Data Processing and Refinement

Report for JASS 2006

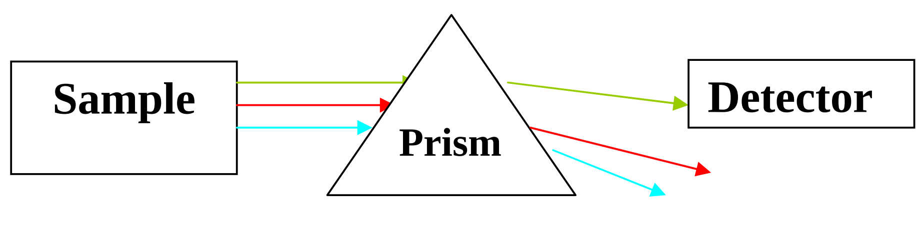
Philipp Schapotschnikow

April 06

Introduction

The human body has two mechanisms for transforming waves into information suitable for the brain. The first one is seeing, the second one is hearing. Although we obtain most of our daily information through seeing, hearing is much more sensitive. In the sequel we will discuss superficially both mechanisms and compare them with the two ways of performing spectroscopy: continuous waves and pulse.
Our eyes have “detectors” (cone cells) for the three basic colours: blue, green and red. Roughly speaking, each of them measures the intensity of light of the corresponding frequency, and different mixtures of those intensities correspond to the different colours we see. On the other hand, we can recognize well the source of light of a certain colour. For this reason we can see two- and three-dimensional objects (as light sources). All together we have a relatively poor frequency resolution and a high spatial resolution. This is a typical example of analogous signal processing.
Our ear on the other hand is able to decompose the mixture of sound waves. We are able to hear and distinguish several voices at the same time. In a noisy underground train we can understand what the neighbour is saying and hear a baby crying at the opposite end of the wagon. In a piece of music we can hear several instruments. But we can hardly tell the direction where the noise comes from, nothing to say about the shape of the source. Summing up, we have a high resolution in frequency and a very poor spatial resolution. This is a typical example of digital signal processing.

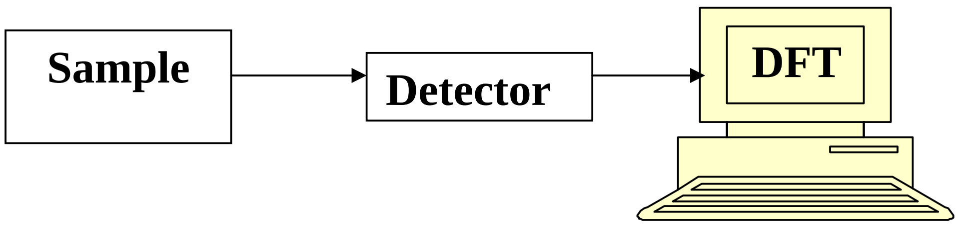
In spectroscopy we either measure the intensity of light sent through the probe (absorption) or the intensity of light coming from the sample after it was excited. In both cases we obtain a signal, which is a linear superposition of several waves, and we would like to know the basic frequencies and in which proportion these waves occur. These different light rays can be physically separated with a monochromator, e.g. a prism, and then focussed on a detector by a lens.



Then the detector measures the intensity of light of the certain “colour” and works therefore similarly to the cone cells in our eyes. That means, that the experiment must be repeated for each frequency, until the desired resolution and band width are reached.
Another way of detecting the spectrum frequency-wise is continuous waves method (CW). Instead of separating light with different frequencies one makes sure that the sample produces monochromatic light.

Until 1965 NMR was performed using CW. But it has serious drawbacks. One needs a high resolution to make sure that no peak was skipped. Hence, the experiment is time-intensive.

Ernst and Anderson suggested 1965 to detect the signal as it is. They needed the “ear” to transform the complicated signal into frequencies and amplitudes. This procedure was well known as Fourier transform and had to be computed by a machine.



With this method, all frequencies in a required range can be resolved simultaneously with one experiment.

