Data Acquisition and Processing
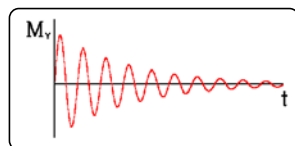
Eugene E. Kwan

February 7, 2012.
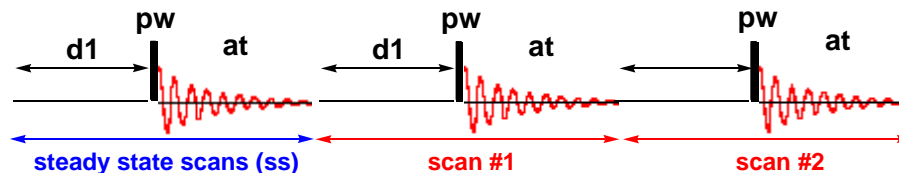
Scope of Lecture

sample preparation
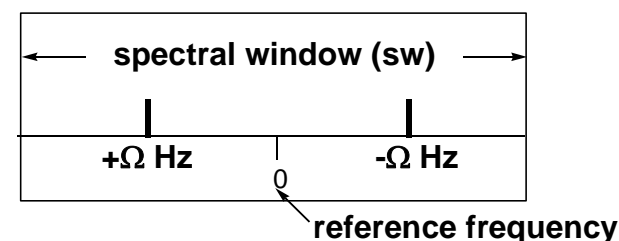
CYCLOPS

locking and shimming

the Nyquist theorem

data acquisition and processing

pulse-acquire

Ernst angle

the ADC and dynamic range

quadrature detection

phase correction

the Fourier transform

Key References

1. Understanding NMR Spectroscopy (2nd Ed.) Keeler, J. Wiley, **2010**. (Chapters 4 and 5)

2. NMR Spectroscopy Explained: Simplified Theory... Jacobsen, N.E. Wiley, **2007**. (Chapter 3)

3. The ABCs of FT-NMR Roberts, J.D. University Science Books, **2000**. (Chapters 3 and 5)

4. High-Resolution NMR Techniques in Organic Chemistry (2nd Ed.) Claridge, T.D.W. Elsevier, **2009**. (Chapter 3)

5. Mathematical Methods for Physicists Chow, T.L. Cambridge, **2000**. (Chapter 4)

Key Questions

(1) What happens in the pulse-acquire experiment?

pw   d1   at   pw   d1   at   pw   at

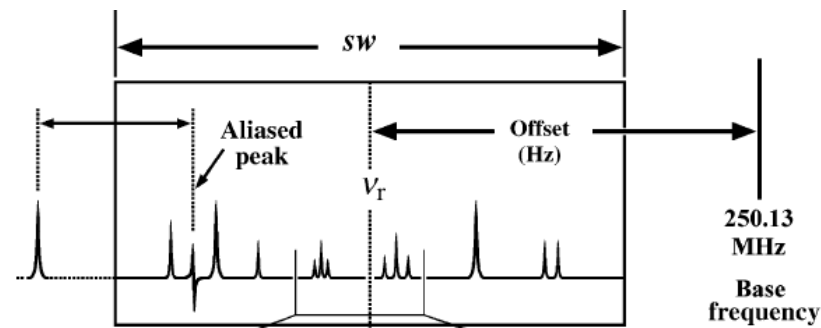steady state scans (ss)   scan #1   scan #2

(2) What's a Fourier transform?

$$g(\omega) = \frac{1}{\sqrt{2\pi}} \int_{-\infty}^{\infty} f(t) e^{-i\omega t} dt$$

(3) What is quadrature detection?

spectral window (sw)

$+\Omega$ Hz   0 $-\Omega$ Hz

reference frequency

(4) How often does signal need to be sampled?
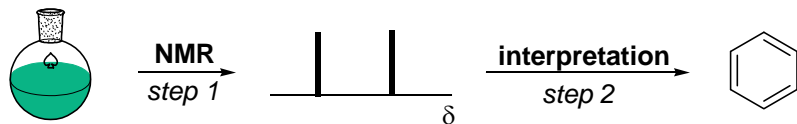
sw

Aliased peak

Offset (Hz)

v_r

250.13 MHz

Base frequency

I thank Dr. Neil Jacobsen (Arizona) and Dr. James Keeler (Cambridge) for providing many of the beautiful diagrams in this lecture.

Sample Preparation

In previous lectures, we analyzed step 2 of the following cartoon:

Today, we will figure out what happen in step 1. Most, but not all spectra are taken in deuterated solutions, and we'll assume this is the case for this lecture.

(1) **How much sample?** 1-5 mg is reasonable for a standard organic molecule with ^{1}H NMR. 5-50 mg is reasonable for ^{13}C NMR. Too little, and there won't be enough signal to give a spectrum. Too much, and the solution will be too viscous, and the peaks will be broad.

(2) **What kind of container?** Most NMR samples are held in cylindrical tubes 5 mm wide. Smaller diameters are available and more sensitive, but require special hardware. The glass itself is a especially uniform, which is why NMR tubes are expensive. Bad quality tubes give bad spectra.

(3) **How much solvent?** In a standard 5 mm tube, you should make up samples to a 5 cm height. NMR tubes are inserted into the magnet via a sample holder called a "spinner." This device gets spun around by air (if it gets dirty, the spinning will not work). You can get away with a smaller height by raising the tube up in the spinner, but the sample will be very difficult to shim. At 5 cm, the sample is effectively of infinite height, and you will not waste a lot of time shimming!

(4) **Can I get the sample back?** Unless you are forced to use a high-boiling solvent like DMSO or DMF, then you can simply rotovap off the solvent with no ill effects (other than those specific to the chemistry of your molecule). Note that $CDCl_3$ is slightly acidic, but CD_2Cl_2 is not. You can add molecular sieves to $CDCl_3$ to cut down on the acidity.

(5) **What safety considerations are there?** Invariably, the tops of NMR tubes break. When this happens, they get jagged edges, which can cut your fingers. You should throw out such tubes. Also, do not twist the tube too hard!

(6) **Should I add internal standard?** Many NMR solvents are available with a small amount of added tetramethylsilane (TMS), but you can reference the spectrum to residual solvent. Technically, this is only supposed to be done when the sample is very dilute. (If the sample is in a solvent mixture, TMS will be required.)

(7) **Oven drying?** No! This warps tubes, giving bad spectra.

The NMR tube is placed into the spinner, and the vertical position of the tube is set by the depth gauge. *If the sample ever breaks, stop what you're doing, put a sign up, and get help!* NMR hardware is breathtakingly expensive, and simply walking off and pretending nothing happened will cause the next user to cause further damage. It can also get you suspended from the NMR facility. This is bad news!

www.chemistry.nmsu.edu/Instrumentation/NMSU_NMR300_SampleDepth.html
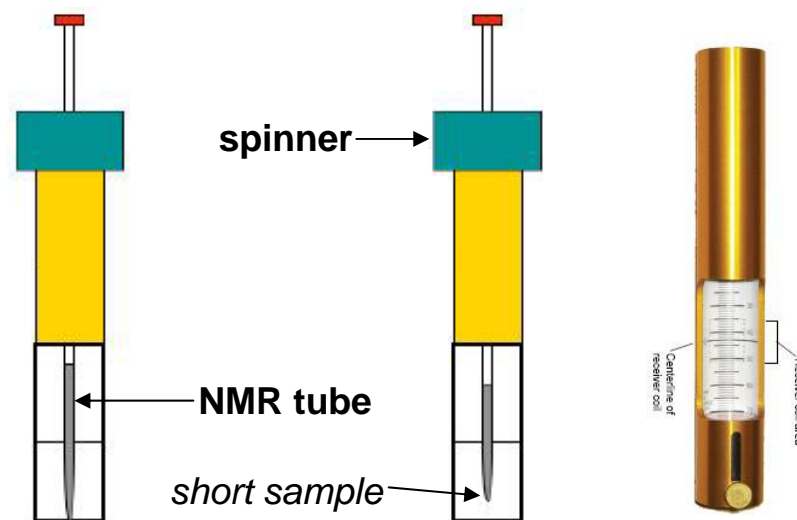
The Magnet

The strong magnetic fields in today's NMR spectrometers come from superconducting magnets which do not require a constant input of electricity to maintain their fields. They are made of miles(!) of $(NbTaTi)_3Sn$ embedded in copper wire. This material is only superconducing at low temperatures, so it is immersed in a double jacket of liquid helium and liquid nitrogen. Here is an *actual* cut-away view of the magnet:

http://www.cis.rit.edu/htbooks/nmr/chap-7/i2-3.htm
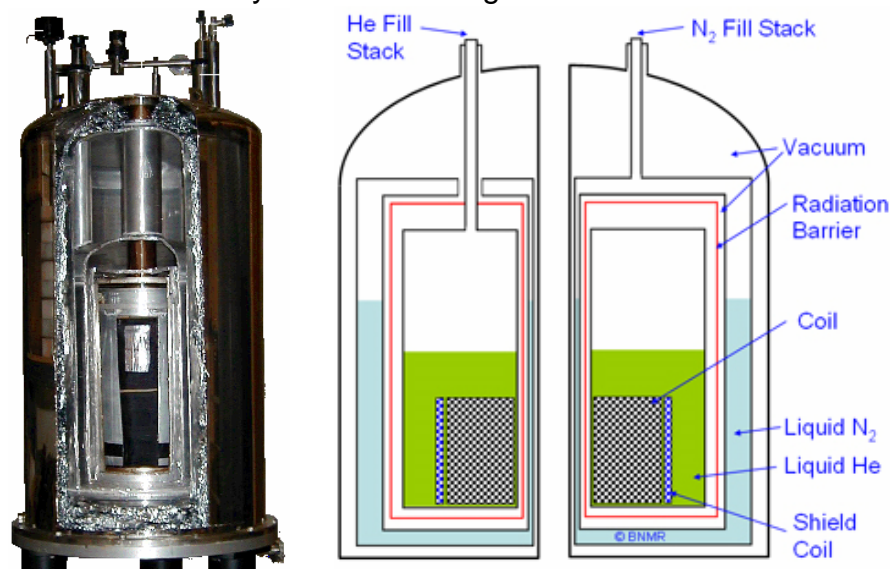
Many magnets are "shielded," which means that there is an additional superconducting coil outside of the main coil which cancels out much of the fringe field. Nonetheless, the NMR magnet has a considerable stray field, which has been known to affect medical implants, electronics, and iron-containing tools. Thus, keep your pacemaker, iPhone, and wrench away from the magnet! Portable dewars must be non-magnetic as well.

