensive and/or have low quality, manifested by poor usability, limited features and a large number of defects. Similar effects can also be seen in open source software; while the software is always free, the most used projects tend to

gather most people working on them, and subsequently also gather most testing, donations and sponsorships.

For these reasons, the testing and adoption of newly published methods and algorithms is usually limited by the available software, not the scientific knowledge itself. While the publication of a new method or idea usually includes source code for some sort of implementation, typically a strong IT background and/or programming expertise is required just to install and use the provided tools. Understandably the authors can't be required to support the software or create a commercial product, and as data-analysis and processing is becoming more important in many fields, more specialists with strong IT expertise and understanding of the scientific context are needed.

4.2 Why automate?

The explosion of information gathering and storage capability has not only made automated processing and data-analysis important in scientific context, it has made them a practical necessity. Not much of the current science could be done without the aid of modern computing; the astonishing advances in genetics during the recent decades have been just as dependent on modern computing and algorithm development as they have been on discoveries made in the laboratory. At the same time processing power and data storage are cheaper than ever, and free open source tools such as R [210] or NumPy/SciPy [27, 32] are challenging expensive commercial packages. Things that would have required significant resources a two decades ago are now possible with low-cost hardware and free software.

From the perspective of humans, almost any tedious repetitive task is desirable to be done automatically, if it can be performed adequately. However, tasks can be defined very vaguely, and the correct or desired result can depend on many factors, which are difficult to express accurately in terms necessary for a computer program. Therefore, automation is easiest to do for tasks in which the process and desired result can be clearly defined, and in such context the main advantages of automation can perhaps be summarized as follows:

- **Consistency and robustness**: The results don't depend on the operator. The same treatment is given for every unit of work.

- **Cheapness**: human labor, especially expert human labor, is expensive.
- **Scalability:** more machines or processing power can be produced and bought usually more easily than seeking suitable expert individuals.

With such clear advantages, the reasons for not automating a *suitable problem* boils mostly down to one thing: cost (in time, money and other resources).

One indirect practical consequence of automating data-analysis is that with automation, refining the analysis becomes much cheaper: if automated processing of a certain data set is implemented, small additions or tweaks to the analysis can be done with minor inconvenience. As a result, the tendency of doing these adjustments increases, sometimes yielding better results. If the analysis involves manual work, the cognitive cost for redoing analysis iteration is much higher and sometimes even small errors might be overlooked, because correcting them involves too much work.

4.2.1 The price of automation

The basic relationship between costs and gains can be based on the work involved and the saved resources. This can be illustrated in a common situation, where one has to do a repetitive task on a computer and considers automating the task with a simple script or program. In this scenario, time is usually the only relevant resource, and if it is assumed that the exact solution is not useful for anyone else, the benefit of automation can directly be evaluated by comparing the time it takes to automate the task to the time it saves, multiplied by how many times the task is likely to be performed. This basic relationship is nicely demonstrated in a XKCD cartoon by Randall Munroe (Figure 4.1): an expensive one-off solution is justified if it saves a lot of work, and even miniscule savings are important in tasks which are repeated many times every day. In practice, cognitive costs can effectively deter the implementation of even a clearly beneficial automation, as it involves more effort: assessing the problem and designing a solution (possibly using new tools or techniques) is cognitively harder than repeating an inefficient but familiar procedure.

In larger problems, more factors come to play, while the same principle still holds. How many places or people benefit from the solution and how

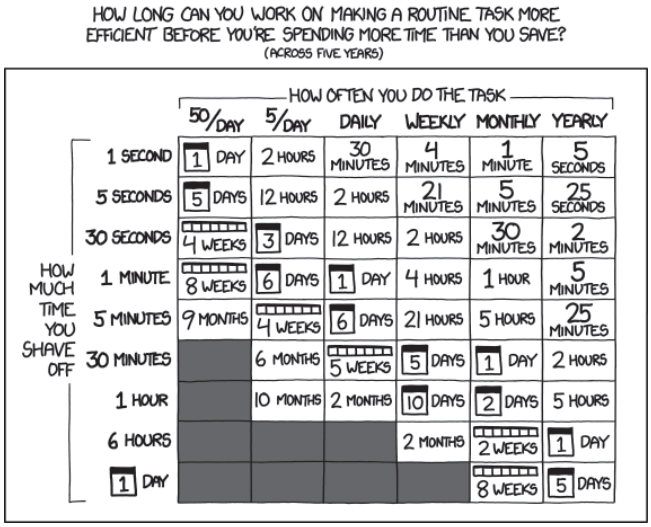

FIGURE 4.1: "Is It Worth the Time?". Comic by Randall Munroe. Licensed by CC BY-NC 2.5 [211]

much less resources are consumed as a result (time, money, raw material)? How much it will cost to scale the solution to all relevant places (training, equipment, premises)? As said earlier, the software development costs can easily be the largest factor, especially when only information processing is needed: usually the hardware is already present and with the hourly wage of a developer at about $50 (median income in the United States [212]), one simply can't buy much development time with the a cost of a modern desktop computer or a few servers.

4.3 Automation in NMR spectroscopy

Obtaining a high-quality NMR spectrum is a quite complex combination of mechanical tasks, high performance electronics and information processing. While the process can be automated almost fully, additional hardware and/or software is usually needed. Unfortunately scientists are quite dependent on the spectrometer vendors to provide any necessary additions, as an individual lab or company rarely has the resources or knowledge to alter a spectrometer or to develop additional components to it.

The cost of acquiring and maintaining the additional equipment should be always compared to the acquired gain. Added complexity nearly always comes with increased costs, and these costs can be difficult to measure, such as unexpected downtime or the learning curve of a more complex interface. A high-capacity sample changer might be very suitable and cost-effective in high-throughput metabolomic analysis, but almost a nuisance in routine organic synthesis work.

Obtaining NMR spectra can broadly be divided into two phases, *acquisition*, in which the actual physical phenomenon is observed, digitized and stored, and *processing*, in which the data is transformed into a spectrum and other meaningful information. These aspects and their automation potential is discussed in the following sections in detail.

4.4 Automation of acquisition

The acquisition contains all of the mechanical tasks of obtaining an NMR spectrum, and automating these is especially important in high-throughput environments. Automated acquisition can also help non-expert users in routine use, as it might mask some of the complex details of acquiring a high-quality NMR spectrum.

4.4.1 Sample handling

Changing the sample in the magnet is traditionally performed manually, with the help of pressurized air to move the sample in and out of the magnet. Automating the sample change is a relatively simple mechanical task for modern robotics, and very beneficial if a large batch of samples with short acquisition times are examined. However, automated sample changing is really only beneficial if the other steps involved in the acquisition process can also be reliably automated.

Modern sample changers can contain up to several hundred samples and have other helping features, such as temperature control and bar-code detection, mostly driven by the requirements of metabolomic analysis [213]. Even with a high-end sample changer, the preparation of samples can also be a limiting factor, but with the help of liquid handling robots constant throughput of nearly 200 samples / day has been demonstrated [214], while even faster throughput is possible using flow-injection probes [215, 216].

4.4.2 Tuning

Tuning involves adjusting the *tune* and *match* capacitors to maximize the transmission of the RF signal to and from the sample. This process is also quite simple to automate: the algorithm of sequential tune/match adjustments based on reflected power to find correct tuning is not very complicated [16], and the adjustments to the capacitors can be done electronically with servo motors or by other means. Making this process easy, robust and reliable for different types of probes (broadband probes usually require large changes done with capacitor sticks) and challenging samples (high salt content can affect the tuning quite dramatically) is of course not entirely trivial, and it is still an optional feature in most spectrometers. However, as mentioned in Chapter 1, adjusting tuning for every sample might not even be required, as similar samples probably have comparable tuning characteristics.

4.4.3 Shimming

Automatic shimming using the lock signal can be done by modeling the manual iterative shimming process of maximizing the lock signal with sequential adjustment of different shims. This simple method has been implemented in Varian/Agilent and Bruker systems, but it is slow and prone to local minima, just like manual shimming [19, 218]. Some improvement can be achieved by using the FID area and Simplex minimization to adjust several shims together [219, 220, 221], but a modern and much faster technique is *gradient shimming*, which is based on the use of PFG to encode the spatial location of signals arising from different parts of the sample [222, 223, 224]. This can then be used to map the magnetic field inhomogeneity in the sample, and make required corrections to the shims.

Gradient shimming originated from MRI, and is closely related to ele-

mentary imaging techniques. In the basic experiment a gradient-echo is recorded with z-axis PFG pulse given during acquisition, which encodes the z-axis spatial location to the frequency of the signal, broadening the signal into a wide frequency range and resulting in an 1D image of the sample [222, 223, 217]. While the 1D profile already tells something about field homogeneity, the field can be mapped nicely by measuring two gradient-echo experiments with different delays: the local inhomogeneities generate differing phase drifts during the delays, and combining the results of the two experiments a field map can be constructed from the phase differences (Figure 4.2). The inhomogeneities can then be corrected by adjusting several shims simultaneously, whose individual accurate effects in the field can be mapped beforehand by a similar experiment. Gradient shimming requires that only one dominant signal is present in order to measure the profile well, and it is usually performed using the deuterium channel [224].

As the process is repeated iteratively a few times, the magnetic field homogeneity can be optimizer rapidly. Gradient shimming is usually performed in the z-axis, but if the spectrometer and probe support x and y-axis PFG, the phase encoding can be done in these dimensions as well. Gradient shimming accomplishes a rapid automatic shimming, which can be done even in seconds in optimal circumstances. This allows rapid throughput compared to lock or FID based shimming, which require several minutes or tens of minutes to complete [217].

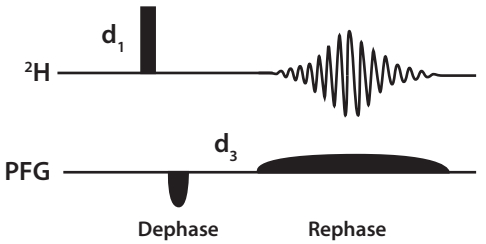

Figure 4.2: The basic gradient-echo pulse sequence used in gradient shimming. Two experiments with differing d_3 values are measured, and the resulting phase drifts are used to construct the field map [217].

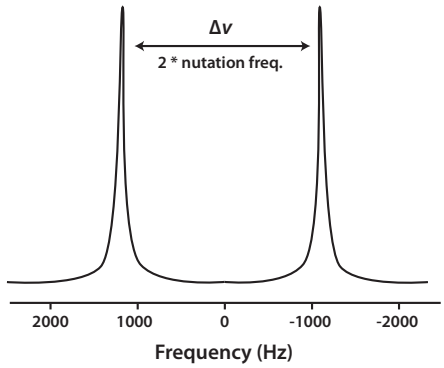

FIGURE 4.3: An illustrative nutation spectrum, based on the work of Otting and Wu[225]. Note that the figure presents a spectrum processed in absolute value mode, so the signals are not in antiphase formation.

4.4.4 Pulse width calibration

The traditional type of pulse calibration based on incrementing pulse lengths could of course be automated, but it would require a set of spectra analyzed with signal detection, phasing and so on. A much simpler method is the nutation experiment originating from MRI instruments [225]. In this experiment, no pulse is given before acquisition, but during the acquisition small pulses are transmitted within dwell time, the time in between quantified data points. This makes the magnetization rotate or *nutate* around the magnetic field generated by the RF radiation, as discussed in Section 1.1.3. For example, RF pulses in the x-axis will produce rotation in the z-y plane, with the speed depending on the RF power and the fraction of time the transmitter is on (the *duty cycle*). The frequency of the rotation is called the *nutation frequency*.

The observed spectrum contains two signals with the nutation frequency, located at equal distance from the transmitter, as shown on Figure 4.3. The signal is mirrored because magnetization is detected only in y-axis, rendering it impossible to detect the sign of the rotation by quadrature detection, leading to an antiphase doublet after performing DFT. The split between these signals is twice the nutation frequency, which is the frequency of full 360° rotations of the magnetization. The reciprocal of

the nutation frequency is the length of a 360° pulse, so taking into account the duty cycle, the RF power and length of the 90° pulse can then be derived easily.

This kind of experiment can be nicely automated, and requires only the frequency difference of two signals and one spectrum. However for in situations with a low signal-to-noise ratio, such as ^{13}C pulse calibration, performing the experiment might be difficult.

4.4.5 NMR experiments

The automation or batch acquisition of various NMR experiments is the easiest thing to automate within acquisition, as the spectrometer is already computer controlled. After the previous steps of calibrating and optimizing the acquisition equipment, deducing required acquisition parameters, such as spectral windows or mixing times, is quite straightforward: usually a batch of similar samples is being measured, and the characteristics of the samples are already well established or a representative sample can be used to derive reasonable parameters for all experiments. The automation of acquisition is then achieved by simply implementing control software, which supports an experiment queue, control of the sample changer and saving of spectra. These are all standard features of modern spectrometer control software, such as IconNMR (Bruker Biospin).

4.5 Automation of processing

The processing of NMR spectra involves transforming raw digitized time domain FID(s) into a frequency domain spectrum using Fourier transform with some pre- and postprocessing steps, as discussed in Chapter 1. In contrast to acquisition, processing involves only computation, and can be easily automated and customized without any additional hardware. While various NMR experiments require a diverse set of processing parameters and steps, the processing pipeline and related parameters can usually be decided quite easily in advance for a set of spectra, and many values can be derived from the acquisition parameters and the type of experiment.

4.5.1 Apodization

Window functions can be chosen according to many different objectives, but automating this process is usually quite straightforward as the optimal type of window function is usually known or can be decided beforehand, as it mostly depends on the experiment at hand. Some parameters of the function might require tweaking, but in many cases they can be determined or estimated from the acquisition parameters, such as the length of the FID or the approximate T_2.

For basic quantitative spectra, using only exponential decay (aka LB) window function is recommended, as it resembles the natural exponential decay and won't change the proportional intensities of signals depending on relaxation rate. The maximum signal-to-noise ratio can be reached when the window matches the signal envelope (*matched window* or *matched filter*) [16, 24, 226]. In most experiments this is achieved simply by matching the exponential function to the decay rate of the signal(s), and the LB window function is commonly described in Hz corresponding to the linewidth of a signal decaying identically with the function (and also describes the additional broadening it imposes). Thus the automatic application of an "optimal" LB window requires only an approximate estimation of the transverse relaxation rates, as the T_2 coefficient corresponds to the linewidth (at half height) by [1]:

$$linewidth = \frac{1}{\pi T_2} \tag{4.1}$$

4.5.2 Zero filling and linear prediction

Automating zero filling is trivial, as zero fill up to a certain amount of data points can be easily done and requires no further parameters or decisions. Linear prediction is more complex, but for example reasonable parameters to predict twice the increments in indirect dimension of a HSQC experiment can be deducted easily from the acquisition parameters.

4.5.3 Phase correction

Phase correction is arguably the most difficult aspect of automating NMR spectrum processing, and one which is still performed manually on a regular basis. While absolute value and power spectra circumvent this problem, they have significant drawbacks in resolution and integration as outlined in Chapter 1. Optimal phase correction depends on many experimental imperfections, and the estimation of exact parameters beforehand is practically impossible. Furthermore, there is no straightforward definition of a *perfect* phase correction, which works well for all real-world spectra. Human operators tend to balance several factors, and can utilize prior knowledge of the sample and signals to interpret the baseline and decide which parts of the spectrum are important.

A spectrum with good phase can be characterized with the following properties, although they can't be considered as a definitive description [38, 227]:

- **Narrow signals**: Signals consist purely of the absorptive component, with zero dispersive characteristic.
- **Symmetrical signals**: The signals are symmetric with the baseline being at the same level on both sides of the signal.
- **Flat baseline**: related to the previous characteristics, the baseline should be straight with no baseline roll or discontinuities near the signals.

In good quality spectra with distinct signals and high SNR these features are quite clear-cut. In practice, many problems can arise: overlap and complex coupling can create unsymmetrical signals, especially with complex substances and mixtures. Broad signals can be difficult to separate from

baseline, and baseline errors complicate the evaluation of the baseline itself. Some signals might not even be possible to phase properly by linear phase correction, such as water suppression residual or other artifacts.