Fourier Transform

Assume that we have recorded a signal $S(t)$ on the sampling interval [0,T] of the form

$$S(t) = b_1 \sin(\frac{2\pi}{T} t) + a_2 \cos(\frac{2\pi}{T} 2x) + b_5 \sin(\frac{2\pi}{T} 5t) .$$

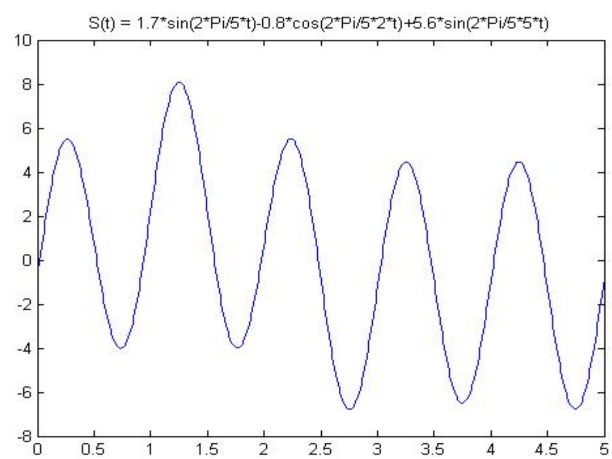


Figure 1: An example signal

Now we would like to know the coefficient b_1, which describes the part of the frequency $\frac{1}{T}$ on our signal. This task can be solved with the following trick:

$$\int_0^T S(t)\sin(\frac{2\pi}{T} t)dt = \int_0^T b_1 \sin^2(\frac{2\pi}{T} t)dt + \int_0^T a_2 \sin(\frac{2\pi}{T} t)\cos(\frac{2\pi}{T} 2x)dt + \int_0^T b_5 \sin(\frac{2\pi}{T} t)\sin(\frac{2\pi}{T} 5x)dt =$$

$$\frac{T}{2} b_1 + 0 + 0.$$

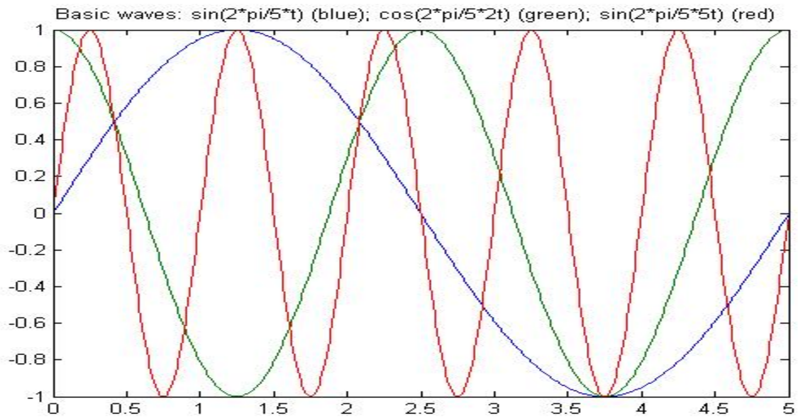


Figure 2: Basic waves

This trick works only for frequencies of the kind $\frac{k}{T}$ where k is an integer. Otherwise the disturbing integrals do not vanish nicely. From these simple calculations we have learned an important principle:

> If the signal is recorded over a time period T, then the resolved frequencies are discrete and of the form $\frac{k}{T}$

This means that with increasing time of measurement we expect a better result, and in the limit case of infinite time we can resolve all frequencies. But usually signals become weaker with time, and noise prevails. In this case, increasing the resolution we do not increase the accuracy.

What happens if the signal S has a component of the form $a_1 \cos(\frac{2\pi}{T}t)$? Then the frequency $\frac{1}{T}$ contributes with the pair (a_1, b_1) to the signal. Computing the integral as above we would still obtain 0 because $\int_0^T \sin(\frac{2\pi}{T}t)\cos(\frac{2\pi}{T}t)dt = 0.$ In order to find a_1 we should calculate $\int_0^T S(t)\cos(\frac{2\pi}{T}t)dt = \frac{T}{2}a_1.$

Now we know how to extract coefficients out of (finite) sums of sine's and cosines with frequency $\frac{k}{T}$. The practical problem is calculating the integrals; we will deal with it later.

Conversely, the coefficients (for k=1,2,….)