(If you prefer not to be murdered by the NMR facility manager, *never* put anything containing iron or other magnetic substances in the magnet like a stirbar. Solutions containing metals are OK, but might give bad spectra due to paramagnetic relaxation.)

The Lock

Many NMR experiments require more than one "scan," which means that data is collected from many identical experiments and added up. The idea is that the signal will add up, but the noise will cancel out. For timescales of a couple of days, **the signal to noise ratio (S/N) increases as the square root of the number of scans**.

For this to work, the magnetic field must be *extremely stable*; its strength must not "drift." Despite the advances in modern spectrometers, they can still drift by ppb over time, which is enough to mess up signal averaging. To reduce this drift, it is customary to "lock." The lock signal is usually the signal NMR resonance arising from the deuterium in the solvent (because D has a gyromagnetic ratio 1/7th that of H, there is separate hardware in the form of another "channel" for the lock). The sample is surrounded by a series of electric coils named "shims" which can be adjusted by a feedback loop to maintain the strength of the magnetic field at a constant value:
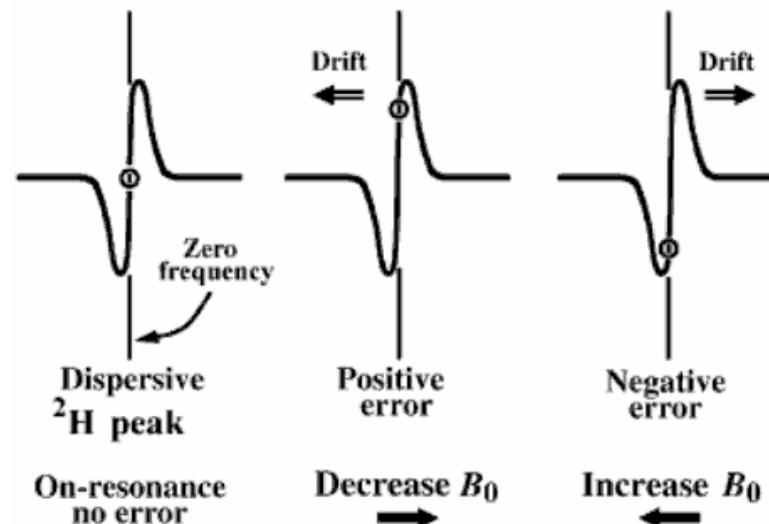
The strength of this shim correction to the overall field strength is called "Z0" on Varian spectrometers. The "lock phase" adjusts the phase of the peak so that it is dispersive as shown (the lock system will not work if the phase is wrong).

The Lock

On Varian spectrometers, locking is controlled from this window:

http://www2.chemistry.msu.edu/facilities/nmr/how%20do%20i.html#locking
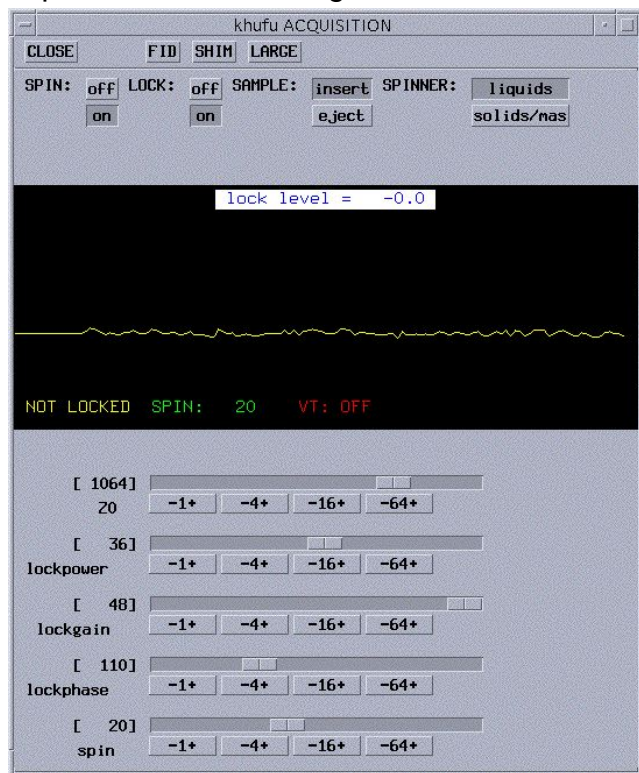
(1) Retrieve the default shims. Various values have been pre-determined for a standard sample for that particular spectrometer, and should be close to what you need for your sample.

(2) Make sure the lock is turned off. Currently, you see a flat, noisy line, which is a realtime display of the audio signal coming from the deuterium channel.

(3) Default values for lock power, lock gain, and Z0 are usually written down for the common NMR solvents. Lock phase usually does not require adjustment.

(4) Increase lock power and lock gain to about 80% of their maximum values.

(5) Start by using the slider to move Z0 into the neighborhood of the default value. This should give a stepped sinuosoidal wave like the one below. The more cycles you see, the farther away from resonance you are.
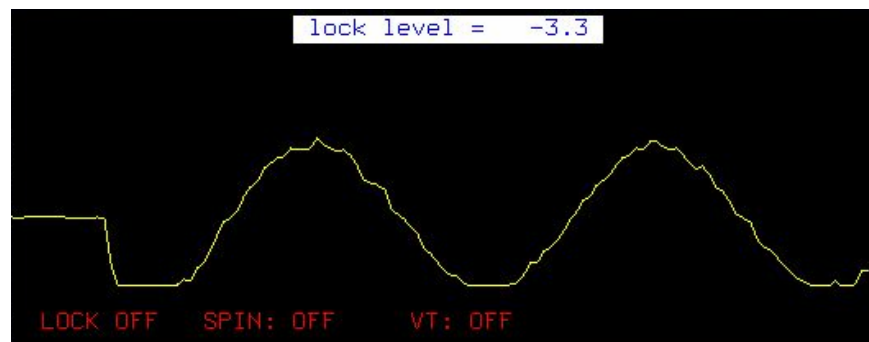
(6) If the signal increases and decreases in intensity rapidly, this is a symptom of *lock saturation*; decrease lock power and continue. As you approach resonance, make smaller changes. When you are at resonance, a step function appears:
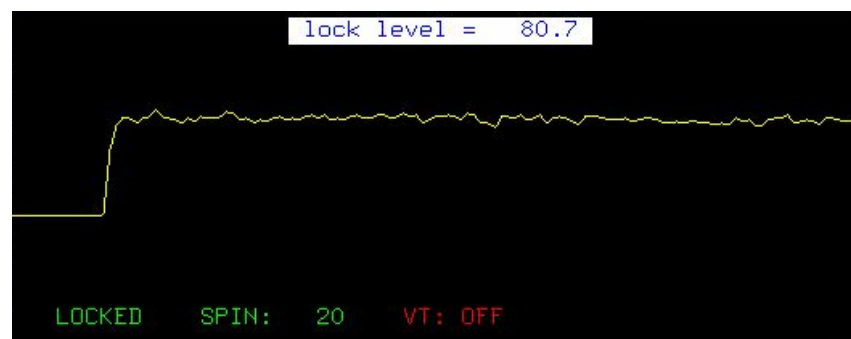

(7) Turn the lock on. This activates the feedback system. Reduce the lock power and lock gain to the typical values for that spectrometer. If the lock drops below 15-20, you will lose the lock.

Shimming

An excellent reference can be found at http://nmr.chem.uiowa.edu/manuals/Shimming-GAP-NMR-magnet.pdf (G. A. Pearson).