4.5.3.1 Autophasing algorithms

Many algorithms have been suggested for performing autophasing during the past years. As early as 1969 Ernst suggested finding the correct phase by using maximum integral in the absorptive (real) part and zero integral in the dispersive (imaginary) part [228]. Since then, several other methods based on various desired and expected properties of NMR spectra have been introduced, such as maximizing points recognizable as baseline [229], fitting signals to Lorentzian lineshapes [230], signal symmetry [231] and baseline continuity [232]. Others depend on the mathematical relationship between real and imaginary components, such as calculating phase errors from integrals of absorptive and dispersive spectra [233] or creating dispersion vs. absorption plots [234].

All of these methods can work well, but expect and require certain features, for instance a particular line shape or singlet signals [230, 231], signals present in both ends of the spectra [232, 233] or a stable baseline [229]. Unfortunately, the samples which would benefit most from automatic phasing are those which are run in large batches with the aim of quantitative or qualitative analysis, and this usually means complex mixtures rarely meeting these requirements.

Recent algorithms are more robust and make fewer assumptions, perhaps because the use of autophasing has shifted from aid in spectrum processing to automation in batch processing of complex spectra. Modern algorithms usually combine several sophisticated methods and are more computationally intensive, such as ACME, which is based on minimization of entropy and negative points [235]. In 2009, de Brouwer evaluated it and other prior algorithms, and ultimately constructed a more accurate method for the analysis of synthetic polymers [38]. The algorithm consists of minimizing the total integral area and negative points, with baseline correction in between. De Brouwer was able to match the performance of expert human operators, producing errors less than 0.1 % with the used polymer samples. A modified version of the algorithm is also used in the automatic phase correction in SimpeleNMR, which was built and utilized in **Paper III** and **Paper IV** of this thesis.

An even more complex algorithm based on two step process of coarse and fine optimization was demonstrated by Bao et al. in 2013 [227]. In the first step, a sophisticated baseline recognition and continuity optimization is performed to achieve rough phasing, while in the second step negative data points are minimized after baseline correction of the initial result. In contrast to the method presented by Chen and de Brouwer [38, 235], the method detects distorted and negative signals after the first step and is thus able to phase spectra with mixed signals present. This is useful for example when phasing spectra with a mangled water suppression residual signal, which is easy for human operators to ignore, but can cause trouble with methods assuming an all-positive signal.

Also algorithms for automatic phasing of 2D spectra have been proposed, such as one based on whitening or the minimization of visible points in the spectrum by Balacco. et al. [236]. This kind of process is robust and can yield nice results for visual inspection of spectra, but might be inadequate when exact quantification is desired and baseline errors are present. Fortunately the phasing errors in the indirect dimensions do not suffer from most of the experimental limitations and effects responsible for the phasing problems in 1D spectra, and as such phase errors are much more controllable and deterministic.

The fact that baseline correction algorithms are still recently being published tells a lot about the difficulty of this seemingly simple problem. Small deviations of 2° degrees in the phasing angles can produce 2.5 % deviations in integration results [38], so accurate phasing is important when small differences are being measured. Also, the intimate relationship of phase correction and baseline creates additional complexity, and the proper handling of both is crucial for accurate quantification. This is evident with the integration of baseline correction or recognition in the recent phasing algorithms, as noted above.

4.5.4 Baseline correction

The simplest baseline correction is the linear (or drift) correction, in which the baseline error B is assumed to follow the linear equation:

$$B = ax_i + b$$

where a and b are coefficients, and x_i is the index of the data point. This kind of flaw is usually a result of errors in the first data points of the FID, and it is quite easily removed automatically by fitting a linear equation to the baseline and subtracting it. The spectrum edges can usually be used safely for the simple linear regression, as they should not (and rarely do) contain any signals. For example ImatraNMR uses 2.5 % of spectrum points from each spectrum end, computes the average values and fits the linear equation [237].

4.5.4.1 Generic baseline correction

The method just described is very simple and unable to correct many common baseline errors. However, it is an example of a more generic three-stage process to correct baseline errors, introduced already in early work by Pearson [238]:

1. **Establishing baseline:** Finding which parts of the spectrum can be considered baseline or used to model it.
2. **Modelling baseline:** Creating a model describing the baseline error, for example fitting a spline function to certain points in the spectrum.
3. **Removing baseline error:** The model is subtracted from the spectrum, producing a baseline corrected spectrum.

The first stage can be done quite easily by hand, which is indeed an option in many processing software packages. After the selection, a suitable mathematical function can be used to model the chosen points. While this is feasible for a small number of spectra, it is of course not very suitable for automated processing. The implementation of the first two stages are the core of automatic baseline correction algorithms, and they are discussed in detail in the next two sections, while the third stage is merely a trivial subtraction.

4.5.4.2 Recognizing baseline points

There are myriad ways to implement the classification of spectrum data points. As an example, an algorithm described by Golotvin et. al. [239] is quite simple and consists of two phases:

1. **Determining the noise level.** The spectrum is divided into k regions (Golotvin used $k = 32$), the standard deviation of which is calculated. The smallest result is selected and then multiplied by a constant value (Golotvin used a factor of 2-4).

2. **Classification.** For each point, a window of N points is centered on the point and the difference of the largest and smallest value is calculated. If this value is smaller than the noise level determined in previous phase, and the surrounding two points also satisfy this rule, the point is considered to be part of the baseline.

This remarkably simple method doesn't filter the spectrum before classification, while several other approaches utilize a derivative of the spectrum to remove constant errors and de-emphasize slow changes in spectra, such as broad signals or baseline errors. The signals can then be detected by iteratively determining a threshold level [240, 241]. A simple numerical calculation of derivatives increases noise, but this can be improved significantly by utilizing *Continuous Wavelet Transform* (CWT), as demonstrated by Cobas et al. [242]. The wavelet transform yields less noisy derivatives than traditional smoothing methods such as Savitzky–Golay [243, 244]. This approach was recently developed further by Bao et al. by combining it to the sliding window approach of Golotvin in order to improve detection broad signals. These were then taken into account through fitting a model to the tails of the signals [239, 245]. Quite similarly to computing derivatives and smoothing, broad signals can be removed just by utilizing a high pass filter on the spectrum [246].

The classification can be made also by probabilistic approaches, such as estimating the standard deviation of noise and signals and calculating the probability of data points containing signal based on the deviation from baseline, while iteratively adjusting the baseline model [247]. Another statistical style approach is calculating the standard deviations of sliding windows and observing their distributions, reasoning that most windows should contain only noise [248].

All of the methods described should work well for high-quality spectra containing sharp, symmetrical and distinct signals. The differences are in implementation complexity, speed, parameters and especially in the made assumptions: some techniques for example don't handle negative signals [249], assume that most of the spectra doesn't contain signal [248] or

classify broad signals or tails easily as baseline [239, 242]. However, the desired distinction between broad signals and baseline errors is ultimately solvable only based on prior knowledge about the samples/spectra, as the effects in the spectra themselves can be practically indistinguishable. As such, there is no best method suitable for any type of spectra, at least not without some adjustments on the parameters of the method. Most of the recent methods are aimed for metabolic spectra [245, 246, 248], while others are more generic and are proposed for use in other analytical techniques as well, such as Raman spectroscopy and Chromatography [249].

4.5.4.3 Modelling the baseline

After the non-signal points have been classified as a binary mask or a vector of weights, a model describing the baseline can be built. In many cases, spline-type functions (polynomials, cubic splines, Bernstein polynomials) have been utilized [238, 240, 241, 246, 250]. However, linear interpolation of the removed parts and smoothing using Gaussian weighted [239] or plain mean function [248] has also worked well, and can correct much more diverse shapes easily. Whittaker smoother [242, 251] has the same advantage and can be used directly without filling the missing signal parts, and also offers a user-defined smoothness parameter.

All of these methods aim at separating the high frequency Gaussian noise present in the data points from the low-frequency errors in the baseline. Therefore even simply applying a high-pass digital filter can work as a baseline correction, such as a "model-free" approach based on calculating the medians for sections by Friedrichs et al. [252]. However, this kind of approach only works with sharp signals and no congested areas, as large deviations from baseline quickly affect the calculated median.

4.5.5 The origin of phase and baseline errors and corrections in the time domain

The phase and baseline errors in 1D and in directly detected dimension are principally a result of two experimental limitations [91, 92]:

1. **Corruption of initial data points:** The first few data points have errors arising from the transient response of the used analog filters and amplifiers.
2. **Delayed sampling**: A delay is needed between the last pulse and the start of acquisition in all experiments which end in a pulse, as the electronics need a few microseconds for recovery in order to stabilize and stop the bleeding of the strong RF pulse to acquisition. Delays and phase shifts are also caused by applied analog filters.

Most NMR pulse sequences end with an RF-pulse, and the delayed start of sampling leads to evolution of the magnetization during the delay. This causes a linear, frequency dependent phase error [255], which is correctable via 1st order phase correction. However, the phase correction causes *baseline roll*, a sinusoidal type oscillation of the baseline dependent on the magnitude of the 1st order phase correction [16, 255]. The mechanism is related to aliasing/folding of the long dispersive tails of the signals to the spectrum area as the phase correction is only correct in the center of each signal [256]. Secondly, analog filters are needed to eliminate high frequency components in order to prevent aliasing in ADC sampling [23], and these filters also introduce a small delay to the signal which is inversely proportional to the width of the filter bandwidth.

The combination of the aforementioned effects was examined by Hoult et al. in 1983 along with a counteractive strategy: by carefully adjusting the delays between the last pulse, receiver gate opening and the start of sampling, the needed 1st order phase correction and errors in the first few data points could be minimized [253, 254]. The principal idea is to start sampling when the delay between the center of the last pulse is equivalent to the theoretical delay introduced by the analog filter (Figure 4.4); in this point of time, the phase error introduced by delayed sampling is compensated by the analog filter. As the signals are (nearly) in phase, the need for 1st order phase correction is minimized and baseline errors are avoided.

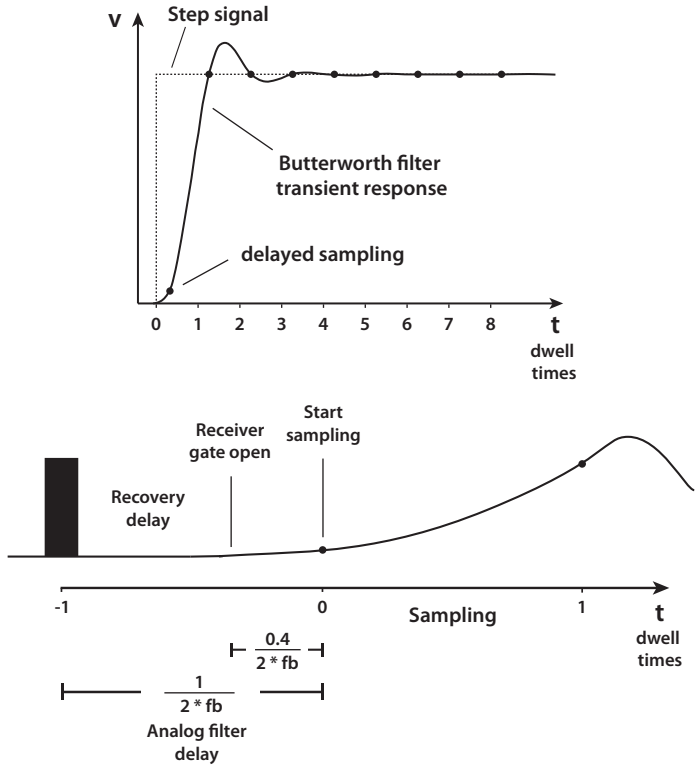

FIGURE 4.4: The theoretical transient response of an analog Butterworth filter (top) and the acquisition scheme (bottom) suggested by Hoult et al. [253, 254]. The transient response of the used analog filter will corrupt the first few data points, but with clever timing (depending on the filter), the response nearly matches the ideal response at the sampled time points, except for the first data point. The sampling is timed appropriately so that the delay of the analog filter compensates for the delayed sampling, and the signals are in phase at $t = 0$. The suggested delays for minimizing both the phase and transient errors for analog Butterworth filters are marked at the lower part, in which fb is the filter bandwidth.

Furthermore, the gating of the receiver is timed so that the transient response of the analog filter has minimal errors in the sampled time points, except for the first data point. This results in only constant error in the baseline, which is easily corrected (the first data point affects all frequencies equally in DFT, see Figure 1.9).

These optimizations are widely implemented, for example Varian/Agilent VNMR/VNMRJ includes commands `hoult` and `calfa` to adjust the acquisition delays (`alfa`, `rof2`) correctly, with the latter utilizing a test spectrum to account for experimental variation [257]. Similarly, Bruker Topspin calculates the pre-scan delay `DE` automatically when digital filtering isn't used [88].

Oversampling can also be used to minimize phase and baseline errors [91]: a higher sampling rate allows opening of the anti-aliasing filters, resulting in a much shorter transient response of the analog filter. If a high oversampling factor is used, the transient response affects only the first data point after the conversion (decimation) to the base sampling rate. Again, the subsequent constant baseline error can be easily rectified. If high enough oversampling is combined with digital low-pass filtering, the transient response is essentially removed. This approach has been employed in the more recent Agilent/Varian VNMRS and DD2 consoles, which utilize a very high oversampling rate (80 MHz) and time-corrected digital filters [23, 26, 91]. The `setrc` command can still be used to adjust the start of sampling to minimize linear phase errors and set linear prediction to compute the first data points if needed.

The errors stemming from the first few data points can also be corrected by reconstructing them: by aligning the sampling points so that the time point $t = 0$ is a multiple of the dwell time, the missing/erroneous data points can be reconstructed by *backward linear prediction* or other methods [37, 258, 259]. This is similar to extending the data by linear prediction, and the "perfect" sampling removes the phase and baseline errors. One example of this approach is Bruker `baseopt` digitizer mode introduced with Topspin 2.0 in 2006, which uses the pre-scan delay `DE` to match the delay between start of sampling and $t = 0$ to the dwell time of the oversampled data points and reconstructs the missing points in the spectrometer console ([88] and email correspondence with Bruker). After this, regular digital filtering is applied, and as a result only minimal 1st order baseline correction is needed and a very flat baseline is obtained.

Even in the ideal case of starting sampling at $t = 0$, the properties of DFT lead to a constant offset in the baseline [256, 260]. This can be understood by considering that each discrete data point represents the integral of ± 0.5 dwell times surrounding the point, while the point in $t = 0$ doesn't have any integral in $t < 0$, as there is no signal. In the ideal case of starting sampling at $t = 0$, this can be corrected by multiplying the first data point with 0.5 without any adverse effects [256]. This situation can be nearly achieved in indirect dimensions in multidimensional spectra, in which the electronic limitations don't restrict the sampling much, as only timings between pulses are adjusted. The finite length of pulses still account for some evolution (approximately $2t_{90}/\pi$ for a 90° pulse of length t_{90} [258]), so even for a zero delay between two 90° pulses such as in NOESY, there is some evolution. However, in real world use, adjusting the 1st data point can still effectively reduce t_1 ridges in the 2D spectra.

While the time-domain correction of phase and baseline errors is very beneficial, it can't fix errors arising from other sources, such as non-linear phase errors of analog filters or phase errors rising from off-resonance effects. This makes small 1st order phase corrections still useful in many situations. Undesired baseline shape can result from other imperfections as well, such as unwanted broad resonances, justifying also the use of advanced baseline correction algorithms.

4.6 Automating data extraction and statistical analysis

After an optimally processed spectrum is available, the signals in the spectrum must be quantified in order to extract concentrations or other meaningful information for further analysis. Different methods can be chosen depending on the type of information and amount of spectra in question: checking the existence of a single signal from a few spectra can easily be done even visually, but statistical analysis of all signals present in tens or hundreds of spectra requires more advanced planning and tools.

4.6.1 Integration

The basic method for quantifying a signal is simple numerical integration: the area of a signal is calculated numerically using the discrete data points of the spectrum with extended trapezoidal rule or other numerical integration algorithms in 1D or 2D [22]. The area corresponds to the intensity of the signal and therefore in the case of quantitative NMR, the relative concentration.

In the optimal case of high quality spectra with distinct signals and a flat baseline, integration is very straightforward, and data from a few spectra and signals can be produced easily with practically any NMR processing program. For larger sample sets, a predefined set of integration areas can be saved and loaded, or scripting can be used to automate the integration of spectra. However, special tools designed for handling a large number of spectra are usually beneficial, as discussed later in Section 4.6.6.