$\frac{2}{T}\int_0^T S(t)\cos(\frac{2\pi}{T}kt)dt = a_k$; $\frac{2}{T}\int_0^T S(t)\sin(\frac{2\pi}{T}kt)dt = b_k$ and $\frac{1}{T}\int_0^T S(t)dt = a_0$

can be computed for any bounded integrable function S. Does it mean that any function R(t) with computable coefficients $(a_0, a_1, b_1, a_2, b_2, ...)$ can be represented over the interval [0,T] as the infinite sum?

$$R(t) = a_0 + a_1 \cos(\frac{2\pi}{T}t) + b_1 \sin(\frac{2\pi}{T}t) + a_2 \cos(\frac{2\pi}{T}2t) + b_2 \sin(\frac{2\pi}{T}2t) + ...$$

The answer is yes, and this representation is called sine/cosine Fourier series, see [4] for further details. Hence, there exists a one-to-one correspondence between sequences $(a_0, a_1, b_1, a_2, b_2, ...)$ and functions R(t).

We still have the problem that a single frequency is represented by two real numbers. Recall, that a pair of real numbers a, b can be combined to complex numbers $c_+ = \frac{1}{2}(a - bi)$ and $c_- = \frac{1}{2}(a + bi)$ where i is the imaginary unit, $i^2 = -1$. Then $a = c_+ + c_-$ and $b = i(c_+ - c_-)$.

The reason of this odd labelling will become clear later. We can now plug it in sine/cosine Fourier series putting $c_0 = a_0$ and using de Moivre's formula $\exp(ix) = \cos(x) + i\sin(x)$:

$$R(t) = a_0 + a_1 \cos(\frac{2\pi}{T} t) + b_1 \sin(\frac{2\pi}{T} t) + a_2 \cos(\frac{2\pi}{T} 2t) + b_2 \sin(\frac{2\pi}{T} 2t) + ... =$$

$$c_0 + (c_{+1} + c_{-1}) \cos(\frac{2\pi}{T} t) + i(c_{+1} - c_{-1}) \sin(\frac{2\pi}{T} t) + (c_{+1} + c_{-1}) \cos(\frac{2\pi}{T} 2t) + i(c_{+2} - c_{-2}) \sin(\frac{2\pi}{T} 2t) + ...$$

$$= c_0 + c_{+1}(\cos(\frac{2\pi}{T} t) + i\sin(\frac{2\pi}{T} t)) + c_{-1}(\cos(\frac{2\pi}{T} t) - i\sin(\frac{2\pi}{T} t)) +$$

$$c_{+2}(\cos(\frac{2\pi}{T} 2t) + i\sin(\frac{2\pi}{T} 2t)) + c_{-2}(\cos(\frac{2\pi}{T} 2t) - i\sin(\frac{2\pi}{T} 2t)) + ... =$$

$$c_0 + c_{+1} \exp(+i\frac{2\pi}{T} t) + c_{-1} \exp(-i\frac{2\pi}{T} t) + c_{+2} \exp(+i\frac{2\pi}{T} 2t) + c_{-2} \exp(-i\frac{2\pi}{T} 2t) + ...$$

The coefficients $c_{\pm k}$ are called Fourier coefficients and the series representation accordingly Fourier series. They can be calculated as

$$c_{\pm k} = \frac{1}{T} \int_0^T R(t) \exp(\mp ikt) dt .$$

Before moving on to DFT, we should broach the issue of the double continuous Fourier transform. Assume that our signal cannot be represented by an infinite sum of countably many sine- and cosine-waves. Then it can still be represented as a weighted sum of all possible frequencies; k becomes continuous and a sum becomes an integral:

$$R(t) = \frac{1}{\sqrt{2\pi}} \int_{-\infty}^{\infty} c(k) \exp(ikt) dk . \qquad (4)$$

The weight function $c(k)$ can be calculated by the Fourier transform

$$c(k) = \frac{1}{\sqrt{2\pi}} \int_{-\infty}^{\infty} R(t) \exp(-ikt) dt . \qquad (5)$$

Discrete Fourier Transform

In general, continuous signals cannot be measured and continuous functions cannot be processed. Assume that the signals is received on N time points with constant intervals Δt and the first time point is zero. For a reason which will be explained below, we can assume that $T = N\Delta t$.