As mentioned earlier, the primary source of linewidth for typical organic samples is magnetic inhomogeneity. If the field strengths experienced by different parts of the sample are different, the nuclei will have different precession frequencies and dephase. To maximize the homogeneity, the magnet has dozens of electric coils inside the probe called "shims." On Varian spectrometers, shims are controlled from this panel:
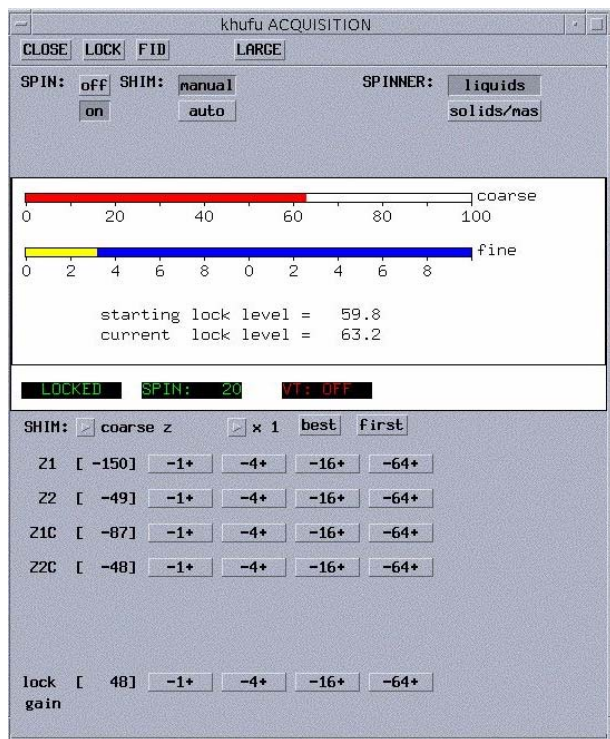
The variation of a magnetic field with position is called a **gradient**. The shim gradients must be adjusted to exactly cancel the gradient in the sample. Each shim gradient is labeled by a simple polynomial in Cartesian coordinates. The total correction is the sum of all these corrections.
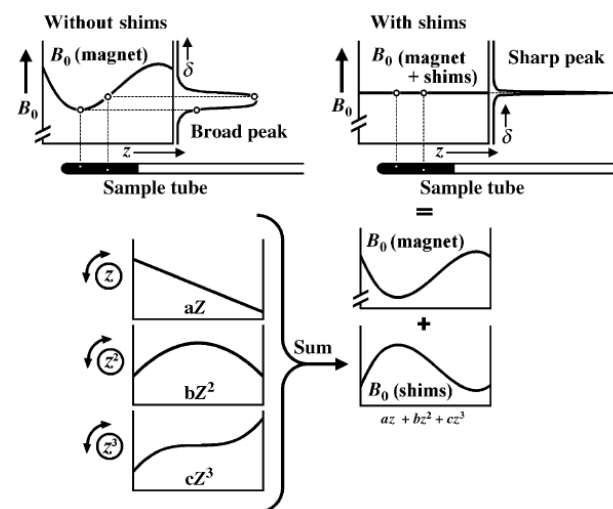
Most of the time, you will only need to adjust the first- and second-order shims (e.g., Z1 and Z2). With sample spinning, only the "spinning gradients" Z_n need to be adjusted. The effect of large errors in the shims is represented here:
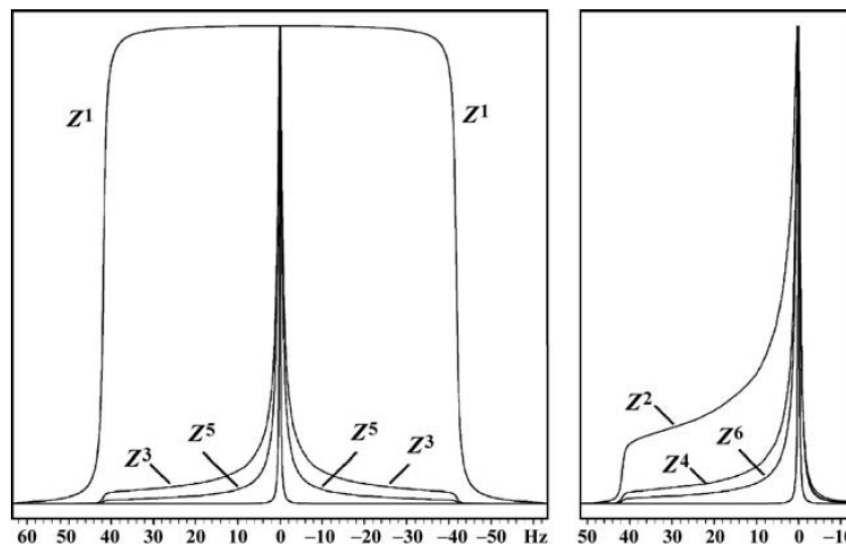
The lower the order of the shim, the bigger of an effect there is. If the error in the even-order shims changes sign, the error will move to the other side of the peak.

Shimming

The standard method is to **shim on the lock signal**. The more uniform the field, the stronger the lock signal will be. Thus, there is *no particular lock level that is optimal.* Any lock shimming is only complete when no further improvements can be made.

Shimming is challenging because different shims interact: the optimal lock level as a function of Z1 depends on the value of Z2.

Here is a zero-order procedure which assumes that the shims do not interact:

(0) Does the sample have any particulates? Is it cloudy? Is it too short (< 5 cm)? This will make shimming impossible.

(1) Turn the spinning on.

(2) Adjust Z1 to get a maximum reading. Start with large changes (64) and go to small changes (1).

(3) Repeat (2) for Z2.

Here is a better procedure:

(1) Adjust Z1 to a maximum.

(2) Move Z2 in the negative direction by say, 20 units.

(3) Adjust Z1 again. Is the maximum better? If so, return to step 2.

(4) Is the maximum worse? Move Z2 in a positive direction and adjust Z1 again.

(5) Repeat this until a maximum level is found.

If the lock level exceeds the maximum, reduce the lock power and continue the process.

Automated Shimming

Q: How good is "good"?

It depends on what you want. For a routine proton spectrum, peaks with a full width at half maximum (FWHM) of 1 Hz are acceptable. Dr. Pearson, evidently a master at shimming, says "Proton line widths in a standard sample should be less than 0.5 Hz with the spinner on. Line widths of 0.2 to 0.3 Hz are not difficult to achieve." However, I normally don't have the patience to adjust them to that level.

Spinning side bands can appear when the shims are not quite right. They go away when the spinning is turned off. On our fancy instruments, the shims are usually so good that it is not really necessary to spin at all. Spinning interferes with fancy NMR experiments, and is not used anyways.

If you're not satisfied, you can use a method called "simplex autoshimming." You can tell the instrument which shims to optimize and whether to make large (L), medium (M), or small (S) adjustments. This takes a couple of minutes to ten minutes.

For organic samples, you can use **deuterium gradient shimming** to adjust the Z-shims on any instrument with deuterium gradients (for example, our I600 does not have this). The idea is to apply a *deliberate* field gradient to a well shimmed reference sample. This creates a "shim map." Then, when any subsequent samples are perturbed in the same way in inhomogeneous fields, the difference between the known shim map response and the actual response can be calculated. The required corrections to the shims can then be made. **This works *really* well.**

A big advantage to deuterium gradient shimming is that it does not require the sample be locked first. Since shimming improves the lock level, but locking requires a good shim, one is normally faced with a "chicken and egg" problem. This problem is gone with gradient shimming, which I use for almost every sample.

Getting Faster Results

Q: I'm too impatient for all this. How do I get my spectrum right *now*? What if I want accurate integrals, too?

The primary cause of bad integrals is incomplete longitudinal (T_1) relaxation between scans. Here is what happens in a standard **pulse-acquire experiment:**
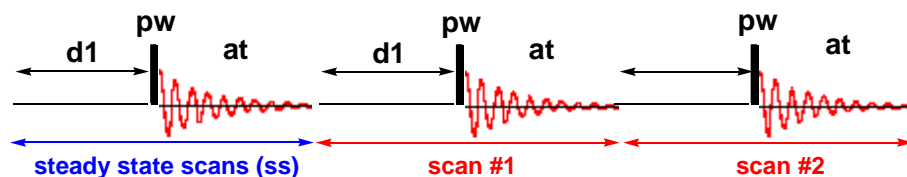
You can see all of this if you type "**dps**" at the VNMR console. "**dg**" will display all of these acquisition parameters. Here is what happens within each scan cycle:

(1) The spectrometer waits for a time **d1** seconds for the sample's spins to relax back to equilibrium.

(2) A transverse pulse of **pw** microseconds is applied. (More on how long to set this pulse in a moment.)

(3) Transverse magnetization/single quantum coherence is created, resulting in an oscillating signal: the free induction decay (FID). The signal is collected for **at** seconds (more on how that works in a moment, too).

(4) Every scan is called a "transient."

What happens in the whole experiment?

(1) By default (gain='n'), the first thing that happens is that the gain is automatically set. The gain is a setting that tells the hardware how much to pre-amplify the signal before detection. Too much signal, and the receiver gets overwhelmed ("ADC overflow"); too little, and the signal is too weak.

(2) The **autogain procedure** is run with the settings you have chosen: the particular combination of **d1**, **pw**, **at**. The status window will read "autogain" while this happens.

(3) Because relaxation is not necessarily complete between scans, a number of "steady state" scans are run. Data is *not* stored from these runs.

(4) A "sum to memory" of all of the scans is made for the final stored FID on the computer. Remember, S/N ~ Sqrt[**nt**]. A total of **nt** transients are run during the experiment.

(5) The CYCLOPS phase-cycling procedure is normally used to remove artifacts. To make it work properly, make sure **nt** is set to a multiple of 4.

Fast *and* Accurate?