The principal limitation of basic integration is signal overlap, as it can't separate the contributions of the individual signals. This can sometimes be circumvented by using only well-separated "characteristic" signals with no overlap, but with complex mixtures finding these signals is unlikely. Using 2D methods or observing other type of nuclei (such as ^{13}C) can provide more resolution, but obtaining quantitative spectra is usually more difficult for several reasons, as discussed in Chapter 3.

4.6.1.1 Aligning spectra

Automated integration of predefined areas assumes that chemical shifts don't change across different spectra, and signals with the same chemical shift correspond to the same chemical structures. Unfortunately, a lot of factors can affect the chemical shift (temperature, pH, solvent and other

intermolecular interactions), and a small sample to sample to variation usually exists. This can be minimized by standardizing sample preparation and stabilizing the measurement conditions, but still some form of spectral alignment is always needed.

The simplest approach is to align the spectra using a reference signal: one signal is chosen and the chemical shift scale is set accordingly to match a given value in each spectrum. This kind of alignment is required for accurate chemical shift measurements, and while it corrects changes uniformly affecting all chemical shifts, it can't correct relative deviations of chemical shifts. This is a much harder problem, and several approaches for more advanced alignment have been suggested. Almost all of the methods still follow the basic two-step principle:

1. **Splitting:** The spectrum is divided into segments, either uniform or depending on spectral features.
2. **Shifting/scaling:** The segments are aligned to a reference spectrum by shifting and/or scaling (stretch/compression). The possibly formed gaps between segments are filled in some way.

The first methods following these basic principles were based on *Correlation Optimised Warping* (COW) or *Dynamic Time Warping* (DTW) applied initially for chromatograms and having roots in speech recognition [261, 262, 263]. In these methods the spectrum is divided into equal segments, which are then scaled to obtain an optimal correlation with the reference spectrum, while the segments remain connected, so no gaps are formed. The resulting warping is quite obviously suboptimal for quantitative NMR spectra, even with conservative constraints in min/max scaling. Subsequently, methods aligning separate segments based on a genetic algorithm or beam search were developed, introducing the practice of using spectral features to establish the segment borders [264, 265]. Also methods simplifying the alignment process by extracting locations of the signals were introduced, but these deviate from the basic algorithm by handling individual signals instead of slices of spectra [266, 267].

All of the above methods were quite complex and slow, and didn't gain widespread use. This was addressed by Wong et al. in 2005 by utilizing cross correlation calculated speedily with FFT to align the segments, making the alignment of large datasets feasible [268]. The cross correlation

measures the match between two signals by multiplying them point-wise and calculating the sum for each offset (or *lag*). The offset with the maximum sum then corresponds to the best match. This is closely related to convolution and can be calculated very efficiently utilizing FFT and inverse FFT, a common technique in digital filtering [24]. Subsequently improved methods were based on this principle, with one of the most prominent being *i*coshift [269, 270], which aligns several segments in spectra simultaneously, provided that the slicing is done identically in every spectrum and the spectra can fit in the available memory. More recently the *Cluster-based Peak Alignment* (CluPA) technique combined signal detection and hierarchical clustering to the cross correlation alignment, claiming improved accuracy at a somewhat slower speed [271]. Both *i*coshift and CluPA implementations are provided with the source code (in MATLAB and R, respectively). When aligning large datasets of several hundred hi-resolution spectra, these methods can reduce computation times from minutes or even hours to seconds compared to classic COW methods.

ImatraNMR supports the basic alignment on a single signal, and the alignment of separate slices of spectrum, but no totally automatic slice alignment method has been implemented yet. However, very flexible *dynamic integration areas* can be used, which are adjusted depending on signals present in each spectrum [237].

4.6.1.2 Binning

Another solution to small deviations in chemical shift is *binning* (or *bucketing*): reducing the resolution by combining several data points into *bins*, so small variations of chemical shifts inside the bin do not matter. This is especially useful when performing statistical analysis on complex mixtures and no particular signals are targeted (e.g. in metabolomics and chemometry). The signal areas can also be even defined by hand utilizing prior knowledge.

In the basic form of this method, the spectrum area of interest is divided into uniformly equidistant sections of typically 0.04 ppm in ^{1}H spectra, which are then integrated [272, 273]. Solvent signals or signal free areas can be neglected or filtered afterward, but the end result is a reduced dataset of a few hundred bins instead of several thousand data points. This *dimensionality reduction* is an another desired outcome: multiple data points describe a single signal in NMR spectrum, while the intensity information of the signal could be represented ultimately by only a single value. The

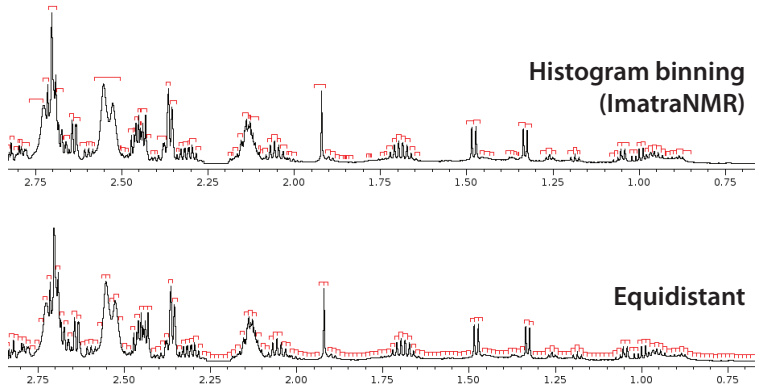

FIGURE 4.5: Two examples of binning. In the lower spectrum traditional equidistant binning has been used, while the upper histogram demonstrates binning implemented in ImatraNMR, based on grouping the detected signals.

correlated data points making up the signal are in this sense redundant information, which can be dealt with by fusing the data points if no overlap is present. However, as simple equidistant binning doesn't consider the shape of the spectra, it can just as easily combine two separate signals or split a single signal depending on how the bin borders happen to align.

To improve the equidistant binning method, several algorithms have been published quite recently trying to match the bins more appropriately to the signals present. *Adaptive binning* creates a combined spectrum from maximum data point values and uses an undecimated wavelet transform to model the spectrum and decide the bin borders, but this is quite complex, and requires tuning the wavelet transform in order to achieve the desired level of binning [274]. *Adaptive Intelligent Binning* (AI-Binning) recursively splits each bin into two bins and decides if the created bins are more appropriate than the original one using a cost function [275]. This can result in quite optimal and automatic binning, but the presented algorithm uses brute force search when dividing the bins, and as such doesn't scale well to large datasets of high-resolution spectra [276].

Binning can also be based on found signals: *Dynamic adaptive binning* smooths the spectra and identifies signals by local maxima, and then tries to choose bins which contain only one signal in each spectrum [276]. The *histogram binning* present in the ImatraNMR software (**Paper II**) also forms bins from detected signals, classifying them into bins based on proximity tolerance, and choosing the bin area based on the original signals (Figure 4.5). The nice side-effect of this is that no bins are formed in areas where no signals are detected, so it is easily usable in 2D spectra as well without further filtering, as demonstrated in **Paper III**. The implementation details can be found in the ImatraNMR user guide [237], and the improved performance over equidistant binning is demonstrated in **Paper IV** (Figure 4.6).

Some simpler algorithms improving the equidistant binning have also been devised, such as using Gaussian weighted integrals instead of strict rectangular bins [277] or optimizing the basic equidistant bins by moving the bin borders to low intensity regions [278]. Neither of these methods allocate more bins to crowded areas and can yield areas with only noise present, but still they avoid coarse splitting of signals and are quite simple to implement.

All of the above methods have been shown to improve results over equidistant binning in subsequent statistical analysis, but choosing the optimal method probably depends on the features of the spectra at hand. For automation, the algorithm should be robust and require no careful optimization of parameters to work well, and with large datasets, speed can be critical. Most likely choosing the algorithm is still dictated by the used software, as usually only one algorithm is implemented in addition to the basic equidistant binning. The real-world differences between the binning algorithms are probably quite small, as even equidistant binning works surprisingly well. Slightly suboptimal binning is also usually not a major concern, as the analysis is done by multivariate methods, which assume that the observed values can be affected by several underlying variables. Furthermore, signals might overlap anyway and even optimal binning can't resolve that.

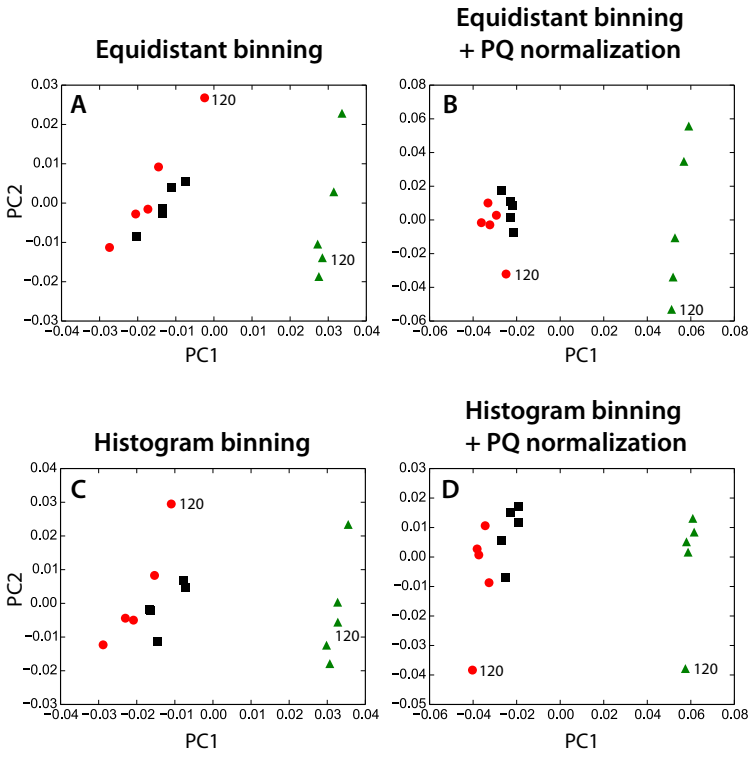

FIGURE 4.6: The significance of binning and normalization procedures can be easily seen in the four different principal component analyzes (see Section 4.6.5) with different binning and normalization procedures performed in **Paper IV**. The JRES spectra was acquired from three different plant lines (denoted by the circle, square and triangle markers), which were exposed to ozone for varying amounts of time (for details, see **Paper IV**). Both regular equidistant and ImatraNMR histogram binning produce quite similar results (A and C), but when combined with Probabilistic Quotient Normalization (PQN), histogram binning provides clearly superior results (B and D): the plant lines are separated cleanly with PC1, while the lines sensitive to ozone are clearly separated with PC2. The two 120 min exposures for ozone sensitive species are marked in each plot.

4.6.2 Normalization

Comparing even relative concentration differences in a set of spectra requires scaling the signals properly, since various factors can affect the absolute values produced by the spectrometer (as discussed earlier in Section 3.2). The easiest way is to add an internal standard, but basically any signal with an (approximately) constant concentration relative to the sample material can be used to normalize the concentrations to some arbitrary scale.

In some cases, an internal standard is difficult to use, because the measurement of the sample material is inaccurate: the amount of uninteresting or undetected compounds, such as residual water, can affect the weighting of the sample and result in errors of the concentration of the standard. Urine samples are an extreme example of this; using a regular internal standard would accurately gauge the dilution of the sample, but the constituents and their proportions are what is probably more interesting.

If an internal standard is not present or is not meaningful, there are several alternate strategies which can be used to normalize the integral/bin values, such as:

- **Combined signal/integral normalization**: Assumes that all material is detected, and that the combined intensity of all signals is thus correlated to the mass/concentration of the sample material. Values can then be normalized to the combined intensity and measured as fractions of the total signal.
- **"Stable compound"**: Assumes that some feature, such as a specific compound or solvent residue, is constant. This is then used as a standard, similarly to a proper internal standard.
- **Distribution**: Assumes that some distribution of features in the spectra should be constant across samples. This is then used to derive normalization factors.

All of the above make some sort of assumptions in order to normalize the data, which can be more or less justified depending on the case. A combined signal can be correlated very well to the mass of the sample material when similar molecules form the samples, but it can be strongly affected

by changes in composition: different compounds can have different proportions of the observed type of nuclei, or even lack it completely. Further, calculating the total signal is very sensitive to baseline errors.

4.6.2.1 Probabilistic Quotient Normalization

More advanced normalization algorithms are usually based on the analysis of distributions derived from sample data. For example, *Probabilistic Quotient Normalization* (PQN) introduced by Dieterle et al. in 2006, assumes that the samples have similar composition, and that only a minority of the signals contain significant deviations [279]. This assumption is used to devise a distribution of intensity ratios between the spectrum and reference, from which a probable correction factor is devised. The algorithm can be summarized by the following steps:

1. **Construction of the reference spectrum:** This can be a single selected representative spectrum or a median spectrum, computed by taking the median value of each data point (intensity at each chemical shift or bin). All spectra are normalized against this reference.
2. **Integral normalization**: The data values are normalized to the total detected signal, which provides a coarse initial normalization.
3. **Signal ratios**: The ratio of each data point to the reference spectrum is calculated, and if the signals match in intensity the ratios are close to 1.0. If the signal intensities are predominantly smaller, the ratios are mostly <1.0 (less sample material than reference) and correspondingly, larger signals produce ratios > 1.0 (more sample material).
4. **Probable ratio and normalization**: The distribution of ratios is used to devise a correction factor. Most of the data points are assumed to represent the same concentration as the reference, so the distribution of the ratios should be narrow and concentrate around particular value. The median of this distribution is considered to be the most likely ratio, representing the concentration relation between the current and reference sample. This value is then used for normalization.

In essence, PQN tries to scale the spectrum or bin values in a way that most of them match the reference spectrum in intensity, quite close to what

visual matching of the spectra might produce. This kind of normalization requires that there are a large number of signals and that most of them indeed don't vary much between samples. As such it seems to be well suited for biofluid samples such as urine or blood serum [279, 280]. It was also found to be efficient in normalizing the spectra of plant extracts in **Paper IV** (Figure 4.6).

There are many other methods based on distributions [281], such as *quantile normalization*, which assumes (and forces) the intensity distribution to be the same for each sample [282]. Many of these can easily be automated and implemented in numerical software such as Matlab or Python/NumPy, and some of them are readily available in analysis software. ImatraNMR has the PQN built in for both spectra and integral/binning results.

4.6.3 Line shape fitting ("deconvolution")

Signals in NMR spectra should follow the *Lorentzian* line shape, which can be derived from the fact that the signals are exponentially decaying sinusoidal waves processed with the Fourier transform. The signals can also follow the more familiar Gaussian line shape, if the Gaussian window function is used, or a combination of these basic line shapes (*Voigt profile* [283]). In any case, the theoretical shape of NMR signals can be represented with mathematical functions, which can then be used to model signals in a real NMR spectrum, a process which is referred to as line shape fitting or "deconvolution" in NMR literature.

The main advantage of line shape fitting is that the components of the resulting model can be subsequently analyzed separately and used to extract the chemical shifts, intensities and line widths of overlapping signals, a process which would be impossible using the basic integration approach [284]. This kind of functionality is readily available in many processing programs such as PERCH (PERCH Solutions) or Mnova (Mestrelab research) [285, 286]. However, the exact line shape is affected by many factors (field inhomogeneity/drift, temperature, relaxation, molecular interactions) and the fitting represents a complex multidimensional optimization problem. Therefore, in addition to iterative fitting, manual adjustments are usually required to achieve good results.

Any manual involvement is of course undesirable in automated processing of large data sets, and some recent developments have tried to alleviate

this problem. Chenomx NMR Suite (Chenomx Inc.), aimed for metabolite analysis, includes a library of reference compound profiles consisting of chemical shifts and signal intensities, extracted from authentic standards in different magnetic field strengths [287, 288]. These profiles can be used to construct spectrum templates, which take into account a variety of aspects, such as linewidth and chemical shift deviations, caused by sample pH or other factors. This simplifies and improves fitting by grouping several resonances together, making it possible to quantify metabolites from complex overlapping patterns. Batch fitting multiple metabolites and files is possible, but Chenomx is still clearly aimed for "targeted profiling", accurate quantification of specific known metabolites.