As we have seen in the previous section, finite time range is equivalent to discrete frequencies. Conversely, discrete time is equivalent to finite frequency range. The next question will be, how large the frequency range can be. From N pieces of information we can determine N Fourier coefficients. To each frequency except 0 correspond 2 coefficients. Thus, with N measurements we can resolve exactly $\frac{N+1}{2}$ frequencies. Since our resolution is $\frac{1}{T}=\frac{1}{N\Delta t}$, we obtain a broad width of $\frac{N-1}{2}\cdot\frac{1}{N\Delta t}\approx\frac{1}{2\Delta t}$. This has an important consequence:

Band width only depends on the time resolution and not on the time range

It remains to compute the integrals in (3). As we know the function R only at the points $n\Delta t$, where $n=0,1,..,N-1$, we can only apply quadrature:

$$c_{\pm k}=\frac{1}{T}\int_0^T R(t)\exp(\mp ikt)dt\approx\frac{1}{T}\sum_{n=0}^{N-1}R(n\Delta t)\exp(\mp ikn\Delta t)\Delta t \tag{6}$$

The approach in (6) provides a $\frac{1}{N^2}$-order approximation (midpoint rule) for an integral over the interval $[-\frac{\Delta t}{2},N-\frac{\Delta t}{2}]$, and this is the promised explanation for $T=N\Delta t$. Recall that $R(n\Delta t)$ are exactly the measured data, denote them by $r_0,r_1,...,r_{N-1}$ After rescaling time and frequency we obtain the Discrete Fourier Transform (DFT):

$$c_k=\sum_{n=0}^{N-1}\exp(-i2\pi kn/N) \tag{7}$$

where k may be positive or negative.

Consider $k'=k+mN$ where m is an integer. Then $c_k=c_{k'}$ always because

$$\exp(-i2\pi kn/N-i2\pi m)=\exp(-i2\pi kn/N)\exp(-i2\pi m)=\exp(-i2\pi kn/N). \tag{8}$$

Hence, DFT has one very important property: it is periodic in the frequency domain. This phenomenon is called aliasing. It implies that a high-frequency signal may appear as its alias if we do not choose our time resolution small enough.

Why is the button called FFT?

FFT stands for Fast Fourier Transform. It provides a method for a quick computation of DFT, but <u>not</u> an alternative transform!

If we calculate Fourier coefficients straight forward as in (7), we will require N operations for each c_k and, correspondingly, N^2 operations for the entire transform. Rearranging the terms and using the structure of (7), however, the costs can be reduced to $C\cdot N\log(N)$ where C is a constant. This is much less, especially for large N. We will not discuss in detail how it works. It should only be mentioned that modern FFT codes are perfect in terms of both underlying mathematics and data structures; the best way to perform DFT is using a FFT code [5].

Note that FFT only works for powers of 2 (N=2,4,8,…) This will be achieved by zero-filling.

Examples

We perform FFT to the Figure 1. The result is shown below:

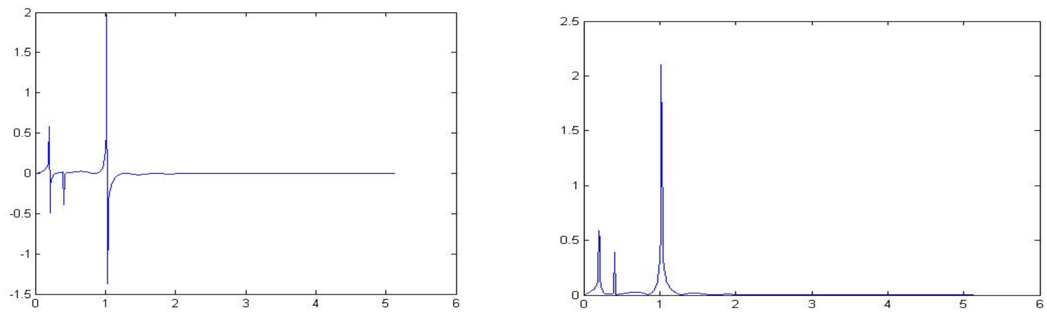


Figure 3: The real part (left) and the modulus (right) of the FFT of Figure 1

The function was sampled on 1024 points with resolution $\Delta t = 0.1$. The Fourier transform consists in 1024 complex numbers, but only the first 513 are important, the rest is redundant. Obviously our time resolution was too good: for frequencies larger than 2 nothing happens. The time range seems to satisfy.
As we can see, the three waves in the example can be identified with the three peaks, the frequencies are found correctly. In the real part we can see for the first and third peak a rapid sign change to a negative peak. This is sometimes called Lorentzian shape.
The real part gives us some information about the phase of the signal. A sharp Lorentzian indicates a sine, and no sign change as for the second peak indicates a cosine. This phenomenon is related to the complex representation of the Fourier coefficients. The phase of a signal is relevant if we want to trace back the birth of a wave.
The modulus of the FFT tells us in which proportion which frequencies contribute to the signal. This information is crucial for spectroscopy.

Uncertainty principle for DFT

We finish the theory of NMR data processing with a brief discussion of uncertainty. From

$$\Delta f = \frac{1}{T} = \frac{1}{N\Delta t}$$

we immediately conclude

$$\Delta t \Delta f = \frac{1}{N}.$$

This is the uncertainty principle for Fourier analysis. It means that having a finite number of measurements we cannot resolve time and frequency with an arbitrary accuracy at the same time. Moreover, this relation remains unchanged after refining manipulations of data. We cannot improve the bandwidth and frequency resolution without making new measurements.

Decaying noisy signals

Unlike our example, NMR signals exponentially decay due to the exponential decay of excited states. Additionally, we have to analyse effects of noise, which is unavoidable.
For this purpose we multiply the function on Figure 1 with a slowly decaying exponential and add some noise.

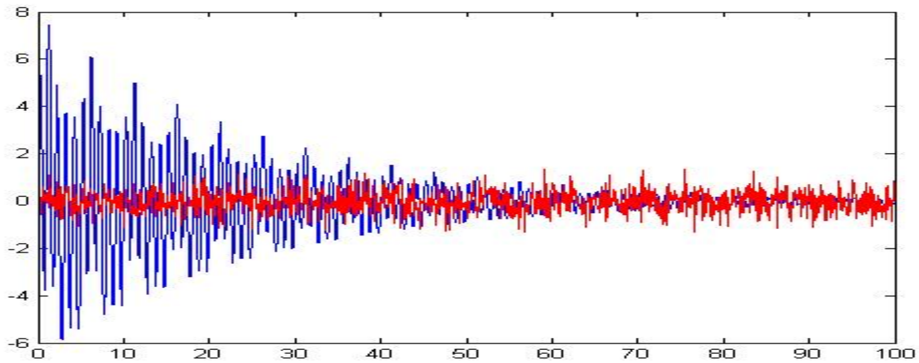


Figure 4: A regular damped signal (blue) and noise (red)

As we can see, at the beginning the true signal dominates while at later time points noise prevails. The FFT of the sum of the blue and red lines is the following:

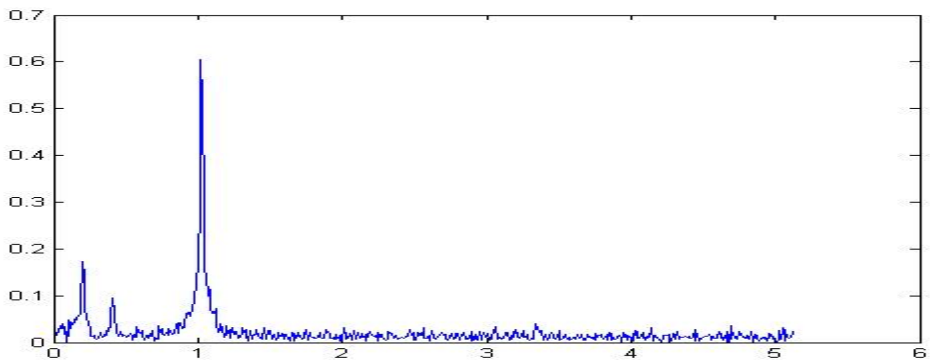


Figure 5: FFT of a noisy signal (modulus)

Compare Figures 3 and 5. Due to the exponential decay, the proportion of peak heights is changed. Noise can be clearly separated from the spectrum. Now if we know that our spectrum lies between 0 and 2, we can put the rest 0 and perform an inverse transform. This procedure is implicitly done by a digital filter (also called sinc-filter).