There are two key points to realize here:

(1) At typical concentrations of several mg per tube, perfectly good S/N is realized *within one scan*. Thus, **signal averaging is unnecessary for proton spectra.**

(2) Taking more than one scan results in incomplete relaxation, which results in unequal integrals (since different protons have different T_1 times).

Thus, here is a fast procedure (I once managed to take seven good NMR spectra in one ten minute slot using this):

(1) Insert sample.

(2) Select defaults for the nucleus solvent (e.g., ^{1}H, $CDCl_3$). Recall the default shims.

(3) Main Menu...Shim...Gradient Autoshim on Z. The shims will be adjusted iteratively. If this takes more than three cycles, it probably won't work. Abort and use the standard lock, then lock shim procedure.

Getting Faster Results

(4) If you feel very impatient, you can change directories while the gradient shimming is in progress and type in the next set of commands (without pressing enter).

(5) Click on "acqi." In most cases, the instrument will already be locked. If not, you will need to lock it manually. A lock level of 30-80 or so is fine. A super nice lock is unnecessary.

side note: For a short experiment, there is no real need to lock the instrument at all if you have the gradient autoshim available. However, if the referencing is too far off, this can mess up the spectral window, and peaks may not appear in the spectrum properly. (See *aliasing* later in this lecture.)

(6) nt=1 bs=1 d1=0 ss=0 ga (ga starts acquisition.)

(7) When acquisition completes, type "wft aph f cz vsadj." wft requests a weighted Fourier transform. aph performs automatic phase correction. f displays the whole spectral window. cz clears the integral resets. vsadj automatically adjusts the vertical scale.

(8) Reference the spectrum if needed. While you do all this, the sample will largely relax to equilibrium.

(9) Type "gain?". The response will be "Not used (xx)." Now, type "gain=xx," where xx is the number it gave you.

(10) Reacquire the spectrum using "ga." By typing gain=xx, you are telling the spectromter to skip the autogain process and use the value you tell it. Make sure you've waited at least one minute since the end of the previous acquisition.

side note: You should save every NMR spectrum you take using a detailed filing system of your choice. Storage space is essentially free these days, and you may regret it if you one day want to see your spectrum again and have lost your printout.

Baseline correction is requested by commands like **dc** (drift correction) or **bc** (baseline correction). These subtract linear-(dc) or polynomial-type (bc) corrections from the spectrum. Unless you know what you're doing, don't bother with this. Avoid the **hammer** macro, which is equivalent to **dc cz region bc** (automatic integration and baseline correction), as it is unlikely to find the baseline properly in a complicated spectrum.
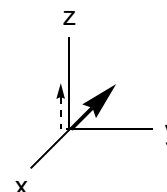
The Ernst Angle

Q: What pw should I use?

There are times when multiple scans are **required**, like 1D ^{13}C spectra. Should you wait longer between scans and use a full 90° pulse, or wait less time between scans but use a smaller "tip angle"?

A 90° pulse gives maximum intensity, but you have to wait longer between scans for M_z to decay:
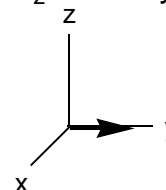
A 45° pulse, for example, gives less intensity, but there is less M_z to decay:


Richard Ernst (Nobel Prize, 1991) has conclusively demonstrated that **smaller tip angles and no delay between scans is optimal**. What is the best tip angle (also called the "nutation angle")? This is given by the "Ernst angle":

$$\cos(\theta) = e^{-(d1+at)/T_1}$$

T_1 can be measured by inversion recovery. The command **ernst(T_1)** will set the Ernst tip angle for a particular value of T_1.

Pulse Width Calibration

Once you've chosen a tip angle, how do you make sure that the pulse width delivers the desired tip angle? The simplest method is to deliver what you *think* is a 360° pulse, and then see if you observe any signal:
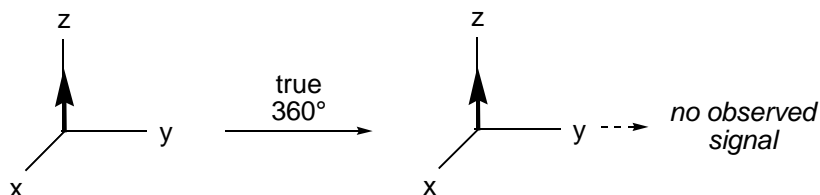

If the pulse is too short, the signal will appear to be negative:
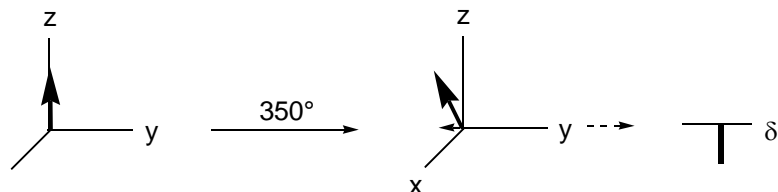
Conversely, if the pulse is too long, the signals will be positive. To find the true 90° pulse angle, the calibrated 360° pulse is just divided by four.

Tuning and Matching the Probe

We would like to transfer the maximum amount of energy (generate the most current) for a given power level. The more efficient the transfer, the shorter the pulse has to be, and the more sensitivity there is in the experiment. This is done using a resonant RC circuit:
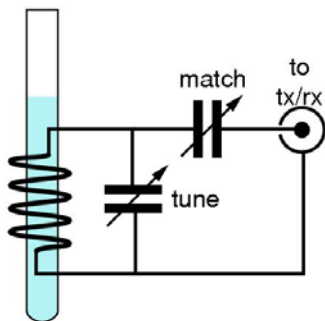
Practically, this means twisting knobs which control variable capacitors ("match" and "tune") so that the reflected power from the probe is minimized. These capacitors are very delicate, expensive, and are not to be treated lightly!
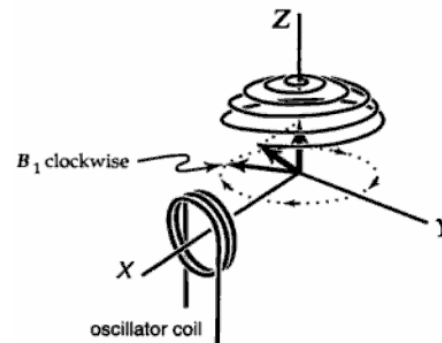
The Receiver

Q: How does the signal get from the spectrometer to the computer?
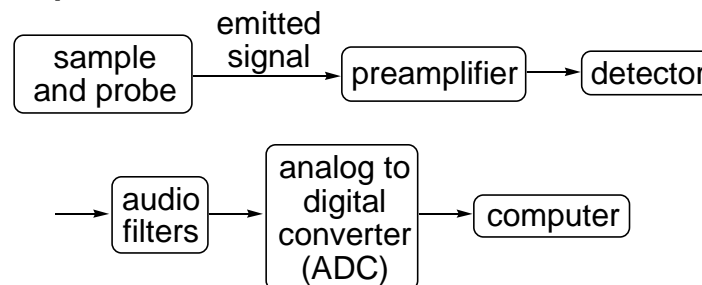
Preamplification. The signal is amplified immediately. The longer the signal goes through copper wires, the more thermal noise it picks up. In **cryoprobes**, the wires in the probe are cooled to further reduce thermal noise, giving an increase in S/N of 3-4x (but they're expensive).

Detector. Remember, this is what the vector model looks like after a 90° pulse in the laboratory frame (Roberts, page 21):


Now, the precession frequency of a proton is really, really fast: say, 500 MHz. This is much too fast to be detected, even by top-of-the-line electronics. How is the signal detected?

The Receiver

Mixing. The sample is mixed with a reference RF signal. The beat frequency of the resulting mixture is in the audio range and is suitable for the electronics (Wikipedia):

$$\sin(\omega_1 t)\cos(\omega_2 t) = \frac{1}{2}\left[\sin((\omega_1 + \omega_2)t) + \sin((\omega_1 - \omega_2)t)\right]$$

One of the beat frequencies is much faster (the oscillations inside the envelope) and one of them is much slower (the envelope):
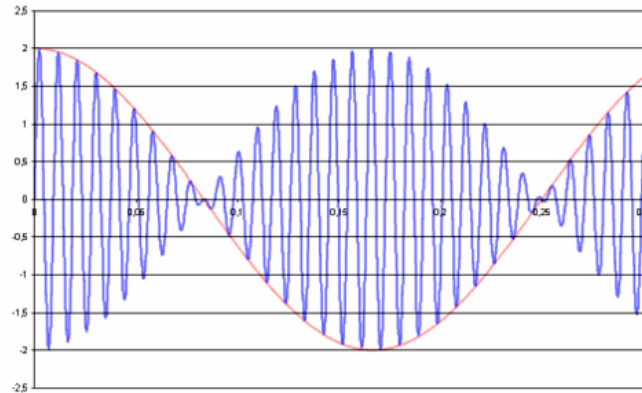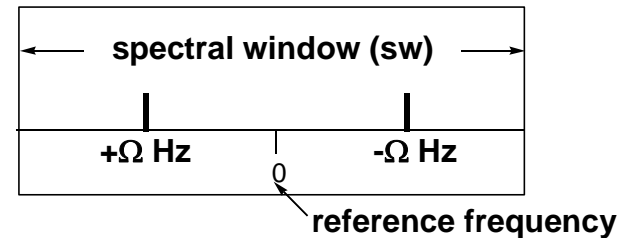