A similar approach is provided by *Bayesian AuTomated Metabolite ANalyser* (BATMAN) [289, 290], a free software package based on the R statistics environment and programming language. The package is aimed for more automated processing, and seems to be capable of yielding comparable results with human operators using Chenomx [291]. For specific type of samples (e.g. blood plasma), fully automated processing schemes have been demonstrated by combining targeted profiling with advanced statistical tools such as Bayesian models, *Markov Chain Monte Carlo* (MCMC) and *Sequential Monte Carlo* (SMC) [214, 292, 293].

Another solution is provided by Mnova, which implements the *Global Spectral Deconvolution* (GSD) algorithm, providing automatic line shape fitting for the full spectrum area [286]. The resulting intensities and chemical shifts can then be exported and analyzed using other tools, such as with multivariate statistical methods. However, it seems that there is not much research data available on how well this method can handle complex real-world spectra, such as those encountered in metabolite analysis.

4.6.4 Time-domain analysis

Another approach for extracting individual signals is *time-domain detection*, in which DFT is not used, but instead the FID is modeled as a combination of exponentially damped sinusoids [294, 295]. This approach has many advantages: the baseline and phase correction related problems are irrelevant, as no spectrum is even formed, and all parameters such as decay rate (linewidth) are also extracted.

Deriving the model parameters for a FID containing hundreds of signals with noise is of course a very complex problem. The current methods are

based on advanced Bayesian models utilizing MCMC methods in addition of filtering the spectrum into smaller regions (or sub-FIDs), which can treated more simply. Nevertheless, the approach has been shown to work in real-world metabolic analysis, providing improved results over traditional analysis and equidistant binning [296]. Unfortunately the recent promising implementation of this method dubbed *Complete Reduction to Amplitude Frequency Table* (CRAFT) implemented in Agilent/Varian VNMRJ 4 is probably lost, as Agilent closed its NMR business in 2014, ending the development [30, 295].

4.6.5 Statistical methods

If a single or a few specific signals are targeted, their intensity can be plotted easily and simple tools, such as regular linear regression, can be utilized. This is suitable for simple mixtures and cases such as reaction monitoring. In many cases however, the problem is to find out which spectral features can be used to differentiate or cluster samples in a large set of complex spectra.

Every NMR spectrum usually consists of thousands of data points, and even after binning, several hundred values might define each observation. The statistical analysis of this kind of data is not trivial, and is most commonly achieved with *Principal Component Analysis* (PCA) and *PLS Discriminant Analysis* (PLS-DA) [192, 273, 297, 298], both of which are multivariate statistical methods, suitable for dealing the hundreds or thousands of collinear parameters derived from NMR spectra [299, 300, 301].

The aim of PCA is to find orthogonal *principal components*, which are chosen to explain the maximal variance in the dataset (Figure 4.7). The quantity of a component present in each sample (component *score*) can be used to project the dataset into fewer dimensions, hopefully resulting in a meaningful distribution or clustering of the samples. The chemical relevance of the components can be derived by examining the *loadings* of the components, which correspond to the original data points or bins, probably matching signal patterns of identifiable compounds. Because PCA treats all variables as equally important, changes in strong signals can easily dominate in the components, as relative changes in a strong signal can easily be larger in absolute terms than even the total disappearance of a small signal [299]. This can be overcome by normalizing the variables by *unit variance scaling*, with the risk of amplifying small signals and noise.

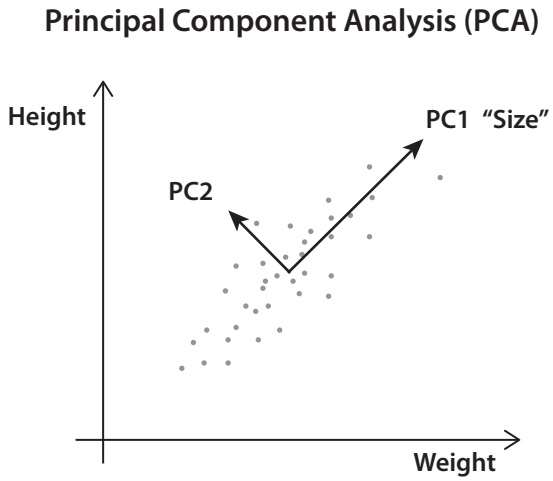

FIGURE 4.7: The basic principle of the Principal Component Analysis (PCA). In this simple demonstration two variables of height and weight are measured from a group of people. While they describe different characteristics, they are obviously correlated: taller people tend to be heavier and vice versa. PCA finds the axis explaining the most variance between samples, which is called the *first principal component* (PC1). The score on this component can then be used to describe each person in the dataset, and in this case the PC1 can quite intuitively reflect the general size of the person, while it doesn't describe the height or weight exactly. The second principal component is orthogonal to the first, and describes the rest of the variance. PCA thus produces two scores similarly to the original data set, but ordered by decreasing explanatory power. Usually the number of measured variables is much larger, and PCA is used to find few principal components describing most of the variance in the dataset, which might or might not represent meaningful characteristics of the original samples.

PCA is an *unsupervised* technique, meaning that no prior information about the data is used, and that the maximum variance might or might not provide an optimal clustering of the samples. The closely related technique PLS-DA can be used to incorporate sample class information in

order to improve the separation. PLS-DA is based on *Partial Least Squares projection to latent structures* (PLS) regression [301], in which every observation (sample) is associated with one or multiple *responses*. Similarly to PCA, both the observations and responses are projected onto scores, and a linear model which tries to predict the responses is built. PLS-DA is again just PLS with the responses set to dummy variables representing the classes [299]. The constructed projections can be plotted as with PCA, or the model can be used for prediction.

PCA and PLS-DA are widely implemented in statistical packages, and the use of PCA is demonstrated in **Paper II** and **Paper IV**.

4.6.6 Software tools for quantitative NMR

While regular processing software packages such as Mnova (Mestrelab research) include features aimed for quantitative NMR [286], and some commercial software designed for mixture analysis, such as AMIX (Bruker Biospin), Chenomx NMR Suite (Chenomx Inc. [287]) exist, majority of "classic" NMR processing software, such as TopSpin (Bruker Biospin), VNMRJ (Agilent/Varian) or SpinWorks (Kirk Marat [302]) are simply not designed from the point of view of mass processing and data-analysis.

Driven by the increased research interest in metabolomics and chemometry, several free tools for automated analysis have been released during the recent years. These tools are mostly designed for metabolomics use, and are primarily concerned with spectrum handling, data extraction and metabolite identification: rNMR enables the handling, viewing and quantification of 2D spectra [303], MetaboMiner offers peak picking and automated assignment in 2D [304], while other tools also include NMR processing capabilities [305, 306]. Also several public metabolite/spectrum databases have been created such as *BioMagResBank* (BMRB) and *Human Metabolome Database* (HMDB) [307, 308], and some tools offer integrated support for using them [309]. Surprisingly, even web-applications incorporating processing such as MetaboAnalyst and Bayesil for biofluids are offered [293, 310], and while these are free to use, they run and depend on resources offered by the original developer.

For more general use cases, raw processing tools with *Command-line interface* (CLI), such as NMRPipe [311] can be used for batch processing and preparation of spectra, albeit further data-analysis must then be performed with more universal data-analysis software such as MATLAB [312], the R

statistics package [210] or Python with Numpy/Scipy [27, 32]. Processing extensions to numerical analysis environments such as MVAPACK [313] for GNU Octave and NMRGlue [314] for NumPy/SciPy can combine these two aspects in single environment, but the use of these kind of low-level tools require scripting and/or programming knowledge and might be out of reach for many users. On the other hand, the open-source and low-level nature of these tools make it possible to build custom analysis pipelines, test new processing schemes and extend the provided functionality.

Likewise, the tools presented in this thesis, ImatraNMR (**Paper II**) and SimpleNMR (**Paper III**), are designed for more generic NMR analysis with the emphasis on batch processing, and as such require more expertise than traditional NMR software. While not as low-level, they can still be customized fairly easily and are able to process large datasets.

Most of the tools presented here can perform only some of the desired operations, and the full analysis stack from FIDs to visualization requires a combination of software with intermediate results stored in files. This allows a lot of customization and freedom, but in many cases it would be desirable to incorporate the full analysis pipeline in a single environment, with the ability to update the results whenever analysis parameters or data is changed. This kind of approach has been taken by Pathomx [306], in which the analysis pipeline is defined graphically using connected processing steps, with the results updated automatically after any adjustments. Full analysis from FIDs to statistical tools such as PCA is possible, and the code for each stage, most of which is based on Python/NumPy, can be customized freely within the program. While this work seems promising and many features are available, initial testing conducted by the author revealed many small defects and only partial implementation of some features. However, the open-source nature and heavy customization possibilities mean that the problems are usually solvable with programming expertise, and hopefully a more polished version will be released in the future.

Conclusions 5

"You ever notice that? Any time you see two groups of people who really hate each other, chances are good they're wearing different kind of hats. Keep an eye on that, it might be important."

--George Carlin

In the preceding chapters, many aspects and difficulties associated with quantitative NMR and automated analysis have been discussed, along with insights into the underlying reasons and principles. While NMR remains a quite complex method, the advances in NMR techniques and equipment during the last two decades have largely diminished the traditional weakness of NMR spectroscopy: sensitivity. This in conjunction with easy sample preparation and non-specificity has established the use of NMR in high-throughput settings and steadily fuelled the use of quantitative NMR, with over 700 publications released annually during the recent years [42, 216]. The appearance of easy-to-use benchtop NMR spectrometers will probably further help in reducing the barrier of using NMR due to high costs and maintenance fees, and make it a feasible technique in smaller companies and facilities as well.

In this thesis, several techniques and tools for quantitative NMR are introduced. In **Paper I**, the novel quantitative polarization transfer experiment Q-INEPT-CT is presented, boosting the sensitivity and/or reducing the time requirements of acquiring quantitative ^{13}C spectra. A 2D version of this technique is also presented in **Paper III**.

Two tools to aid the batch processing and analysis of NMR spectra are presented, ImatraNMR for spectrum analysis (**Paper II**) and SimpeleNMR for NMR processing (**Paper III**). Applications for hydrocarbon analysis (**Paper III**) and plant extracts (**Paper IV**) are demonstrated, along with insights regarding the type of analysis performed. In (**Paper III**), powerful automated binning and integration is utilized in conjunction with model compounds to find cycloalkane signals in base oil samples, a novel feat for an NMR based method, while the computation of several average structural parameters for routine analysis are presented as well. **Paper IV** shows the superiority of histogram binning (implemented in ImatraNMR) compared to traditional equidistant binning for PCA analysis, and new JRES processing and PQN normalization functionality are demonstrated.

The presented tools are generic and also suited for many other use cases in which batch processing and basic analysis of numerous NMR spectra are required. The tools feature basic NMR processing for 1D and 2D spectra (apodization, FT, baseline correction) and analysis (signal search, integration, binning), and produce their output in the standard CSV and text formats, which can be easily utilized in many data analysis and spreadsheet programs.

Bibliography

[1] M. H. Levitt, *Spin Dynamics: Basics of Nuclear Magnetic Resonance*, John Wiley & Sons, Ltd., **2001**.

[2] P. Atkins, J. de Paula, *Atkins' Physical Chemistry 7th ed.*, Oxford University Press, **2002**.

[3] J. Cavanagh, W. J. Fairbrother, A. G. Palmer, M. Rance, N. J. Skelton, *Protein NMR Spectroscopy: Principles and Practice*, Elsevier Academic Press, **2007**.

[4] J. H. Lee, Y. Okuno, S. Cavagnero. Sensitivity enhancement in solution NMR: Emerging ideas and new frontiers. *Journal of Magnetic Resonance* **2014**, *241*, 18--31.

[5] D. I. Hoult, B. Bhakar. NMR signal reception: Virtual photons and coherent spontaneous emission. *Concepts in Magnetic Resonance* **1997**, *9*, 277--297.

[6] T. D. Claridge, *High-Resolution NMR Techniques in Organic Chemistry 1st ed.*, Vol. 19 of *Tetrahedron Organic Chemistry Series*, Elsevier Ltd, **1999**.

[7] O. Sørensen, G. Eich, M. Levitt, G. Bodenhausen, R. Ernst. Product operator formalism for the description of NMR pulse experiments. *Progress in Nuclear Magnetic Resonance Spectroscopy* **1984**, *16*, 163--192.

[8] D. P. Goldenberg. The product operator formalism: A physical and graphical interpretation. *Concepts in Magnetic Resonance Part A* **2010**, *36A*, 49--83.

[9] R. R. Ernst, G. Bodenhausen, A. Wokaun, *Principles of Nuclear Magnetic Resonance in One and Two Dimensions*, of *International Series of Monographs on Chemistry*, Oxford University Press, **1992**.

[10] G. Pagès, A. Gerdova, D. Williamson, V. Gilard, R. Martino, M. Malet-Martino. Evaluation of a Benchtop Cryogen-Free Low-Field ^{1}H NMR Spectrometer for the Analysis of Sexual Enhancement and Weight Loss Dietary Supplements Adulterated with Pharmaceutical Substances. *Analytical Chemistry* **2014**, *86*, 11897--11904.

[11] T. Parker, E. Limer, A. Watson, M. Defernez, D. Williamson, E. K. Kemsley. 60 MHz ^{1}H NMR spectroscopy for the analysis of edible oils. *TrAC Trends in Analytical Chemistry* **2014**, *57*, 147 -- 158.

[12] W. Jakes, A. Gerdova, M. Defernez, A. Watson, C. McCallum, E. Limer, I. Colquhoun, D. Williamson, E. Kemsley. Authentication of beef versus horse meat using 60 MHz ^{1}H NMR spectroscopy. *Food Chemistry* **2015**, *175*, 1 -- 9.

[13] K. Takeda. OPENCORE NMR: Open-source core modules for implementing an integrated FPGA-based NMR spectrometer. *Journal of Magnetic Resonance* **2008**, *192*, 218--229.

[14] H.-Y. Chen, Y. Kim, P. Nath, C. Hilty. An ultra-low cost NMR device with arbitrary pulse programming. *Journal of Magnetic Resonance* **2015**, *255*, 100--105.

[15] S. Berger, S. Braun, *200 and More NMR Experiments: A Practical Course*, Wiley-VCH, Weinheim, **2004**.

[16] W. R. Croasmun, R. M. K. Carlson, *Two-Dimensional NMR Spectroscopy: Applications for Chemists and Biochemists 2nd ed.*, John Wiley & Sons, inc., **1994**.

[17] D. Hoult. The principle of reciprocity in signal strength calculations—a mathematical guide. *Concepts in Magnetic Resonance* **2000**, *12*, 173--187.

[18] R. D. Farrant, J. C. Hollerton, S. M. Lynn, S. Provera, P. J. Sidebottom, R. J. Upton. NMR quantification using an artificial signal. *Magnetic Resonance in Chemistry* **2010**, *48*, 753--762.

[19] W. E. Hull. NMR Tips for Shimming, Part I. Computerized shimming with the Tuning algorithm. *Bruker SpinReport* **2003**, *152/153*, 53--61.

[20] Thermo Fisher Scientific Inc., *picoSpin Spectrometer Frequently Asked Questions*, **2013**.

[21] P. A. Keifer. 90° pulse width calibrations: How to read a pulse width array. *Concepts in Magnetic Resonance* **1999**, *11*, 165--180.

[22] W. H. Press, S. A. Teukolsky, W. T. Vetterling, B. P. Flannery, *Numerical Recipes: The Art of Scientific Computing 3rd ed.*, Cambridge University Press, **2007**.

[23] D. Moskau. Application of real time digital filters in NMR spectroscopy. *Concepts in Magnetic Resonance* **2002**, *15*, 164--176.

[24] S. W. Smith, *The Scientist and Engineer's Guide to Digital Signal Processing*, California Technical Publishing, San Diego, California, **1997**.

[25] L. Gengying, X. Haibin. Digital quadrature detection in nuclear magnetic resonance spectroscopy. *Review of Scientific Instruments* **1999**, *70*, 1511--1513.

[26] Agilent Technologies, *Agilent DirectDrive Datasheet*, **2013**.

[27] E. Jones, T. Oliphant, P. Peterson, et al., *SciPy: Open source scientific tools for Python*, **2001**--, [Online; accessed 2015-04-02]. `http://www.scipy.org/`.

[28] R. R. Ernst, W. A. Anderson. Application of Fourier Transform Spectroscopy to Magnetic Resonance. *Review of Scientific Instruments* **1966**, *37*, 93--102.

[29] R. R. Ernst. Zurich's Contributions to 50 Years Development of Bruker. *Angewandte Chemie International Edition* **2010**, *49*, 8310--8315.