Quadrature detector

In NMR oscillations are created in both x and y directions. Hence the incoming signal is complex. It has to be compared with a reference signal first in a device called quadrature detector. It contains two doubly balanced mixers, two amplifiers and one 90° phase shifter.

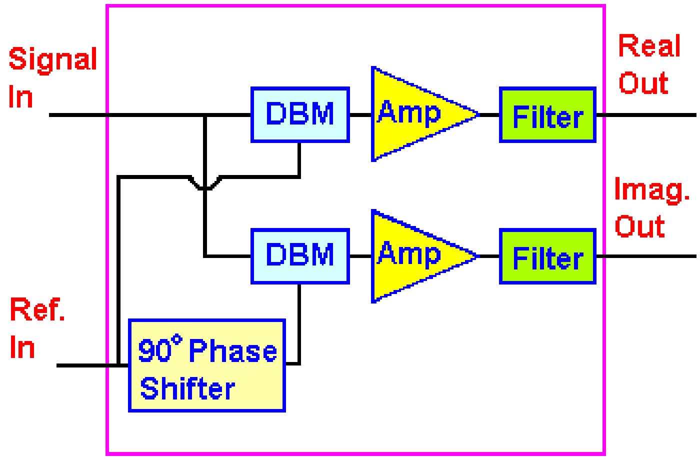


Figure 6: Quadrature detector, © see [3]

The DBM has two inputs and one output. When the input signals are $\cos(A)$ and $\cos(B)$, then the output will be $\cos(A)\cos(B) = \frac{1}{2}\cos(A+B) + \frac{1}{2}\cos(A-B)$.

The digital filters reduce noise and cut out unnecessary frequency intervals.

First point extrapolation

For technical reasons it is impossible to measure the NMR signal from the very beginning of the relaxation. Therefore the first data point must be extrapolated from succeeding data, usually via linear extrapolation.

Adopization

As we can see on Figure 4, data at the beginning are more accurate than at the end. Therefore it makes sometimes sense to weight them more, for instance by multiplying the measured data with a decaying function. This leads to resolution enhancement, but may improve the spectrum. See [1] for details.

Zero filling

In order to smoothen lines in the spectrum, one can add some zeros to the end of the measured data. With this method one can ensure that the data have the for FFT required length of 2^m. Caution! Zero filling does not improve the quality of the spectrum, it does not add or remove peaks. It only makes the spectrum look nicer.

Base line correction

If in the spectrum the background noise is not completely random but follows some pattern (baseline), it has simply to be subtracted in the frequency domain. Automatic recognition of a baseline is a hard task and is not solved yet completely. A method based on linear prediction is presented in [1], another method is presented in [2].

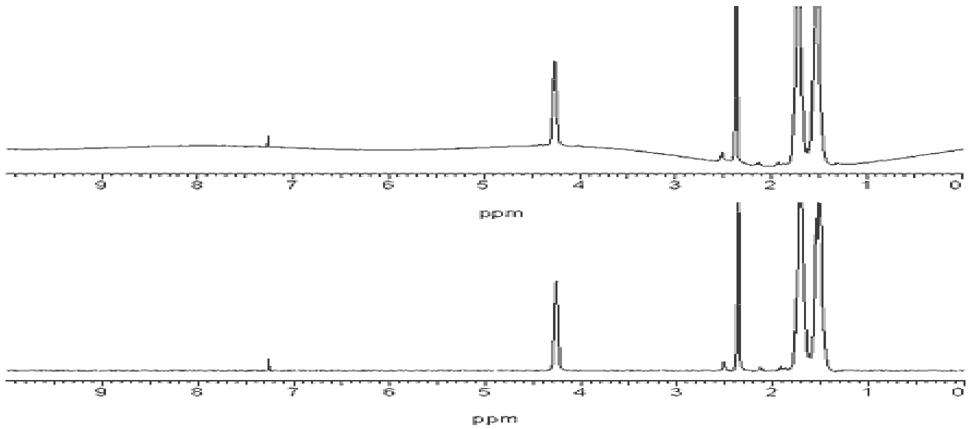


Figure 7: Baseline correction, © see [2]