This is the physical analog of moving from the laboratory frame to the rotating frame! The vector picture is (Roberts, page 23):
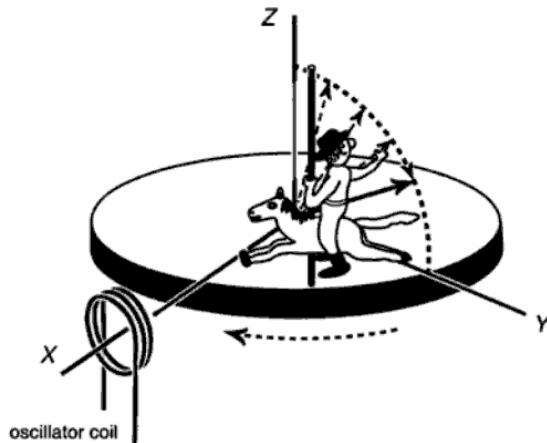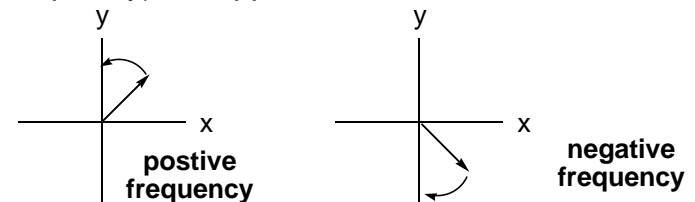

Quadrature Detection

Usually, the reference frequency (the zero of the new rotating frame frequency scale) is placed in the middle of the spectrum.

(If the reference frequency were to be placed on one side of the spectrum, then half of the spectral window would be wasted space.) Now, we have a problem:

Q: How can positive and negative frequencies be distinguished?

To see why this is a problem, imagine the detector is positioned at +x. Imagine we're in the rotating frame, and the vector has a *higher* frequency than the reference (a positive frequency). It then appears to rotate counterclockwise. Conversely, a vector that is *lower* than the reference frequency by the same amount (a negative frequency) will appear to rotate clockwise:

From the perspective of the detector at +y, these vectors are **indistinguishable!** Obviously, this is a problem.

The solution is to have detectors at *both* +x and +y, so that we can tell if the vector passes +y or -y first. This is called **quadrature detection**.

Quadrature Detection
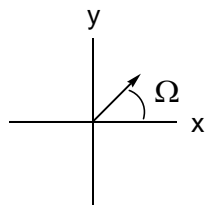
In our simple model, we have *two* input signals:

$M_x = M_0 \cos(\Omega t)$ and $M_y = M_0 \sin(\Omega t)$

where M_0 is the equilibrium magnetization (this is right after the 90° pulse).

By convention, one forms the signal S(t) as:

$S(t) = M_x + i\, M_y = M_0 \exp(i\, \Omega t)$.

With relaxation, this must be amended to:

$S(t) = M_0 \exp(i\, \Omega t) \exp(-Rt)$

where R is a time constant for relaxation ($1/T_2{}^*$). If there is more than one signal present, one has a linear combination of signals. This gives rise to a very complicated FID. Our task is to deconvolute such an FID into its component frequencies.

The Fourier Transform

I won't go into the mathematical details of the Fourier transform here. For detailed discussions, please see:

(1) Weber and Arfken <u>Essential Mathematical Methods for Physicists</u>

(2) Chow <u>Mathematical Methods for Physicsts</u>

On the next problem set, you will investigate the behavior of the Fourier transform for yourself. Here, we follow the treatment in Chow.

Consider a periodic time-domain function f(t):

$$f(t+T) = f(t)$$

T is the period of the function.

The **Fourier series representation** of a function f(t) defined on an interval $[-\pi, \pi]$ is defined as:

$$f(t) = \frac{1}{2} a_0 + \sum_{n=1}^{\infty} a_n \cos(nt) + \sum_{n=1}^{\infty} b_n \sin(nt)$$

where the a_n and b_n are coefficients. In English, this means we write the function as a sum of cosine and sine functions. In linear algebra language, this is a projection of f(t) onto an orthogonal basis of oscillations.

We're not mathematicians, so we will assume that f(t) and its Fourier series are well behaved. Specifically, we will assume that the series exists, converges, and can be integrated by term.

Multiply both sides by $\cos(mt)$ and integrate over $[-\pi, \pi]$:

$$\int_{-\pi}^{\pi} f(t) \cos(mt)\, dt = \frac{a_0}{2} \int_{-\pi}^{\pi} \cos(mt)\, dt +$$

$$\sum_{n=1}^{\infty} a_n \int_{-\pi}^{\pi} \cos(nt) \cos(mt)\, dt +$$

$$\sum_{n=1}^{\infty} b_n \int_{-\pi}^{\pi} \sin(nt) \cos(mt)\, dt$$

(1) The first integral is zero, since the integral of cosine over a full cycle is zero.

(2) Cos(nt) and cos(mt) are orthogonal:

$$\int_{-\pi}^{\pi} \cos(nt) \cos(mt)\, dt = \pi \delta_{nm}$$

(δ_{nm} is the Kronecker delta, which is 1 if n=m and 0 otherwise.)

(3) The third integral is zero since sine and cosine are orthogonal.

The Fourier Transform

Thus, we find that:

$$\int_{-\pi}^{\pi} f(t)\cos(mt)\,dt = \sum_{n=1}^{\infty} a_n \pi \delta_{nm} = a_m$$

$$a_m = \frac{1}{\pi}\int_{-\pi}^{\pi} f(t)\cos(mt)\,dt$$

This gives us a way to figure out what the coefficients a_n are. For b_n, there is a similar expression:

$$b_m = \frac{1}{\pi}\int_{-\pi}^{\pi} f(t)\sin(mt)\,dt$$

These are the **Euler-Fourier formulas**.

Q: Now, what is the meaning of these coefficients a_n and b_n? Are they just numbers in a series?

A: Yes, but they're more than that: they represent how much of each frequency is present in the function.

Now clearly, not all functions are periodic. We can view the **Fourier transform** as the limit of Fourier series as the period T approaches infinity:

$$f(t) = \int_{-\pi}^{\pi} c(\omega)e^{i\omega t}\,d\omega$$

Defining the Fourier transform $g(\omega)$ as:

$$g(\omega) = \sqrt{2\pi}\,c(-\omega)$$

Explicitly:

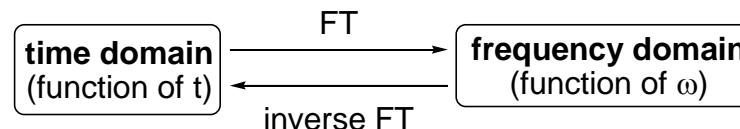$$g(\omega) = \frac{1}{\sqrt{2\pi}}\int_{-\infty}^{\infty} f(t)e^{-i\omega t}\,dt$$

Note that $g(\omega)$ is a function of frequency. The integral is over t so the function that comes out is a function of ω.

We can also take the inverse Fourier transform of $g(\omega)$ to get $f(t)$ back (notice the change in sign in the exponential):

$$f(t) = \frac{1}{\sqrt{2\pi}}\int_{-\infty}^{\infty} g(\omega)e^{i\omega t}\,d\omega$$

Other forms which differ in the pre-factor are possible. The factor of 2π comes from a change of variable: sine and cosine are periodic over $[-\pi,\pi]$, whereas the aperiodic function is periodic over T (as it approaches infinity). For a clearer explanation, see pp 165-166 in Chow.

If the this is all gibberish, here are the essentials:

$$\text{time domain (function of t)} \xrightarrow{\;\;FT\;\;} \text{frequency domain (function of }\omega)$$
inverse FT

(1) The Fourier transform takes a function of time and expresses the same information in terms of frequency.

(2) We want to know about frequencies because they are easier for humans to interpret. Frequency is a natural space to work in, since it has a direct correspondance with energy.

(3) In general, $g(\omega)$ is a complex-valued function. It is customary to display only the real part.

(4) In real life, we work with discrete time series, rather than continuous functions of time. Similar definitions apply.

(5) The **fast Fourier transform (FFT)** is an efficient algorithm for computers. It requires that the number of points n be a power of 2. Computations by the definition take $O(n^2)$ while the FFT takes only $O(N \log N)$. (Actually, there are FFT algorithms for arbitrary n, but traditionally, these are not the ones being used in NMR programs.)