[30] R. Freeman, G. A. Morris. The Varian story. *Journal of Magnetic Resonance* **2015**, *250*, 80 -- 84.

[31] J. W. Cooley, J. W. Tukey. An Algorithm for the Machine Calculation of Complex Fourier Series. *Mathematics of Computation* **1965**, *19*, pp. 297--301.

[32] S. van der Walt, S. Colbert, G. Varoquaux. The NumPy Array: A Structure for Efficient Numerical Computation. *Computing in Science Engineering* **2011**, *13*, 22--30.

[33] T. E. Oliphant. Python for Scientific Computing. *Computing in Science & Engineering* **2007**, *9*, 10--20.

[34] E. Bartholdi, R. Ernst. Fourier spectroscopy and the causality principle. *Journal of Magnetic Resonance (1969)* **1973**, *11*, 9 -- 19.

[35] A. Ebel, W. Dreher, D. Leibfritz. Effects of zero-filling and apodization on spectral integrals in discrete Fourier-transform spectroscopy of noisy data. *Journal of Magnetic Resonance* **2006**, *182*, 330 -- 338.

[36] J. J. Led, H. Gesmar. Application of the linear prediction method to NMR spectroscopy. *Chemical reviews* **1991**, *91*, 1413--1426.

[37] G. Zhu, A. Bax. Improved linear prediction of damped NMR signals using modified "forward-backward" linear prediction. *Journal of Magnetic Resonance (1969)* **1992**, *100*, 202 -- 207.

[38] H. de Brouwer. Evaluation of algorithms for automated phase correction of NMR spectra. *Journal of Magnetic Resonance* **2009**, *201*, 230--238.

[39] H. Kovacs, D. Moskau, M. Spraul. Cryogenically cooled probes—a leap in NMR technology. *Progress in Nuclear Magnetic Resonance Spectroscopy* **2005**, *46*, 131 -- 155.

[40] A. Webb. Microcoil nuclear magnetic resonance spectroscopy. *Journal of Pharmaceutical and Biomedical Analysis* **2005**, *38*, 892 -- 903.

[41] V. Rizzo, V. Pinciroli. Quantitative NMR in synthetic and combinatorial chemistry. *Journal of pharmaceutical and biomedical analysis* **2005**, *38*, 851--857.

[42] G. F. Pauli, T. Gödecke, B. U. Jaki, D. C. Lankin. Quantitative ^{1}H NMR. Development and Potential of an Analytical Method: An Update. *Journal of Natural Products* **2012**, *75*, 834--851.

[43] P. Horowitz, W. Hill, *The Art of Electronics 2nd ed.*, Cambridge Eniversity Press, **1989**.

[44] J. T. Bushberg, J. A. Seibert, E. M. L. Jr., J. M. Boone, *The Essential Physics of Medical Imaging*, Lippincott Williams & Wilkins, **2002**.

[45] D. I. Hoult, R. E. Richards. Critical Factors in the Design of Sensitive High Resolution Nuclear Magnetic Resonance Spectrometers. *Proceedings of the Royal Society of London A: Mathematical, Physical and Engineering Sciences* **1975**, *344*, 311--340.

[46] D. Reichert, G. Hempel. Receiver imperfections and CYCLOPS: An alternative description. *Concepts in Magnetic Resonance* **2002**, *14*, 130--139.

[47] P. Tuominen, *Todenäköisyyslaskenta I*, Limes ry, **2010**.

[48] C. J. Jones, C. K. Larive. Could smaller really be better? Current and future trends in high-resolution microcoil NMR spectroscopy. *Analytical and Bioanalytical Chemistry* **2012**, *402*, 61--68.

[49] *Shigemi Inc.* `http://www.shigeminmr.com`.

[50] K. Hayamizu, W. S. Price. A new type of sample tube for reducing convection effects in PGSE-NMR measurements of self-diffusion coefficients of liquid samples. *Journal of Magnetic Resonance* **2004**, *167*, 328 -- 333.

[51] D. Hoult, R. Richards. The signal-to-noise ratio of the nuclear magnetic resonance experiment. *Journal of Magnetic Resonance (1969)* **1976**, *24*, 71 -- 85.

[52] K. Hashi, S. Ohki, S. Matsumoto, G. Nishijima, A. Goto, K. Deguchi, K. Yamada, T. Noguchi, S. Sakai, M. Takahashi, Y. Yanagisawa, S. Iguchi, T. Yamazaki, H. Maeda, R. Tanaka, T. Nemoto, H. Suematsu, T. Miki, K. Saito, T. Shimizu. Achievement of 1020 MHz NMR. *Journal of Magnetic Resonance* **2015**, *256*, 30 -- 33.

[53] R. K. Harris, E. D. Becker, C. d. Menezes, S. M, R. Goodfellow, P. Granger. NMR nomenclature. Nuclear spin properties and conventions for chemical shifts(IUPAC Recommendations 2001). *Pure and Applied Chemistry* **2001**, *73*, 1795--1818.

[54] A. Overhauser. Polarization of Nuclei in Metals. *Physical Review* **1953**, *92*, 411--415.

[55] J. Jeener, B. H. Meier, P. Bachmann, R. R. Ernst. Investigation of exchange processes by two-dimensional NMR spectroscopy. *The Journal of Chemical Physics* **1979**, *71*, 4546--4553.

[56] R. Wagner, S. Berger. Gradient-selected NOESY — a fourfold reduction of the measurement time for the NOESY experiment. *Journal of Magnetic Resonance, Series A* **1996**, *123*, 119--121.

[57] R. Freeman, *A Handbook of Nuclear Magnetic Resonance 2nd ed.*, Addison Wesley Longman, **1997**.

[58] G. A. Morris, R. Freeman. Enhancement of nuclear magnetic resonance signals by polarization transfer. *Journal of the American Chemical Society* **1979**, *101*, 760--762.

[59] G. A. Morris. Sensitivity enhancement in nitrogen-15 NMR: polarization transfer using the INEPT pulse sequence. *Journal of the American Chemical Society* **1980**, *102*, 428--429.

[60] D. M. Doddrell, D. T. Pegg. Assignment of proton-decoupled carbon-13 spectra of complex molecules by using polarization transfer spectroscopy. A superior method to off-resonance decoupling. *Journal of the American Chemical Society* **1980**, *102*, 6388--6390.

[61] G. Bodenhausen, D. J. Ruben. Natural abundance nitrogen-15 NMR by enhanced heteronuclear spectroscopy. *Chemical Physics Letters* **1980**, *69*, 185 -- 189.

[62] J. Ardenkjær-Larsen, B. Fridlund, A. Gram, G. Hansson, L. Hansson, M. Lerche, R. Servin, M. Thaning, K. Golman. Increase in signal-to-noise ratio of >10,000 times in liquid-state NMR. *Proceedings of the National Academy of Sciences of the United States of America* **2003**, *100*, 10158--10163.

[63] P. R. Jensen, S. Meier, J. H. Ardenkjær-Larsen, J. Ø. Duus, M. Karlsson, M. H. Lerche. Detection of low-populated reaction intermediates with hyperpolarized NMR. *Chemical Communications* **2009**, 5168--5170.

[64] S. Meier, M. Karlsson, P. R. Jensen, M. H. Lerche, J. Ø. Duus. Metabolic pathway visualization in living yeast by DNP-NMR. *Molecular BioSystems* **2011**, *7*, 2834--2836.

[65] T. G. Walker, W. Happer. Spin-exchange optical pumping of noble-gas nuclei. *Reviews of Modern Physics* **1997**, *69*, 629--642.

[66] A. Comment, B. van den Brandt, K. Uffmann, F. Kurdzesau, S. Jannin, J. Konter, P. Hautle, W. Wenckebach, R. Gruetter, J. van der Klink. Design and performance of a DNP prepolarizer coupled to a rodent MRI scanner. *Concepts in Magnetic Resonance Part B: Magnetic Resonance Engineering* **2007**, *31B*, 255--269.

[67] S. Jannin, A. Comment, F. Kurdzesau, J. A. Konter, P. Hautle, B. van den Brandt, J. J. van der Klink. A 140 GHz prepolarizer for dissolution dynamic nuclear polarization. *The Journal of Chemical Physics* **2008**, *128*, 241102.

[68] P. Miéville, S. Jannin, G. Bodenhausen. Relaxometry of insensitive nuclei: Optimizing dissolution dynamic nuclear polarization. *Journal of Magnetic Resonance* **2011**, *210*, 137 -- 140.

[69] J. Leggett, R. Hunter, J. Granwehr, R. Panek, A. J. Perez-Linde, A. J. Horsewill, J. McMaster, G. Smith, W. Kockenberger. A dedicated spectrometer for dissolution DNP NMR spectroscopy. *Phys. Chem. Chem. Phys.* **2010**, *12*, 5883--5892.

[70] P. Miéville, P.ville, P. Ahuja, R. Sarkar, S. Jannin, P. Vasos, S. Gerber-Lemaire, M. Mishkovsky, A. Comment, R. Gruetter, O. Ouari, P. Tordo, G. Bodenhausen. Scavenging Free Radicals To Preserve Enhancement and Extend Relaxation Times in NMR using Dynamic Nuclear Polarization. *Angewandte Chemie* **2010**, *122*, 6318--6321.

[71] S. Jannin, A. Bornet, S. Colombo, G. Bodenhausen. Low-temperature cross polarization in view of enhancing dissolution Dynamic Nuclear Polarization in NMR. *Chemical Physics Letters* **2011**, *517*, 234 -- 236.

[72] L. Frydman, T. Scherf, A. Lupulescu. The acquisition of multidimensional NMR spectra within a single scan. *Proceedings of the National Academy of Sciences* **2002**, *99*, 15858--15862.

[73] M. Mishkovsky, L. Frydman. Progress in Hyperpolarized Ultrafast 2D NMR Spectroscopy. *ChemPhysChem* **2008**, *9*, 2340--2348.

[74] P. Giraudeau, Y. Shrot, L. Frydman. Multiple Ultrafast, Broadband 2D NMR Spectra of Hyperpolarized Natural Products. *Journal of the American Chemical Society* **2009**, *131*, 13902--13903.

[75] C. Griesinger, M. Bennati, H. Vieth, C. Luchinat, G. Parigi, P. Höfer, F. Engelke, S. Glaser, V. Denysenkov, T. Prisner. Dynamic nuclear polarization at high magnetic fields in liquids. *Progress in Nuclear Magnetic Resonance Spectroscopy* **2012**, *64*, 4 -- 28.

[76] S. B. Duckett, C. J. Sleigh. Applications of the parahydrogen phenomenon: A chemical perspective. *Progress in Nuclear Magnetic Resonance Spectroscopy* **1999**, *34*, 71 -- 92.

[77] J. Natterer, J. Bargon. Parahydrogen induced polarization. *Progress in Nuclear Magnetic Resonance Spectroscopy* **1997**, *31*, 293--315.

[78] C. R. Bowers, D. P. Weitekamp. Parahydrogen and synthesis allow dramatically enhanced nuclear alignment. *Journal of the American Chemical Society* **1987**, *109*, 5541--5542.

[79] T. C. Eisenschmid, R. U. Kirss, P. P. Deutsch, S. I. Hommeltoft, R. Eisenberg, J. Bargon, R. G. Lawler, A. L. Balch. Para hydrogen induced polarization in hydrogenation reactions. *Journal of the American Chemical Society* **1987**, *109*, 8089--8091.

[80] M. G. Pravica, D. P. Weitekamp. Net NMR alignment by adiabatic transport of parahydrogen addition products to high magnetic field. *Chemical Physics Letters* **1988**, *145*, 255--258.

[81] M. Haake, J. Natterer, J. Bargon. Efficient NMR Pulse Sequences to Transfer the Parahydrogen-Induced Polarization to Hetero Nuclei. *Journal of the American Chemical Society* **1996**, *118*, 8688--8691.

[82] R. W. Adams, J. A. Aguilar, K. D. Atkinson, M. J. Cowley, P. I. P. Elliott, S. B. Duckett, G. G. R. Green, I. G. Khazal, J. López-Serrano, D. C. Williamson. Reversible Interactions with para-hydrogen Enhance NMR Sensitivity by Polarization Transfer. *Science* **2009**, *323*, 1708--1711.

[83] D. Blazina, S. B. Duckett, T. K. Halstead, C. M. Kozak, R. J. K. Taylor, M. S. Anwar, J. A. Jones, H. A. Carteret. Generation and interrogation of a pure nuclear spin state by parahydrogen-enhanced

NMR spectroscopy: a defined initial state for quantum computation. *Magnetic Resonance in Chemistry* **2005**, *43*, 200--208.

[84] magritek, Philipsstraße 8, 52068 Aachen, Germany, *Spinsolve - The fastest compact NMR spectrometer*, **2013**. `http://www.magritek.com/wp-content/uploads/2013/12/Spinsolve-brochure-web-high-res.pdf`.

[85] P. Styles, N. Soffe. A high-resolution NMR probe in which the coil and preamplifier are cooled with liquid helium. *Journal of Magnetic Resonance* **1984**, *60*, 397–404.

[86] A. E. Kelly, H. D. Ou, R. Withers, V. Dötsch. Low-conductivity buffers for high-sensitivity NMR measurements. *Journal of the American Chemical Society* **2002**, *124*, 12013--12019.

[87] P. Styles, N. F. Soffe, C. A. Scott. The first cryoprobe – Some recollections. *Journal of Magnetic Resonance* **2011**, *213*, 355 -- 356.

[88] *Topspin Acquisition, Commands and Parameters*, Bruker BioSpin GmbH, **2010**, Topspin 3.0.

[89] G. Diestelhorst, M. Miller, ^{UNITY} *INOVA Technical Reference*, Varian, Inc., **2000**.

[90] M. Delsuc, J. Lallemand. Improvement of dynamic range in NMR by oversampling. *Journal of Magnetic Resonance (1969)* **1986**, *69*, 504 -- 507.

[91] G. Wider. Elimination of baseline artifacts in NMR spectra by oversampling. *Journal of Magnetic Resonance (1969)* **1990**, *89*, 406 -- 409.

[92] J. Halámek, V. Vondra, M. Kasal. The elimination of baseline distortions induced by audio filters. *Journal of Magnetic Resonance, Series A* **1994**, *110*, 194--197.

[93] R. Beckman, E. Zuiderweg. Guidelines for the Use of Oversampling in Protein NMR. *Journal of Magnetic Resonance, Series A* **1995**, *113*, 223 -- 231.

[94] H. Koskela, *Quantitative 2D NMR Studies*, Vol. 66 of *Annual Reports on NMR Spectroscopy*, G. A. Webb (Ed.), Academic Press, **2009**, chapter 1, pp. 1 -- 31.

[95] K. Mehr, B. John, D. Russell, D. Avizonis. Electronic referencing techniques for quantitative NMR: pitfalls and how to avoid them using amplitude-corrected referencing through signal injection. *Analytical chemistry* **2008**, *80*, 8320--8323.

[96] D. D. Traficante. Optimum tip angle and relaxation delay for quantitative analysis. *Concepts in Magnetic Resonance* **1992**, *4*, 153--160.

[97] M. H. Levitt. Composite pulses. *Progress in Nuclear Magnetic Resonance Spectroscopy* **1986**, *18*, 61 -- 122.

[98] S. Wimperis. Broadband, Narrowband, and Passband Composite Pulses for Use in Advanced NMR Experiments. *Journal of Magnetic Resonance, Series A* **1994**, *109*, 221 -- 231.

[99] T.-L. Hwang, P. C. van Zijl, M. Garwood. Fast Broadband Inversion by Adiabatic Pulses. *Journal of Magnetic Resonance* **1998**, *133*, 200 -- 203.

[100] M. H. Levitt, R. Freeman. NMR population inversion using a composite pulse. *Journal of Magnetic Resonance (1969)* **1979**, *33*, 473 -- 476.

[101] R. Freeman, S. P. Kempsell, M. H. Levitt. Radiofrequency pulse sequences which compensate their own imperfections. *Journal of Magnetic Resonance (1969)* **1980**, *38*, 453 -- 479.

[102] M. H. Levitt, R. Freeman. Compensation for pulse imperfections in NMR spin-echo experiments. *Journal of Magnetic Resonance (1969)* **1981**, *43*, 65 -- 80.