Repeated Measurements

The best method ever to reduce random noise occurring in an every experiment is repeating the experiment for many times and averaging. Unfortunately, the improvement becomes worse with an increasing number of experiments: the deviation of the noise from 0 converges to 0 with a rate of $N^{-0.5}$. This means that doing the experiment 100 times improves makes the result 10 times more precise. This effect is demonstrated below:

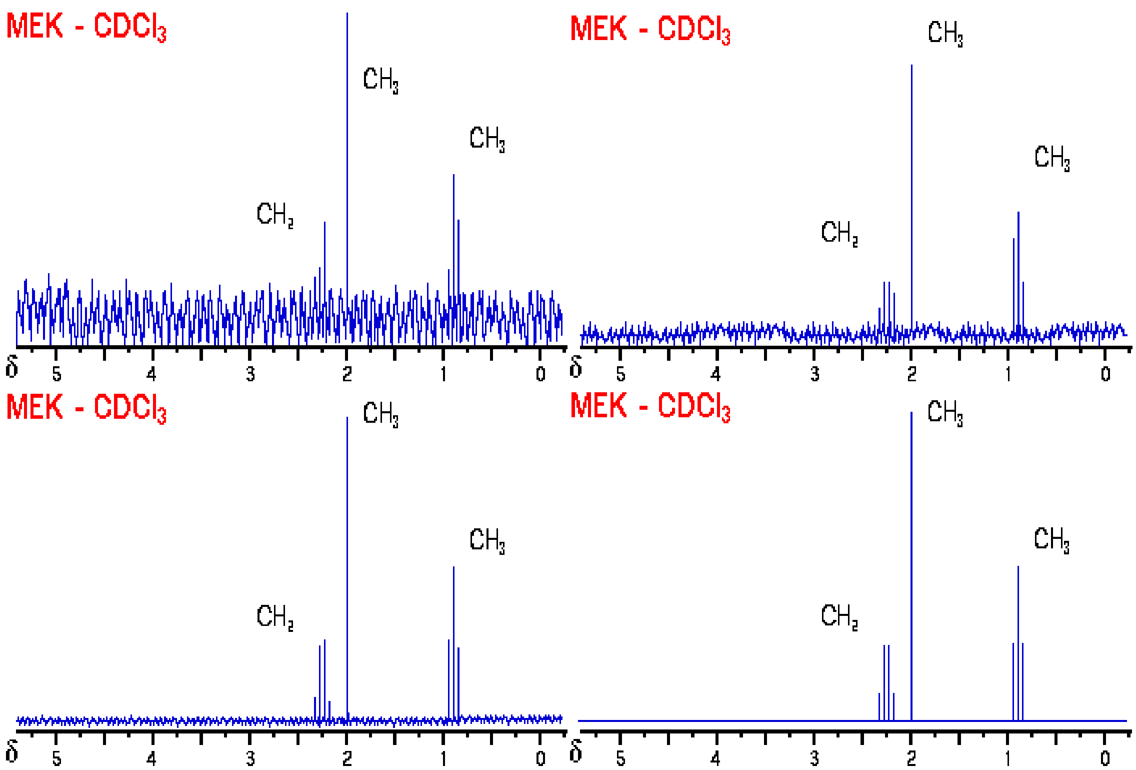


Figure 8: An averaged NMR spectrum with 1, 8, 80, 800 measurements (from left to right and top to bottom) , © see [3]

References:

[1] Hoch, Stern, NMR DATA PROCESSING, 1996, Wiley-Liss
[2] Golotvin, Williams, 1999, http://www.acdlabs.com/publish/nmr_ar.html
[3] Hornak, The Basics of NMR, 1997, http://www.cis.rit.edu/htbooks/nmr/
[4] Nikolsky, Functional analysis, 2001, Fizmatlit (in Russian)
[5] Verzhbitsky, Foundations of Numerical Methods, 2002, V. Shkola (in Russian)