Now, let's take a closer look at the FT of a single oscillation.
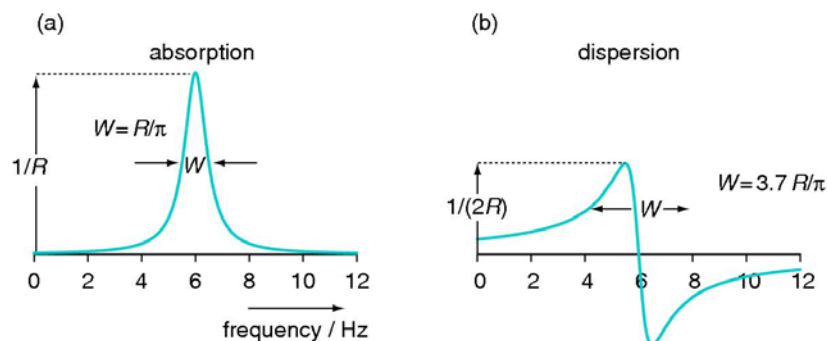
The Fourier Transform

As I mentioned, the time-domain signal arising from a single rotating vector in the rotating frame is:

$$S(t) = M_0 e^{i\Omega t} e^{-Rt}$$

The difference between the frequency of the vector and the reference frequency is called the offset Ω. Note that Ω is a *parameter*, while the frequency ω is a *variable*. The Fourier transform is:

$$\underbrace{\frac{M_0 R}{R^2 + (\omega - \Omega)^2}}_{\textbf{real part}} + i \underbrace{\frac{-M_0 (\omega - \Omega)}{R^2 + (\omega - \Omega)^2}}_{\textbf{imaginary part}}$$

The real part is an **absorption mode Lorentzian** while the imaginary part is a **dispersion mode Lorentzian**. The important message is the absorption lineshape is narrow, while the dispersion lineshape is wide:

In this course, we'll follow the convention in the Keeler book:

$$A(\omega) = \frac{R}{R^2 + (\omega - \Omega)^2} \qquad D(\omega) = \frac{-(\omega - \Omega)}{R^2 + (\omega - \Omega)^2}$$

Phase Correction

The Fourier transform is *linear*, which means multiplication in the time domain by a constant results in multiplication in the frequency domain by the same constant. There is no guarantee that the detector phase will line up with the actual phase of the signal.

Additionally, translation in the time domain (i.e., a delay) corresponds to a phase shift (i.e., a rotation of the real and imaginary parts) in the frequency domain. There is always some dead time after the pulse during which no signal can be collected (due to the disruption of the pulse). All this means that the actal signal collected always has some fixed zeroth-order phase shift ϕ:
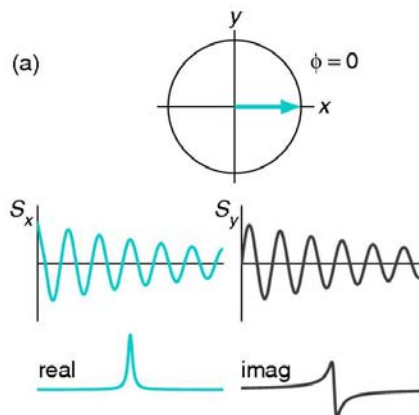
$$S(t) = M_0 e^{i\Omega t} e^{-Rt} e^{i\phi}$$
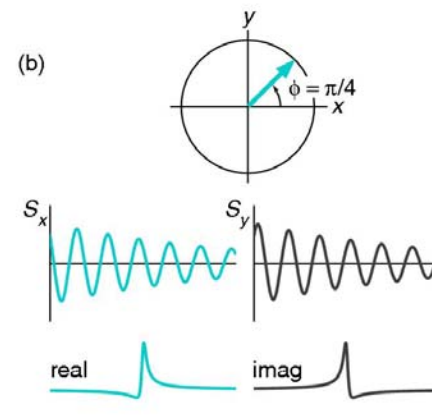
Thus the Fourier transform is:

$$M_0 \left(A(\omega) + iD(\omega) \right) \left(\cos\phi + i\sin\phi \right)$$

This means that the real and imaginary parts will no longer contain purely absorption or dispersion components:

Phase Correction

The solution is simple: the signal is just multiplied by a **phase correction** before Fourier transformation. This can be done manually or automatically. In VNMR, this is what the **aph** command does. Automatic phasing has trouble with spectra that have low S/N. Manual phasing is advised in such cases. Poor phasing will cause significant integration errors.

So far, these are **zero-order phase errors**: every peak has the same phase error regardless of its frequency. An additional complication is **first-order phase errors**: the phase error in each peak depends on its offset. These arise because pulses do not tip every spin the exact same amount. In fact, they get tipped by an amount that depends on their offset. I'll discuss these **off-resonance effects** later in the course. Thus, we must apply an overall linear correction:
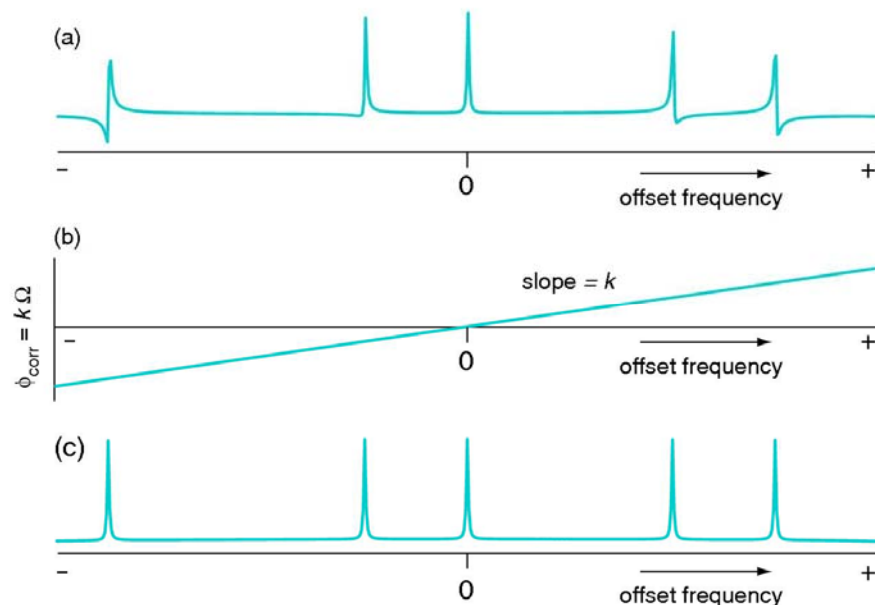
Artifacts like aliased/folded peaks, quad images, etc. will be impossible to phase. Spectra that are hard to phase can benefit from a small amount of backwards linear prediction.

Data Acquisition

Let's take a step back and revisit the instrument diagram:
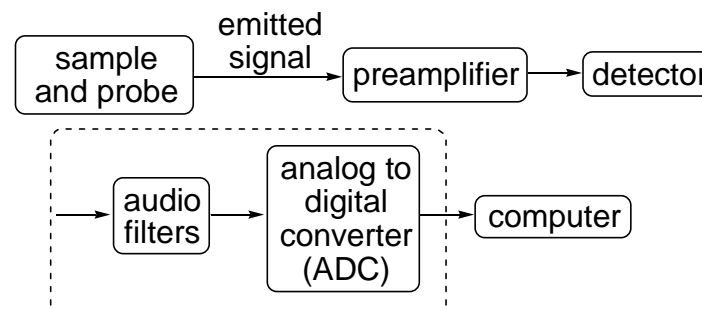
In this section, we'll look at how the signal gets from the detector to the computer. At the moment, we have an analog signal. As such, it has an infinite frequency range and cannot be understood by the computer, which only accepts zeroes and ones.

The **audio filters** are electronic widgets that only let in signals that outside of the spectral window: the range of frequencies we want to know about. There shouldn't be any signals outside of the spectral window, but there's always noise. By filtering these signals out, the S/N of the spectrum is dramatically improved.

The **analog to digital converter** turns the signals into zeroes and ones. The ADC samples the signal at regular intervals and assigns an intensity value to it. The values can't just be anything. A four bit ADC would be able to give 2^4 intensity values: 0000,0001,0010, etc. Typical ADCs have 16 to 32 bits.

This gives the concept of **dynamic range**. Suppose a spectrum has a very weak signal and a very intense signal. How different can these intensities be before the ADC can't tell the difference?

Q: What is the dynamic range of a four bit ADC?

Dynamic Range

A: 16 to 1. The smallest signal can have an intensity of $2^0=1$ while the largest signal can have an intensity of $2^4=16$.

Why does this matter? The autogain procedure increases the receiver gain as much as can be done without exceeding the dynamic range of the ADC. (If you exceed this, you get an "ADC overflow" error.) This is why you spectra are typically acquired in deuterated solvents. If they weren't, the analyte signal would be overshadowed by an enormous solvent signal. (Actually, these days, dynamic range is so good that this is not strictly necessary anymore. See the no-D NMR section of Lecture 6.)

Jacobsen puts this in perspective nicely:

S/N is an absolute limitation: to be seen, a signal must rise above the level of the noise, no matter how strong any other signals are.

Dynamic range is a relative limitation: it determines how much variation in signal intensity there can be in a spectrum.

Q: Where does noise come from?

Analog thermal noise comes from all the electronics and wires. Most of this comes from before the pre-amplifier. If noise gets added after the pre-amp, it doesn't matter as much since the signal already got amplified by a lot.

Digitizer noise comes from the fact that the continuous analog signal is being "binned" into discrete increments. In a spectrum with a large dynamic range, the small signals will be poorly digitized. The peaks will look blocky because there are only a few bits to describe changes in intensity.

Q: How often does the ADC need to sample?

The Nyquist Theorem

A: At *twice* the frequency of the highest frequency component.

To see why this is true, imagine you are watching Fido run around a track:
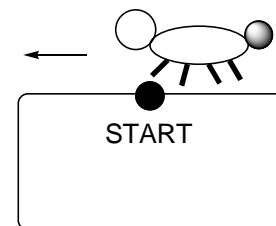
(*You* try drawing a dog in ChemDraw. I'm a chemist, not a biologist.) At first, let's say you're very interested, and you note the position of Fido every second. You determine that he crosses the finish line every two minutes.