[103] M. H. Levitt, R. Freeman, T. Frenkiel. Broadband heteronuclear decoupling. *Journal of Magnetic Resonance (1969)* **1982**, *47*, 328 -- 330.

[104] A. J. Shaka, J. Keeler, T. Frenkiel, R. Freeman. An improved sequence for broadband decoupling: WALTZ-16. *J. Magn. Reson.* **1983**, *52*, 335.

[105] A. Shaka, P. Barker, R. Freeman. Computer-optimized decoupling scheme for wideband applications and low-level operation. *Journal of Magnetic Resonance (1969)* **1985**, *64*, 547 -- 552.

[106] R. Tycko, H. Cho, E. Schneider, A. Pines. Composite pulses without phase distortion. *Journal of Magnetic Resonance (1969)* **1985**, *61*, 90 -- 101.

[107] A. Shaka. Composite pulses for ultra-broadband spin inversion. *Chemical Physics Letters* **1985**, *120*, 201 -- 205.

[108] A. J. Shaka, P. B. Barker, R. Freeman. Experimental demonstration of wideband spin inversion. *J. Magn. Reson.* **1986**, *67*, 580--584.

[109] R. Freeman. Selective excitation in high-resolution NMR. *Chemical Reviews* **1991**, *91*, 1397--1412.

[110] H. Geen, R. Freeman. Band-selective radiofrequency pulses. *Journal of Magnetic Resonance (1969)* **1991**, *93*, 93 -- 141.

[111] E. Kupce, J. Boyd, I. Campbell. Short Selective Pulses for Biochemical Applications. *Journal of Magnetic Resonance, Series B* **1995**, *106*, 300 -- 303.

[112] M. A. Smith, H. Hu, A. Shaka. Improved Broadband Inversion Performance for NMR in Liquids. *Journal of Magnetic Resonance* **2001**, *151*, 269 -- 283.

[113] T. E. Skinner, T. O. Reiss, B. Luy, N. Khaneja, S. J. Glaser. Application of optimal control theory to the design of broadband excitation pulses for high-resolution NMR. *Journal of Magnetic Resonance* **2003**, *163*, 8 -- 15.

[114] T. E. Skinner, T. O. Reiss, B. Luy, N. Khaneja, S. J. Glaser. Reducing the duration of broadband excitation pulses using optimal control with limited RF amplitude. *Journal of Magnetic Resonance* **2004**, *167*, 68 -- 74.

[115] T. E. Skinner, K. Kobzar, B. Luy, M. R. Bendall, W. Bermel, N. Khaneja, S. J. Glaser. Optimal control design of constant amplitude phase-modulated pulses: Application to calibration-free broadband excitation. *Journal of Magnetic Resonance* **2006**, *179*, 241 -- 249.

[116] J. Baum, R. Tycko, A. Pines. Broadband population inversion by phase modulated pulses. *The Journal of Chemical Physics* **1983**, *79*, 4643--4644.

[117] A. Tannús, M. Garwood. Adiabatic pulses. *NMR in Biomedicine* **1997**, *10*, 423--434.

[118] M. Garwood, L. DelaBarre. The Return of the Frequency Sweep: Designing Adiabatic Pulses for Contemporary NMR. *Journal of Magnetic Resonance* **2001**, *153*, 155 -- 177.

[119] E. Kupce, R. Freeman. Adiabatic Pulses for Wideband Inversion and Broadband Decoupling. *Journal of Magnetic Resonance, Series A* **1995**, *115*, 273 -- 276.

[120] E. Kupče, R. Freeman. Compensated adiabatic inversion pulses: broadband INEPT and HSQC. *Journal of Magnetic Resonance* **2007**, *187*, 258--265.

[121] R. Freeman, H. Hill, R. Kaptein. Proton-Decoupled NMR. Spectra of Carbon-13 With the Nuclear Overhauser Effect Suppressed. *Journal of Magnetic Resonance (1969)* **1972**, *7*, 327 -- 329.

[122] I. W. Burton, M. A. Quilliam, J. A. Walter. Quantitative ^{1}H NMR with External Standards: Use in Preparation of Calibration Solutions for Algal Toxins and Other Natural Products. *Analytical Chemistry* **2005**, *77*, 3123--3131.

[123] S. Akoka, L. Barantin, M. Trierweiler. Concentration measurement by proton NMR using the ERETIC method. *Analytical Chemistry* **1999**, *71*, 2554--2557.

[124] G. Wider, L. Dreier. Measuring Protein Concentrations by NMR Spectroscopy. *Journal of the American Chemical Society* **2006**, *128*, 2571--2576.

[125] K. I. Marro, D. Lee, E. G. Shankland, C. M. Mathis, C. E. Hayes, C. E. Amara, M. J. Kushmerick. Synthetic signal injection using inductive coupling. *Journal of Magnetic Resonance* **2008**, *194*, 67--75.

[126] F. Ziarelli, S. Viel, S. Caldarelli, D. N. Sobieski, M. P. Augustine. General implementation of the ERETIC™ method for pulsed field gradient probe heads. *Journal of Magnetic Resonance* **2008**, *194*, 307--312.

[127] G. Maniara, K. Rajamoorthi, S. Rajan, G. W. Stockton. Method performance and validation for quantitative analysis by ^{1}H and ^{31}P NMR spectroscopy. Applications to analytical standards and agricultural chemicals. *Analytical chemistry* **1998**, *70*, 4921--4928.

[128] M. Bauer, A. Bertario, G. Boccardi, X. Fontaine, R. Rao, D. Verrier. Reproducibility of ^{1}H NMR integrals: A collaborative study. *Journal of pharmaceutical and biomedical analysis* **1998**, *17*, 419--425.

[129] V. Pinciroli, R. Biancardi, G. Visentin, V. Rizzo. The well-characterized synthetic molecule: a role for quantitative ^{1}H NMR. *Organic process research & development* **2004**, *8*, 381--384.

[130] F. Malz, H. Jancke. Validation of quantitative NMR. *Journal of pharmaceutical and biomedical analysis* **2005**, *38*, 813--823.

[131] R. J. Wells, J. M. Hook, T. S. Al-Deen, D. B. Hibbert. Quantitative nuclear magnetic resonance (QNMR) spectroscopy for assessing the purity of technical grade agrochemicals: 2,4-dichlorophenoxyacetic acid (2,4-D) and sodium 2,2-dichloropropionate (Dalapon sodium). *Journal of agricultural and food chemistry* **2002**, *50*, 3366--3374.

[132] G. F. Pauli, B. U. Jaki, D. C. Lankin. A Routine Experimental Protocol for qHNMR Illustrated with Taxol. *Journal of natural products* **2007**, *70*, 589--595.

[133] D. A. Foley, A. L. Dunn, M. T. Zell. Reaction monitoring using online vs tube NMR spectroscopy: seriously different results. *Magnetic Resonance in Chemistry* **2015**.

[134] L. Griffiths, A. M. Irving. Assay by nuclear magnetic resonance spectroscopy: quantification limits. *Analyst* **1998**, *123*, 1061--1068.

[135] G. D. Tredwell, V. Behrends, F. M. Geier, M. Liebeke, J. G. Bundy. Between-Person Comparison of Metabolite Fitting for NMR-Based Quantitative Metabolomics. *Analytical Chemistry* **2011**, *83*, 8683--8687.

[136] R. Crouch, D. Russell, *Easy, Precise and Accurate Quantitative NMR*, Application Note, Agilent Technologies, Inc., **2011**.

[137] J. Jeener, *Ampére International Summer School in Baško Polje*, Lecture, **1971**.

[138] W. P. Aue, E. Bartholdi, R. R. Ernst. Two-dimensional spectroscopy. Application to nuclear magnetic resonance. *The Journal of Chemical Physics* **1976**, *64*, 2229--2246.

[139] L. Müller, A. Kumar, R. R. Ernst. Two-dimensional carbon-13 NMR spectroscopy. *The Journal of Chemical Physics* **1975**, *63*, 5490--5491.

[140] S. Heikkinen, M. M. Toikka, P. T. Karhunen, I. A. Kilpeläinen. Quantitative 2D HSQC (Q-HSQC) via Suppression of J-Dependence of Polarization Transfer in NMR Spectroscopy: Application to Wood Lignin. *J. Am. Chem. Soc.* **2003**, *125*, 4362--4367.

[141] R. D. Boyer, R. Johnson, K. Krishnamurthy. Compensation of refocusing inefficiency with synchronized inversion sweep (CRISIS) in multiplicity-edited HSQC. *Journal of Magnetic Resonance* **2003**, *165*, 253 -- 259.

[142] H. Koskela, I. Kilpeläinen, S. Heikkinen. Some aspects of quantitative 2D NMR. *J. Magn. Reson.* **2005**, *174*, 237--244.

[143] L. Mueller, P. Legault, A. Pardi. Improved RNA Structure Determination by Detection of NOE Contacts to Exchange-Broadened Amino Protons. *Journal of the American Chemical Society* **1995**, *117*, 11043--11048.

[144] F. Mulder, C. Spronk, M. Slijper, R. Kaptein, R. Boelens. Improved HSQC experiments for the observation of exchange broadened signals. *Journal of Biomolecular NMR* **1996**, *8*, 223--228.

[145] T. Gullion, D. B. Baker, M. S. Conradi. New, compensated Carr-Purcell sequences. *Journal of Magnetic Resonance (1969)* **1990**, *89*, 479 -- 484.

[146] H. Koskela, O. Heikkilä, I. Kilpeläinen, S. Heikkinen. Quantitative two-dimensional HSQC experiment for high magnetic field NMR spectrometers. *Journal of Magnetic Resonance* **2010**, *202*, 24 -- 33.

[147] D. J. Peterson, N. M. Loening. QQ-HSQC: a quick, quantitative heteronuclear correlation experiment for NMR spectroscopy. *Magnetic Resonance in Chemistry* **2007**, *45*, 937--941.

[148] R. Freeman. The concept of spatially-encoded single-scan NMR. *Concepts in Magnetic Resonance Part A* **2011**, *38A*, 1--6.

[149] D. M. Doddrell, D. T. Pegg, M. R. Bendall. Distortionless enhancement of NMR signals by polarization transfer. *J. Magn. Reson* **1982**, *48*, 323--327.

[150] T. J. Henderson. Sensitivity-enchanced quantitative ^{13}C NMR spectroscopy via cancellation of $^{1}J_{CH}$ dependence in DEPT polarization transfers. *J. Am. Chem. Soc.* **2004**, *126*, 3682--3683.

[151] S. L. Patt, J. N. Shoolery. Attached proton test for carbon-13 NMR. *J. Magn. Reson* **1982**, *46*, 535--539.

[152] B. Jiang, N. Xiao, H. Liu, Z. Zhou, X. an Mao, M. Liu. Optimized Quantitative DEPT and Quantitative POMMIE Experiments for ^{13}C NMR. *Anal. Chem.* **2008**, *80*, 8293--8298.

[153] J. Bulsing, W. M. Brooks, J. Field, D. M. Doddrell. Polarization transfer via an intermediate multiple-quantum state of maximum order. *Journal of Magnetic Resonance (1969)* **1984**, *56*, 167 -- 173.

[154] D. Burum, R. Ernst. Net polarization transfer via a J-ordered state for signal enhancement of low-sensitivity nuclei. *Journal of Magnetic Resonance (1969)* **1980**, *39*, 163 -- 168.

[155] M. Bendall, D. T. Pegg, D. M. Doddrell. Polarization transfer pulse sequences for two-dimensional NMR by heisenberg vector analysis. *Journal of Magnetic Resonance (1969)* **1981**, *45*, 8 -- 29.

[156] H. Koskela, T. Väänänen. Quantitative determination of aliphatic hydrocarbon compounds by 2D NMR. *Magnetic Resonance in Chemistry* **2002**, *40*, 705--715.

[157] D. Marion, M. Ikura, R. Tschudin, A. Bax. Rapid recording of 2D NMR spectra without phase cycling. Application to the study of hydrogen exchange in proteins. *J. Magn. Reson* **1989**, *85*, 393--399.

[158] V. Manu, A. Kumar. Fast and accurate quantification using Genetic Algorithm optimized ^{1}H-^{13}C refocused constant-time INEPT. *Journal of Magnetic Resonance* **2013**, *234*, 106 -- 111.

[159] L. Braunschweiler, R. Ernst. Coherence transfer by isotropic mixing: Application to proton correlation spectroscopy. *Journal of Magnetic Resonance (1969)* **1983**, *53*, 521 -- 528.

[160] R. E. Hurd. Gradient-enhanced spectroscopy. *Journal of Magnetic Resonance (1969)* **1990**, *87*, 422 -- 428.

[161] U. Piantini, O. W. Sorensen, R. R. Ernst. Multiple quantum filters for elucidating NMR coupling networks. *Journal of the American Chemical Society* **1982**, *104*, 6800--6801.

[162] P. Giraudeau, N. Guignard, E. Hillion, E. Baguet, S. Akoka. Optimization of homonuclear 2D NMR for fast quantitative analysis: Application to tropine–nortropine mixtures. *Journal of Pharmaceutical and Biomedical Analysis* **2007**, *43*, 1243 -- 1248.

[163] M. J. Thrippleton, J. Keeler. Elimination of Zero-Quantum Interference in Two-Dimensional NMR Spectra. *Angewandte Chemie International Edition* **2003**, *42*, 3938--3941.

[164] J. P. M. Van Duynhoven, J. Goudriaan, C. W. Hilbers, S. S. Wijmenga. Quantitative evaluation of TOCSY data. Application to sugar ring conformational analysis. *Journal of the American Chemical Society* **1992**, *114*, 10055--10056.

[165] K. Hu, W. M. Westler, J. L. Markley. Simultaneous quantification and identification of individual chemicals in metabolite mixtures by two-dimensional extrapolated time-zero ^{1}H–^{13}C HSQC ($HSQC_0$). *Journal of the American Chemical Society* **2011**, *133*, 1662--1665.

[166] J. Barna, E. Laue, M. Mayger, J. Skilling, S. Worrall. Exponential sampling, an alternative method for sampling in two-dimensional NMR experiments. *Journal of Magnetic Resonance (1969)* **1987**, *73*, 69 -- 77.

[167] J. C. Hoch, A. S. Stern, D. L. Donoho, I. M. Johnstone. Maximum entropy reconstruction of complex (phase-sensitive) spectra. *Journal of Magnetic Resonance (1969)* **1990**, *86*, 236 -- 246.

[168] M. Mobli, J. C. Hoch. Maximum entropy spectral reconstruction of nonuniformly sampled data. *Concepts in Magnetic Resonance Part A* **2008**, *32A*, 436--448.

[169] M. Mobli, M. W. Maciejewski, A. D. Schuyler, A. S. Stern, J. C. Hoch. Sparse sampling methods in multidimensional NMR. *Phys. Chem. Chem. Phys.* **2012**, *14*, 10835--10843.

[170] V. A. Jaravine, A. V. Zhuravleva, P. Permi, I. Ibraghimov, V. Y. Orekhov. Hyperdimensional NMR Spectroscopy with Nonlinear Sampling. *Journal of the American Chemical Society* **2008**, *130*, 3927--3936, PMID: 18311971.

[171] S. G. Hyberts, G. J. Heffron, N. G. Tarragona, K. Solanky, K. A. Edmonds, H. Luithardt, J. Fejzo, M. Chorev, H. Aktas, K. Colson, K. H. Falchuk, J. A. Halperin, , G. Wagner. Ultrahigh-Resolution 1H–13C HSQC Spectra of Metabolite Mixtures Using Nonlinear Sampling and Forward Maximum Entropy Reconstruction. *Journal of the American Chemical Society* **2007**, *129*, 5108--5116.

[172] R. K. Rai, N. Sinha. Fast and Accurate Quantitative Metabolic Profiling of Body Fluids by Nonlinear Sampling of 1H–13C Two-Dimensional Nuclear Magnetic Resonance Spectroscopy. *Analytical Chemistry* **2012**, *84*, 10005--10011.

[173] A. L. Guennec, P. Giraudeau, S. Caldarelli. Evaluation of Fast 2D NMR for Metabolomics. *Analytical Chemistry* **2014**, *86*, 5946--5954, PMID: 24856256.

[174] A. Le Guennec, J.-N. Dumez, P. Giraudeau, S. Caldarelli. Resolution-enhanced 2D NMR of complex mixtures by non-uniform sampling. *Magnetic Resonance in Chemistry* **2015**, *53*, 913--920.

[175] K. Zangger, H. Sterk. Homonuclear broadband-decoupled NMR spectra. *Journal of Magnetic Resonance* **1997**, *124*, 486--489.

[176] K. Zangger. Pure shift NMR. *Progress in Nuclear Magnetic Resonance Spectroscopy* **2015**, *86–87*, 1 -- 20.

[177] W. P. Aue, J. Karhan, R. R. Ernst. Homonuclear broad band decoupling and two-dimensional J-resolved NMR spectroscopy. *The Journal of Chemical Physics* **1976**, *64*, 4226--4227.

[178] M. J. Thrippleton, R. A. Edden, J. Keeler. Suppression of strong coupling artefacts in J-spectra. *Journal of Magnetic Resonance* **2005**, *174*, 97 -- 109.

[179] A. Bax, R. Freeman, G. A. Morris. A simple method for suppressing dispersion-mode contributions in NMR spectra: The "pseudo echo". *Journal of Magnetic Resonance (1969)* **1981**, *43*, 333 -- 338.

[180] J.-M. Nuzillard. Time-Reversal of NMR Signals by Linear Prediction. Application to Phase-Sensitive Homonuclear J-Resolved Spectroscopy. *Journal of Magnetic Resonance, Series A* **1996**, *118*, 132--135.

[181] A. Martinez, F. Bourdreux, E. Riguet, J.-M. Nuzillard. High-resolution and high-sensitivity 2D homonuclear J-resolved NMR spectroscopy. *Magnetic Resonance in Chemistry* **2012**, *50*, 28--32.

[182] P. Sakhaii, W. Bermel. Improving the sensitivity of conventional spin echo spectra by preservation of initial signal-to-noise ratio. *Journal of Magnetic Resonance* **2014**, *242*, 220--223.

[183] P. Mutzenhardt, F. Guenneau, D. Canet. A Procedure for Obtaining Pure Absorption 2D J-Spectra: Application to Quantitative Fully J-Decoupled Homonuclear NMR Spectra. *Journal of Magnetic Resonance* **1999**, *141*, 312 -- 321.

[184] V. A. Mandelshtam, H. S. Taylor, A. Shaka. Application of the Filter Diagonalization Method to One- and Two-Dimensional NMR Spectra. *Journal of Magnetic Resonance* **1998**, *133*, 304 -- 312.

[185] V. A. Mandelshtam, N. D. Taylor, H. Hu, M. Smith, A. Shaka. Highly resolved double absorption 2D NMR spectra from complex severely truncated 2D phase-modulated signals by filter-diagonalization-averaging method. *Chemical Physics Letters* **1999**, *305*, 209 -- 216.

[186] S. Simova, H. Sengstschmid, R. Freeman. Proton Chemical-Shift Spectra. *Journal of Magnetic Resonance* **1997**, *124*, 104 -- 121.

[187] A. J. Pell, J. Keeler. Two-dimensional J-spectra with absorption-mode lineshapes. *Journal of Magnetic Resonance* **2007**, *189*, 293 -- 299.

[188] *Alpestre version 0.1.1*. `http://eos.univ-reims.fr/LSD/JmnSoft/Alpestre/`.

[189] P. Foxall, J. Parkinson, I. Sadler, J. Lindon, J. Nicholson. Analysis of biological fluids using 600 MHz proton NMR spectroscopy: Application of homonuclear two-dimensional J-resolved spectroscopy to urine and blood plasma for spectral simplification and assignment. *Journal of Pharmaceutical and Biomedical Analysis* **1993**, *11*, 21 -- 31.

[190] J. K. Nicholson, P. J. D. Foxall, M. Spraul, R. D. Farrant, J. C. Lindon. 750 MHz ^{1}H and ^{1}H-^{13}C NMR Spectroscopy of Human Blood Plasma. *Analytical Chemistry* **1995**, *67*, 793--811.

[191] M. R. Viant. Improved methods for the acquisition and interpretation of NMR metabolomic data. *Biochemical and Biophysical Research Communications* **2003**, *310*, 943 -- 948.

[192] J. C. Lindon, E. Holmes, J. K. Nicholson. So What's the Deal with Metabonomics? *Analytical Chemistry* **2003**, *75*, 384 A--391 A.

[193] P. Krishnan, N. Kruger, R. Ratcliffe. Metabolite fingerprinting and profiling in plants using NMR. *Journal of experimental botany* **2005**, *56*, 255--265.

[194] C. Ludwig, M. R. Viant. Two-dimensional J-resolved NMR spectroscopy: review of a key methodology in the metabolomics toolbox. *Phytochemical Analysis* **2010**, *21*, 22--32.

[195] S. Tiziani, A. Lodi, C. Ludwig, H. M. Parsons, M. R. Viant. Effects of the application of different window functions and projection methods on processing of 1H J-resolved nuclear magnetic resonance spectra for metabolomics. *Analytica Chimica Acta* **2008**, *610*, 80 -- 88.

[196] H. M. Parsons, C. Ludwig, M. R. Viant. Line-shape analysis of J-resolved NMR spectra: application to metabolomics and quantification of intensity errors from signal processing and high signal congestion. *Magnetic Resonance in Chemistry* **2009**, *47*, S86--S95.

[197] H. M. Parsons, D. R. Ekman, T. W. Collette, M. R. Viant. Spectral relative standard deviation: a practical benchmark in metabolomics. *Analyst* **2009**, *134*, 478--485.

[198] N. H. Meyer, K. Zangger. Simplifying proton NMR spectra by instant homonuclear broadband decoupling. *Angewandte Chemie International Edition* **2013**, *52*, 7143--7146.

[199] J. A. Aguilar, S. Faulkner, M. Nilsson, G. A. Morris. Pure shift ^{1}H NMR: a resolution of the resolution problem? *Angewandte Chemie International Edition* **2010**, *49*, 3901--3903.

[200] S. Glanzer, K. Zangger. Directly Decoupled Diffusion-Ordered NMR Spectroscopy for the Analysis of Compound Mixtures. *Chemistry – A European Journal* **2014**, *20*, 11171--11175.

[201] J. Garbow, D. Weitekamp, A. Pines. Bilinear rotation decoupling of homonuclear scalar interactions. *Chemical Physics Letters* **1982**, *93*, 504 -- 509.

[202] J. A. Aguilar, M. Nilsson, G. A. Morris. Simple proton spectra from complex spin systems: pure shift NMR spectroscopy using BIRD. *Angewandte Chemie* **2011**, *123*, 9890--9891.

[203] L. Paudel, R. W. Adams, P. Király, J. A. Aguilar, M. Foroozandeh, M. J. Cliff, M. Nilsson, P. Sándor, J. P. Waltho, G. A. Morris. Simultaneously enhancing spectral resolution and sensitivity in heteronuclear correlation NMR spectroscopy. *Angewandte Chemie International Edition* **2013**, *52*, 11616--11619.

[204] A. J. Pell, R. A. Edden, J. Keeler. Broadband proton-decoupled proton spectra. *Magnetic Resonance in Chemistry* **2007**, *45*, 296--316.

[205] M. Foroozandeh, R. W. Adams, N. J. Meharry, D. Jeannerat, M. Nilsson, G. A. Morris. Ultrahigh-Resolution NMR Spectroscopy. *Angewandte Chemie International Edition* **2014**, *53*, 6990--6992.

[206] K. J. Donovan, L. Frydman. HyperBIRD: A Sensitivity-Enhanced Approach to Collecting Homonuclear-Decoupled Proton NMR Spectra. *Angewandte Chemie International Edition* **2015**, *54*, 594--598.

[207] S. McConnell, *Code Complete: a practical handbook on software construction 2nd ed.*, Microsoft Press, **2004**.

[208] *Information is Beautiful: Million Lines of Code*. `http://www.informationisbeautiful.net/visualizations/million-lines-of-code/`.

[209] Z. Jiang, R. Mangharam, *University of Pennsylvania Develops Electrophysiological Heart Model for Real-Time Closed-Loop Testing of Pacemakers*, **2013**.

[210] R Development Core Team, *R: A Language and Environment for Statistical Computing*, R Foundation for Statistical Computing, Vienna, Austria, **2008**. `http://www.R-project.org`.

[211] *Creative Commons Attribution-NonCommercial 2.5 Generic*. `http://creativecommons.org/licenses/by-nc/2.5/`.

[212] *The average salary for a Sr. Software Engineer / Developer / Programmer*. `http://www.payscale.com/research/US/Job=Sr._Software_Engineer_%2F_Developer_%2F_Programmer/Salary`.

[213] Bruker Biospin, *SampleJET*. `https://www.bruker.com/products/mr/nmr/automation/automation/samplejet/`.

[214] P. Soininen, A. J. Kangas, P. Wurtz, T. Tukiainen, T. Tynkkynen, R. Laatikainen, M.-R. Jarvelin, M. Kahonen, T. Lehtimaki, J. Viikari, O. T. Raitakari, M. J. Savolainen, M. Ala-Korpela. High-throughput serum NMR metabonomics for cost-effective holistic studies on systemic metabolism. *Analyst* **2009**, *134*, 1781--1785.

[215] R. L. Haner, W. Llanos, L. Mueller. Small Volume Flow Probe for Automated Direct-Injection NMR Analysis: Design and Performance. *Journal of Magnetic Resonance* **2000**, *143*, 69 -- 78.

[216] Z. Pan, D. Raftery. Comparing and combining NMR spectroscopy and mass spectrometry in metabolomics. *Analytical and Bioanalytical Chemistry* **2007**, *387*, 525--527.

[217] M. Weiger, D. Moskau, R. Kerssebaum, W. E. Hull. NMR Tips for Shimming, Part III. Gradient shimming: principles and practical aspects. *Bruker SpinReport* **2005**, 1--27.

[218] E. Schreiber, *NMR Spectroscopy, User Guide. Varian NMR Systems with VnmrJ Software*, Varian, Inc., **2007**, VnmrJ 2.2C.

[219] R. R. Ernst. Measurement and Control of Magnetic Field Homogeneity. *Review of Scientific Instruments* **1968**, *39*, 998--1012.

[220] W. E. Hull. NMR Tips for Shimming, Part II. Computerized shimming with the simplex algorithm. *Bruker SpinReport* **2004**, *154/155*, 15--17.

[221] J. A. Nelder, R. Mead. A Simplex Method for Function Minimization. *The Computer Journal* **1965**, *7*, 308--313.

[222] M. G. Prammer, J. C. Haselgrove, M. Shinnar, J. S. Leigh. A new approach to automatic shimming. *Journal of Magnetic Resonance (1969)* **1988**, *77*, 40 -- 52.

[223] P. Vanzijl, S. Sukumar, M. Johnson, P. Webb, R. Hurd. Optimized Shimming for High-Resolution NMR Using Three-Dimensional Image-Based Field Mapping. *Journal of Magnetic Resonance, Series A* **1994**, *111*, 203 -- 207.

[224] S. Sukumar, M. Johnson, R. Hurd, P. van Zijl. Automated Shimming for Deuterated Solvents Using Field Profiling. *Journal of Magnetic Resonance* **1997**, *125*, 159 -- 162.

[225] P. S. Wu, G. Otting. Rapid pulse length determination in high-resolution NMR. *Journal of Magnetic Resonance* **2005**, *176*, 115 -- 119.

[226] R. R. Ernst, *Sensitivity Enhancement in Magnetic Resonance*, Vol. 2 of *Advances in Magnetic and Optical Resonance*, Academic Press, **1966**, pp. 1 -- 135.

[227] Q. Bao, J. Feng, L. Chen, F. Chen, Z. Liu, B. Jiang, C. Liu. A robust automatic phase correction method for signal dense spectra. *Journal of Magnetic Resonance* **2013**, *234*, 82 -- 89.

[228] R. R. Ernst. Numerical Hilbert transform and automatic phase correction in magnetic resonance spectroscopy. *Journal of Magnetic Resonance (1969)* **1969**, *1*, 7--26.

[229] D. E. Brown, T. W. Campbell, R. N. Moore. Automated phase correction of FT NMR spectra by baseline optimization. *Journal of Magnetic Resonance (1969)* **1989**, *85*, 15--23.

[230] F. Montigny, K. Elbayed, J. Brondeau, D. Canet. Automatic phase correction of Fourier-transform NMR data and estimation of peak area by fitting to a Lorentzian shape. *Analytical Chemistry* **1990**, *62*, 864--867.

[231] A. Heuer. A new algorithm for automatic phase correction by symmetrizing lines. *Journal of Magnetic Resonance (1969)* **1991**, *91*, 241--253.

[232] L. L. Sterna, V. P. Tong. Automated method of phasing difficult nuclear magnetic resonance spectra with application to the unsaturated carbon analysis of oils. *Fuel* **1991**, *70*, 941--945.

[233] Z. Džakula. Phase Angle Measurement from Peak Areas (PAMPAS). *Journal of Magnetic Resonance* **2000**, *146*, 20--32.

[234] E. C. Craig, A. G. Marshall. Automated phase correction of FT NMR spectra by means of phase measurement based on dispersion versus absorption relation (DISPA). *Journal of Magnetic Resonance (1969)* **1988**, *76*, 458--475.

[235] L. Chen, Z. Weng, L. Goh, M. Garland. An efficient algorithm for automatic phase correction of NMR spectra based on entropy minimization. *Journal of Magnetic Resonance* **2002**, *158*, 164--168.

[236] G. Balacco, C. Cobas. Automatic phase correction of 2D NMR spectra by a whitening method. *Magnetic Resonance in Chemistry* **2009**, *47*, 322--327.

[237] V. Mäkelä, *ImatraNMR user guide*, **2014**. `http://vltr.fi/imatranmr/`.

[238] G. A. Pearson. A general baseline-recognition and baseline-flattening algorithm. *Journal of Magnetic Resonance (1969)* **1977**, *27*, 265 -- 272.

[239] S. Golotvin, A. Williams. Improved Baseline Recognition and Modeling of FT NMR Spectra. *Journal of Magnetic Resonance* **2000**, *146*, 122 -- 125.

[240] D. Brown. Fully Automated Baseline Correction of 1D and 2D NMR Spectra Using Bernstein Polynomials. *Journal of Magnetic Resonance, Series A* **1995**, *114*, 268 -- 270.

[241] W. Dietrich, C. H. Rüdel, M. Neumann. Fast and precise automatic baseline correction of one- and two-dimensional NMR spectra. *Journal of Magnetic Resonance (1969)* **1991**, *91*, 1 -- 11.

[242] J. C. Cobas, M. A. Bernstein, M. Martín-Pastor, P. G. Tahoces. A new general-purpose fully automatic baseline-correction procedure for 1D and 2D NMR data. *Journal of Magnetic Resonance* **2006**, *183*, 145 -- 151.

[243] A. K. Leung, F. Chau, J. Gao. Wavelet Transform: A Method for Derivative Calculation in Analytical Chemistry. *Analytical Chemistry* **1998**, *70*, 5222--5229.

[244] A. Savitzky, M. J. Golay. Smoothing and differentiation of data by simplified least squares procedures. *Analytical chemistry* **1964**, *36*, 1627--1639.

[245] Q. Bao, J. Feng, F. Chen, W. Mao, Z. Liu, K. Liu, C. Liu. A new automatic baseline correction method based on iterative method. *Journal of Magnetic Resonance* **2012**, *218*, 35 -- 43.

[246] D. Chang, C. D. Banack, S. L. Shah. Robust baseline correction algorithm for signal dense NMR spectra. *Journal of Magnetic Resonance* **2007**, *187*, 288 -- 292.

[247] C. Bartels, P. Güntert, K. Wüthrich. IFLAT - A New Automatic Baseline-Correction Method for Multidimensional NMR Spectra with Strong Solvent Signals. *Journal of Magnetic Resonance, Series A* **1995**, *117*, 330 -- 333.

[248] K. Wang, S. Wang, C. Kuo, Y. J. Tseng. Distribution-Based Classification Method for Baseline Correction of Metabolomic 1D Proton Nuclear Magnetic Resonance Spectra. *Analytical Chemistry* **2013**, *85*, 1231--1239.

[249] Z. Zhang, S. Chen, Y. Liang. Baseline correction using adaptive iteratively reweighted penalized least squares. *Analyst* **2010**, *135*, 1138--1146.

[250] Z. Zolnai, S. Macura, J. L. Markley. Spline method for correcting baseplane distortions in two-dimensional NMR spectra. *Journal of Magnetic Resonance (1969)* **1989**, *82*, 496 -- 504.