After a while, you get bored. You only look up every minute to note his position. Assuming Fido is *actually* going at a regular pace of a lap every two minutes. Below, the dark circle represents Fido's position and the hash represents the finish line:

| t=0 | t=1 | t=2 |
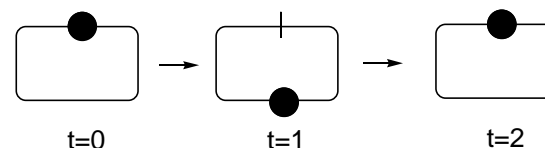
Despite the vastly reduced sampling interval, you still get the right answer: one lap every two minutes. But what if you only look up every *three* minutes?


| t=0 | t=3 | t=6 |
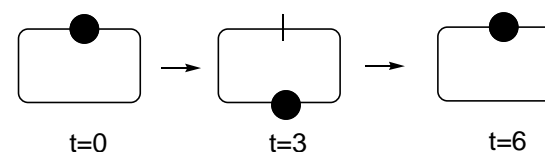
Uh oh. Now it looks like Fido is only going around once every six minutes, not two. We got the wrong answer, because in between the samples, he ran all the way around the track!

The Nyquist Theorem

Here are some illustrations taken from redwood.berkeley.edu/bruno/npb261/aliasing.pdf. (The appearance of a high-frequency signal as a low-frequency signal due to inadequate sampling is called **aliasing**.)

Another way to put the Nyquist theorem: we have to sample a wave once per crest and once per trough to have any idea what its frequency is.
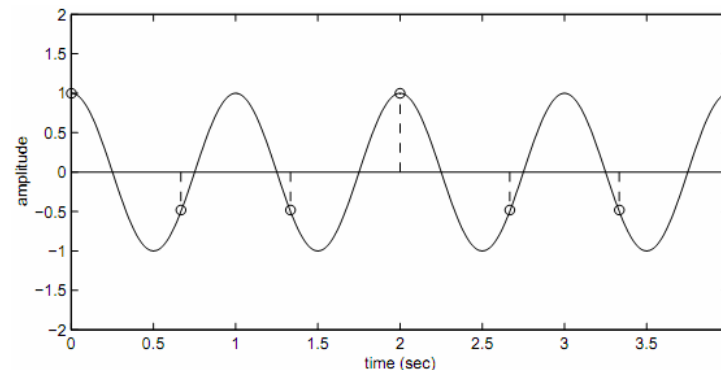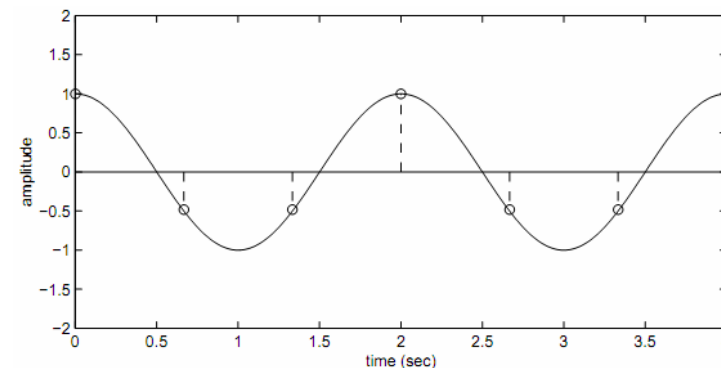
Here is a 1 Hz signal being sampled at 2 Hz:
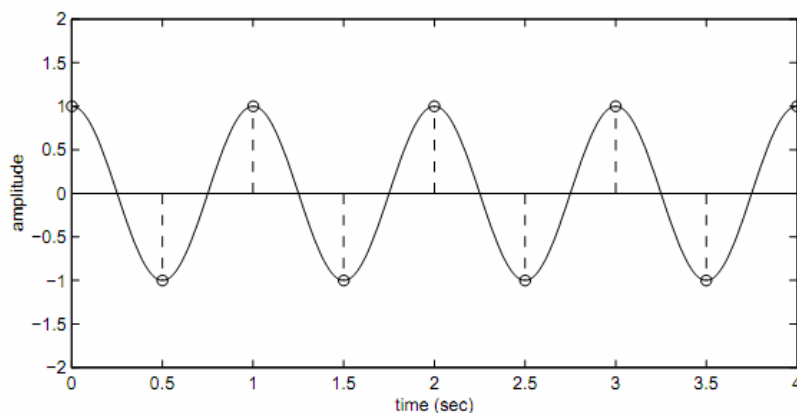
Sampling at 3 Hz is *more* than twice 1 Hz, so we get the frequency quite accurately:
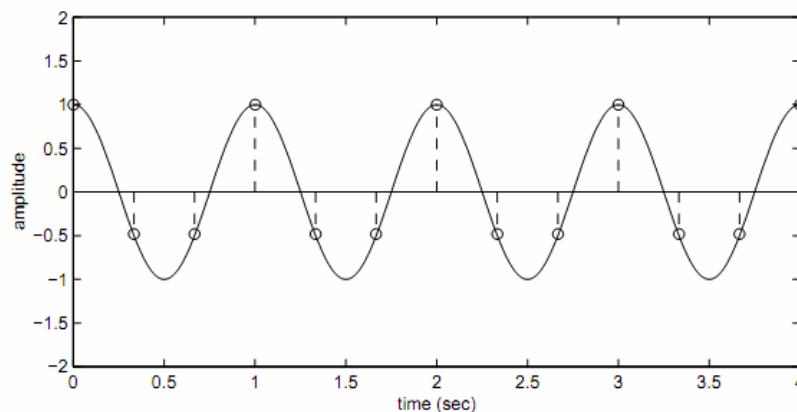
Sampling at 1.5 Hz is *less* than the Nyquist frequency of 2 Hz, so we now get an inaccurate picture of what the wave looks like:

In fact, the wave *seems* to look like this:

You can see this effect in movies where it looks like the wheels of cars seem to go backwards. The wheels are turning too fast for the sampling interval of film: 24 fps. This also occurs in 2D graphics:

The effect is reduced if the text is blurred first. This is what anti-aliasing is:

Acquisition Parameters

Q: What do all the acquisition parameters mean?

Specifically, I want to look at what **at**, **np**, and **sw** are:

at is the total acquisition time in seconds -- how long the FID is being recorded for.

np is the total number of points -- how many points the FID is made of.

Recall that the receiver has (effectively) two channels which record the x and y components of magnetization. The overall signal is formed as $S(t) = M_x + i M_y$.
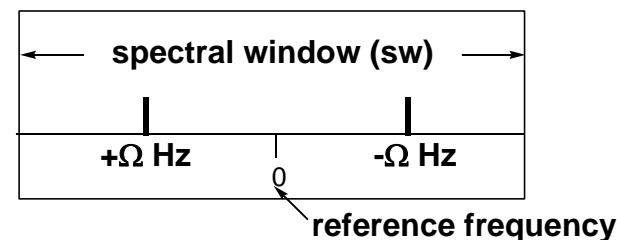
Data points are collected at regular intervals. The time between data points is called the dwell time (**dw**):
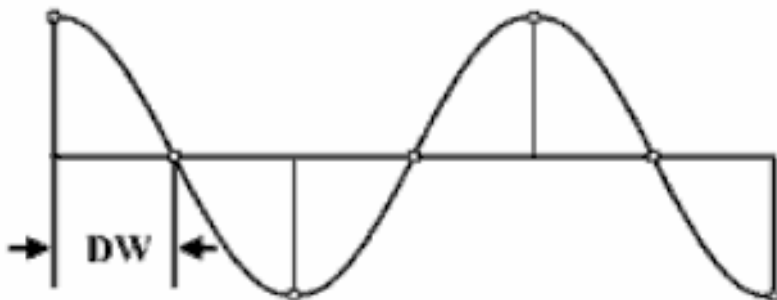

This means we are sampling at a frequency of 1/**dw**. From the Nyquist theorem, the range of frequencies we are gathering accurately is the spectral window **sw**:

$$sw = 1 / 2 \times dw$$

This gives the spectrum:

Where do aliased peaks appear? If the real and imaginary points are sampled simultaneously (Varian), folded peaks appear at an equal distance inside the opposite edge of the spectrum. With alternating acquisition (Bruker), aliases appear reflected from the nearest edge:
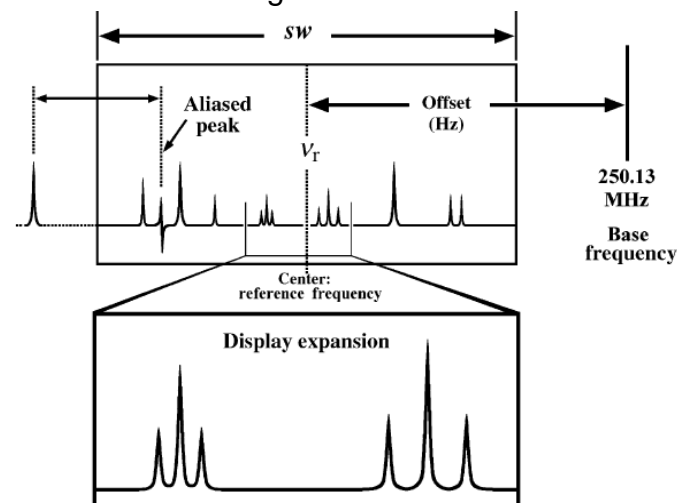
Because of the first-order dependence of phase errors, a linear phase correction will be unable to correct for the phase error in an aliased peak.