[251] P. Eilers. A perfect smoother. *Analytical Chemistry* **2003**, *75*, 3631--3636.

[252] M. Friedrichs. A model-free algorithm for the removal of baseline artifacts. *Journal of Biomolecular NMR* **1995**, *5*, 147--153.

[253] D. Hoult, C.-N. Chen, H. Eden, M. Eden. Elimination of baseline artifacts in spectra and their integrals. *Journal of Magnetic Resonance (1969)* **1983**, *51*, 110 -- 117.

[254] R. Kyburz. What Happens After the Last Pulse? *Magnetic Moments* **1995**, *VII.1 Spring*.

[255] P. Plateau, C. Dumas, M. Guéron. Solvent-peak-suppressed NMR: Correction of baseline distortions and use of strong-pulse excitation. *Journal of Magnetic Resonance (1969)* **1983**, *54*, 46 -- 53.

[256] C. Tang. An Analysis of Baseline Distortion and Offset in NMR Spectra. *Journal of Magnetic Resonance, Series A* **1994**, *109*, 232 -- 240.

[257] *VNMR Command and Parameter Reference*, Varian, Inc., **2001**, Varian NMR Spectrometer Systems With VNMR 6.1C Software.

[258] D. Marion, A. Bax. Baseline correction of 2D FT NMR spectra using a simple linear prediction extrapolation of the time-domain data. *Journal of Magnetic Resonance (1969)* **1989**, *83*, 205 -- 211.

[259] A. Heuer, U. Haeberlen. A new method for suppressing baseline distortions in FT NMR. *Journal of Magnetic Resonance (1969)* **1989**, *85*, 79 -- 94.

[260] G. Otting, H. Widmer, G. Wagner, K. Wüthrich. Origin of t_1 and t_2 ridges in 2D NMR spectra and procedures for suppression. *Journal of Magnetic Resonance (1969)* **1986**, *66*, 187 -- 193.

[261] N.-P. V. Nielsen, J. M. Carstensen, J. Smedsgaard. Aligning of single and multiple wavelength chromatographic profiles for chemometric data analysis using correlation optimised warping. *Journal of Chromatography A* **1998**, *805*, 17 -- 35.

[262] G. Tomasi, F. van den Berg, C. Andersson. Correlation optimized warping and dynamic time warping as preprocessing methods for chromatographic data. *Journal of Chemometrics* **2004**, *18*, 231--241.

[263] F. H. Larsen, F. van den Berg, S. B. Engelsen. An exploratory chemometric study of ^{1}H NMR spectra of table wines. *Journal of Chemometrics* **2006**, *20*, 198--208.

[264] J. Forshed, I. Schuppe-Koistinen, S. P. Jacobsson. Peak alignment of NMR signals by means of a genetic algorithm. *Analytica Chimica Acta* **2003**, *487*, 189 -- 199.

[265] G.-C. Lee, D. L. Woodruff. Beam search for peak alignment of NMR signals. *Analytica Chimica Acta* **2004**, *513*, 413--416.

[266] R. J. O. Torgrip, M. Åberg, B. Karlberg, S. P. Jacobsson. Peak alignment using reduced set mapping. *Journal of Chemometrics* **2003**, *17*, 573--582.

[267] R. Torgrip, J. Lindberg, M. Linder, B. Karlberg, S. Jacobsson, J. Kolmert, I. Gustafsson, I. Schuppe-Koistinen. New modes of data partitioning based on PARS peak alignment for improved multivariate biomarker/biopattern detection in ^{1}H-NMR spectroscopic metabolic profiling of urine. *Metabolomics* **2006**, *2*, 1--19.

[268] J. W. H. Wong, C. Durante, H. M. Cartwright. Application of Fast Fourier Transform Cross-Correlation for the Alignment of Large Chromatographic and Spectral Datasets. *Analytical Chemistry* **2005**, *77*, 5655--5661.

[269] K. A. Veselkov, J. C. Lindon, T. M. Ebbels, D. Crockford, V. V. Volynkin, E. Holmes, D. B. Davies, J. K. Nicholson. Recursive segment-wise peak alignment of biological ^{1}H NMR spectra for improved metabolic biomarker recovery. *Analytical Chemistry* **2008**, *81*, 56--66.

[270] F. Savorani, G. Tomasi, S. Engelsen. icoshift: A versatile tool for the rapid alignment of 1D NMR spectra. *Journal of Magnetic Resonance* **2010**, *202*, 190 -- 202.

[271] T. Vu, D. Valkenborg, K. Smets, K. Verwaest, R. Dommisse, F. Lemiere, A. Verschoren, B. Goethals, K. Laukens. An integrated workflow for robust alignment and simplified quantitative analysis of NMR spectrometry data. *BMC Bioinformatics* **2011**, *12*, 405.

[272] M. Spraul, P. Neidig, U. Klauck, P. Kessler, E. Holmes, J. Nicholson, B. Sweatman, S. Salman, R. Farrant, E. Rahr, C. Beddell, J. Lindon. Automatic reduction of NMR spectroscopic data for statistical and pattern recognition classification of samples. *Journal of Pharmaceutical and Biomedical Analysis* **1994**, *12*, 1215 -- 1225.

[273] J. S. McKenzie, J. A. Donarski, J. C. Wilson, A. J. Charlton. Analysis of complex mixtures using high-resolution nuclear magnetic resonance spectroscopy and chemometrics. *Progress in Nuclear Magnetic Resonance Spectroscopy* **2011**, *59*, 336 -- 359.

[274] R. A. Davis, A. J. Charlton, J. Godward, S. A. Jones, M. Harrison, J. C. Wilson. Adaptive binning: An improved binning method for metabolomics data using the undecimated wavelet transform. *Chemometrics and Intelligent Laboratory Systems* **2007**, *85*, 144 -- 154.

[275] T. De Meyer, D. Sinnaeve, B. Van Gasse, E. Tsiporkova, E. R. Rietzschel, M. L. De Buyzere, T. C. Gillebert, S. Bekaert, J. C. Martins, W. Van Criekinge. NMR-Based Characterization of Metabolic Alterations in Hypertension Using an Adaptive, Intelligent Binning Algorithm. *Analytical Chemistry* **2008**, *80*, 3783--3790.

[276] P. Anderson, D. Mahle, T. Doom, N. Reo, N. DelRaso, M. Raymer. Dynamic adaptive binning: an improved quantification technique for NMR spectroscopic data. *Metabolomics* **2011**, *7*, 179--190.

[277] P. E. Anderson, N. V. Reo, N. J. DelRaso, T. E. Doom, M. L. Raymer. Gaussian binning: a new kernel-based method for processing NMR spectroscopic data for metabolomics. *Metabolomics* **2008**, *4*, 261--272.

[278] S. Sousa, A. Magalhães, M. M. C. Ferreira. Optimized bucketing for NMR spectra: Three case studies. *Chemometrics and Intelligent Laboratory Systems* **2013**, *122*, 93 -- 102.

[279] F. Dieterle, A. Ross, G. Schlotterbeck, H. Senn. Probabilistic Quotient Normalization as Robust Method to Account for Dilution of Complex Biological Mixtures. Application in ^{1}H NMR Metabonomics. *Analytical Chemistry* **2006**, *78*, 4281--4290.

[280] T. De Meyer, D. Sinnaeve, B. Van Gasse, E.-R. Rietzschel, M. De Buyzere, M. Langlois, S. Bekaert, J. Martins, W. Van Criekinge. Evaluation of standard and advanced preprocessing methods for the univariate analysis of blood serum ^{1}H-NMR spectra. *Analytical and Bioanalytical Chemistry* **2010**, *398*, 1781--1790.

[281] S. Kohl, M. Klein, J. Hochrein, P. Oefner, R. Spang, W. Gronwald. State-of-the art data normalization methods improve NMR-based metabolomic analysis. *Metabolomics* **2012**, *8*, 146--160.

[282] B. Bolstad, R. Irizarry, M. Åstrand, T. Speed. A comparison of normalization methods for high density oligonucleotide array data based on variance and bias. *Bioinformatics* **2003**, *19*, 185--193.

[283] J. Olivero, R. Longbothum. Empirical fits to the Voigt line width: A brief review. *Journal of Quantitative Spectroscopy and Radiative Transfer* **1977**, *17*, 233 -- 236.

[284] P. Soininen, J. Haarala, J. Vepsäläinen, M. Niemitz, R. Laatikainen. Strategies for organic impurity quantification by ^{1}H NMR spectroscopy: Constrained total-line-shape fitting. *Analytica Chimica Acta* **2005**, *542*, 178 -- 185.

[285] R. Laatikainen, M. Niemitz, U. Weber, J. Sundelin, T. Hassinen, J. Vepsäläinen. General Strategies for Total-Lineshape-Type Spectral Analysis of NMR Spectra Using Integral-Transform Iterator. *Journal of Magnetic Resonance, Series A* **1996**, *120*, 1--10.

[286] M. A. Bernstein, S. Sýkora, C. Peng, A. Barba, C. Cobas. Optimization and Automation of Quantitative NMR Data Extraction. *Analytical Chemistry* **2013**, *85*, 5778--5786.

[287] A. M. Weljie, J. Newton, P. Mercier, E. Carlson, C. M. Slupsky. Targeted Profiling: Quantitative Analysis of ^{1}H NMR Metabolomics Data. *Analytical Chemistry* **2006**, *78*, 4430--4442.

[288] P. Mercier, M. Lewis, D. Chang, D. Baker, D. Wishart. Towards automatic metabolomic profiling of high-resolution one-dimensional proton NMR spectra. *Journal of Biomolecular NMR* **2011**, *49*, 307--323.

[289] J. Hao, W. Astle, M. De Iorio, T. Ebbels. BATMAN—an R package for the automated quantification of metabolites from NMR spectra using a Bayesian Model. *Bioinformatics* **2012**, *28*, 2088--2090.

[290] J. Hao, M. Liebeke, W. Astle, M. De Iorio, J. G. Bundy, T. M. Ebbels. Bayesian deconvolution and quantification of metabolites in complex 1D NMR spectra using BATMAN. *Nature protocols* **2014**, *9*, 1416--1427.

[291] W. Astle, M. De Iorio, S. Richardson, D. Stephens, T. Ebbels. A Bayesian model of NMR spectra for the deconvolution and quantification of metabolites in complex biological mixtures. *Journal of the American Statistical Association* **2012**, *107*, 1259--1271.

[292] A. Vehtari, V.-P. Mäkinen, P. Soininen, P. Ingman, S. Mäkelä, M. Savolainen, M. Hannuksela, K. Kaski, M. Ala-Korpela. A novel Bayesian approach to quantify clinical variables and to determine their spectroscopic counterparts in ^{1}H NMR metabonomic data. *BMC Bioinformatics* **2007**, *8*, S8.

[293] S. Ravanbakhsh, P. Liu, T. C. Bjordahl, R. Mandal, J. R. Grant, M. Wilson, R. Eisner, I. Sinelnikov, X. Hu, C. Luchinat, R. Greiner, D. S. Wishart. Accurate, Fully-Automated NMR Spectral Profiling for Metabolomics. *PLoS ONE* **2015**, *10*, e0124219.

[294] D. V. Rubtsov, J. L. Griffin. Time-domain Bayesian detection and estimation of noisy damped sinusoidal signals applied to NMR spectroscopy. *Journal of Magnetic Resonance* **2007**, *188*, 367 -- 379.

[295] K. Krishnamurthy. CRAFT (complete reduction to amplitude frequency table) – robust and time-efficient Bayesian approach for quantitative mixture analysis by NMR. *Magnetic Resonance in Chemistry* **2013**, *51*, 821--829.

[296] D. V. Rubtsov, C. Waterman, R. A. Currie, C. Waterfield, J. D. Salazar, J. Wright, J. L. Griffin. Application of a Bayesian Deconvolution Approach for High-Resolution ^{1}H NMR Spectra to Assessing the Metabolic Effects of Acute Phenobarbital Exposure in Liver Tissue. *Analytical Chemistry* **2010**, *82*, 4479--4485.

[297] L. Brennan. NMR-based metabolomics: From sample preparation to applications in nutrition research. *Progress in Nuclear Magnetic Resonance Spectroscopy* **2014**, *83*, 42 -- 49.

[298] H. K. Kim, Y. H. Choi, R. Verpoorte. NMR-based metabolomic analysis of plants. *Nature Protocols* **2010**, *5*, 536--549.

[299] L. Eriksson, T. Byrne, E. Johansson, J. Trygg, C. Vikström, *Multi- and Megavariate Data Analysis: Basic Principles and Applications 3rd ed.*, Umetrics Academy, **2013**.

[300] J. E. Jackson, *A user's guide to principal components*, *Vol. 587*, John Wiley & Sons, **1991**.

[301] S. Wold, M. Sjöström, L. Eriksson. PLS-regression: a basic tool of chemometrics. *Chemometrics and Intelligent Laboratory Systems* **2001**, *58*, 109 -- 130.

[302] K. Marat, *SpinWorks*, University of Manitoba. `http://home.cc.umanitoba.ca/~wolowiec/spinworks/`.

[303] I. A. Lewis, S. C. Schommer, J. L. Markley. rNMR: open source software for identifying and quantifying metabolites in NMR spectra. *Magnetic Resonance in Chemistry* **2009**, *47*, S123--S126.

[304] J. Xia, T. Bjorndahl, P. Tang, D. Wishart. MetaboMiner - semi-automated identification of metabolites from 2D NMR spectra of complex biofluids. *BMC Bioinformatics* **2008**, *9*, 507.

[305] C. Ludwig, U. Gunther. MetaboLab - advanced NMR data processing and analysis for metabolomics. *BMC Bioinformatics* **2011**, *12*, 366.

[306] M. A. Fitzpatrick, C. M. McGrath, S. P. Young. Pathomx: an interactive workflow-based tool for the analysis of metabolomic data. *BMC bioinformatics* **2014**, *15*, 396.

[307] E. L. Ulrich, H. Akutsu, J. F. Doreleijers, Y. Harano, Y. E. Ioannidis, J. Lin, M. Livny, S. Mading, D. Maziuk, Z. Miller, E. Nakatani, C. F. Schulte, D. E. Tolmie, R. Kent Wenger, H. Yao, J. L. Markley. BioMagResBank. *Nucleic Acids Research* **2008**, *36*, D402--D408.

[308] D. S. Wishart, C. Knox, A. C. Guo, R. Eisner, N. Young, B. Gautam, D. D. Hau, N. Psychogios, E. Dong, S. Bouatra, R. Mandal, I. Sinelnikov, J. Xia, L. Jia, J. A. Cruz, E. Lim, C. A. Sobsey, S. Shrivastava, P. Huang, P. Liu, L. Fang, J. Peng, R. Fradette, D. Cheng, D. Tzur, M. Clements, A. Lewis, A. De Souza, A. Zuniga, M. Dawe, Y. Xiong, D. Clive, R. Greiner, A. Nazyrova, R. Shaykhutdinov, L. Li, H. J. Vogel, I. Forsythe. HMDB: a knowledgebase for the human metabolome. *Nucleic Acids Research* **2009**, *37*, D603--D610.

[309] D. Tulpan, S. Leger, L. Belliveau, A. Culf, M. Cuperlovic-Culf. MetaboHunter: an automatic approach for identification of metabolites from ^{1}H-NMR spectra of complex mixtures. *BMC Bioinformatics* **2011**, *12*, 400.

[310] J. Xia, I. V. Sinelnikov, B. Han, D. S. Wishart. MetaboAnalyst 3.0 — making metabolomics more meaningful. *Nucleic Acids Research* **2015**.

[311] F. Delaglio, S. Grzesiek, G. Vuister, G. Zhu, J. Pfeifer, A. Bax. NMRPipe: A multidimensional spectral processing system based on UNIX pipes. *Journal of Biomolecular NMR* **1995**, *6*, 277--293.

[312] MathWorks Inc., Natick, Massachusetts, United States., *MATLAB*, **2015**. `http://mathworks.com`.

[313] B. Worley, R. Powers. MVAPACK: A Complete Data Handling Package for NMR Metabolomics. *ACS Chemical Biology* **2014**, *9*, 1138--1144, PMID: 24576144.

[314] J. J. Helmus, C. P. Jaroniec. Nmrglue: an open source Python package for the analysis of multidimensional NMR data. *Journal of Biomolecular NMR* **2013**, *55*, 355--367.