Q: What is **np** in terms of **sw** and **at**?

Acquisition Parameters

This is a simple units problem:

at (time) = **np** (points) x **dw** (time per point)

Dwell time is related to the spectral window (or "sweep width"):

at = **np** x (1 / 2 x **sw**)

Rearranging, we have **np** = 2 x **sw** x **at**.

Q: What happens if we double the number of points?

A: We either have to double the acquisition time or double the spectral window (or some mixture of the two). In VNMR, changing **sw** gives a new number of points with the same **at**.

We're always interested in maximizing S/N. One way is to take more scans:
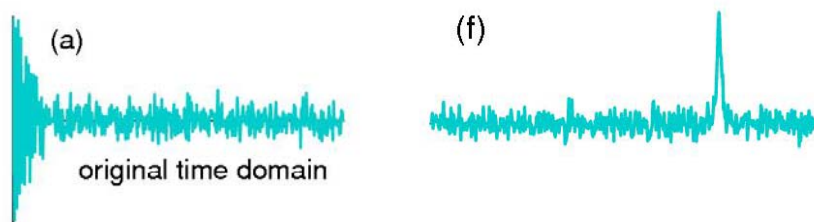
$$S/N \sim Sqrt[nt]$$

When you type the command **time**, VNMR will tell you how long the experiment will take. For the pulse-acquire experiment:

$$time = nt \times (d1 + pw + at)$$

pw, for all intents and purposes, is instantaneous.

Q: But what can we do to maximize S/N for data that's already been collected?

(a)

(f)

original time domain

Digital Resolution

The above is an FID with its spectrum, which we've decided is unsatisfactory. Now, fill in the blanks:

**Most of the signal is in the (early, late) part of the FID.**

Most of the fine details are in the (early, late) part of the FID.

The first one is easier: the signal is exponentially decaying. Because of the inhomogeneity of T_2^*, most of the signal is gone in about half a second. If I collect the FID for ten seconds, the last eight seconds will contain almost all noise.

The second one is harder. We need the concept of **digital resolution**, which tells us how many frequency units are described by a single spectrum point:

$$digital\ resolution = sw / np$$

From the above, we can see that the resolution is 1 / 2 x **at**. This means that at small acquisition times **at**, digital resolution is low. (It's a bit confusing, since a large value of 1/2 x **at** actually means that there is poor resolution.) Here is a picture:

Low digital resolution
3.0 Hz/pt.

**High digital resolution**
0.3 Hz/pt.
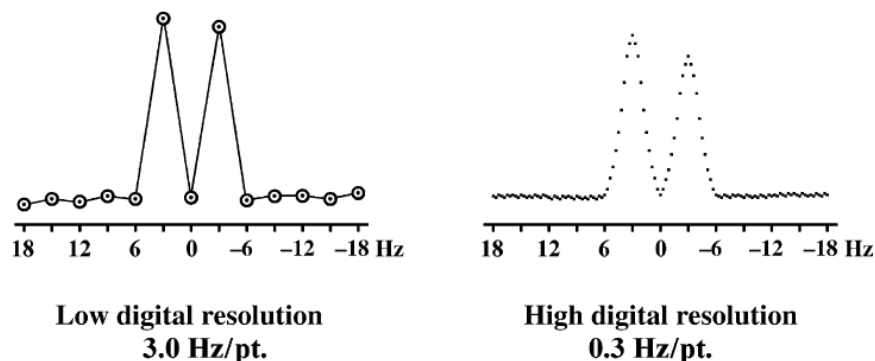
Another way to think about this is that fine details need low frequency components to describe, which require a long **at**.
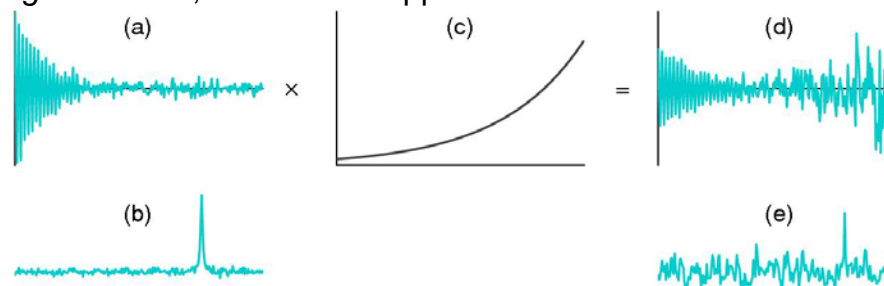
Apodization

One answer is to multiply the FID by a time-domain function a function that emphasizes the early part of the FID (enhances S/N) or the later part (enhances resolution). In time series analysis, this is called *windowing* or **apodization**.

The most common kind of apodization is **line broadening** which multiplies the FID by an exponential function:
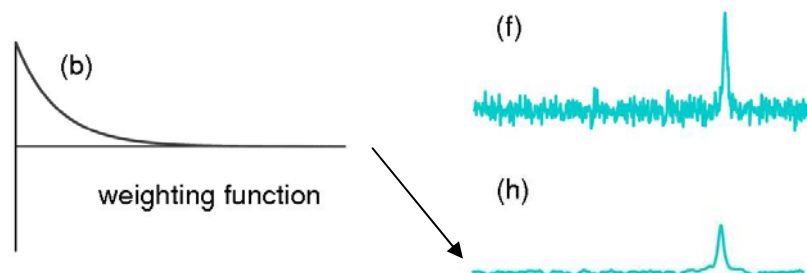

Notice the S/N has increased, but the peak is wider. There's no free lunch. However, some lunches are better than others: it turns out the best improvement in S/N comes from a filter whose time constant is *matched* to the decay envelope of the FID. The degree of line broadening is specified by the amount the width of the peaks will increase. In VNMR, this is the parameter **lb**. For proton spectra, **lb** = 0.1 (Hz) is reasonable. For carbon spectra **lb** = 1 is reasonable. Attempting to use too much line broadening will not work--it's like zooming in on a digital picture.

The exponential function starts out big and gets small, and therefore enhances S/N. If we give the exponential function a *negative* value, it will do the opposite:

Unfortunately, the last part of the FID, which contains a lot of noise, gets emphasized too much, giving a noisy spectrum. A better choice is to use an apodization function is part negative exponential and part Gaussian:
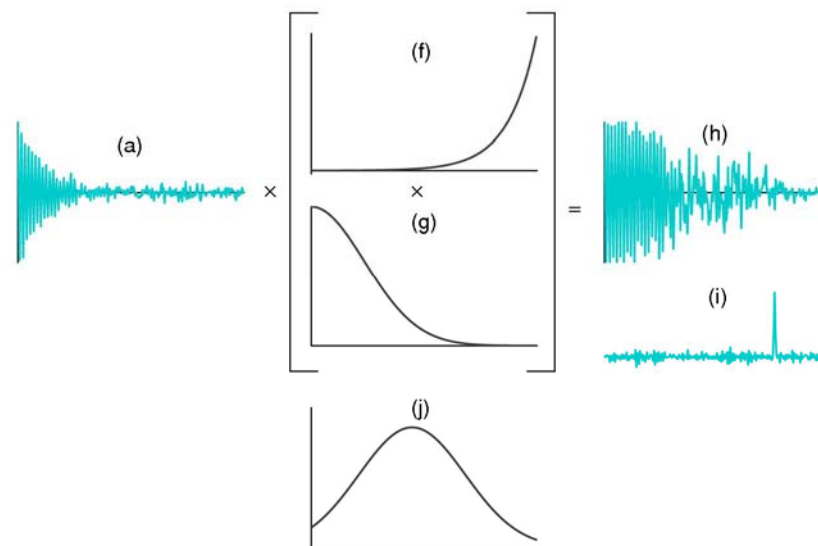
The sine bell and sine bell squared take half of a sine wave. There are some details as to which one is best for which experiment, but I won't go into it here:
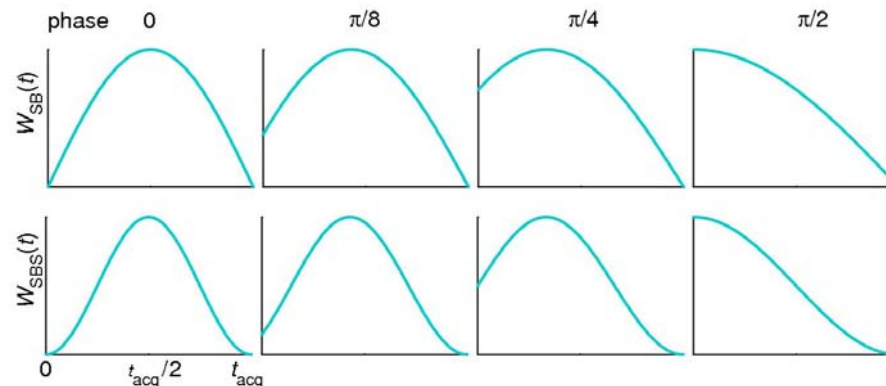

For 1D experiments, line broadening and Gaussian-type functions will be entirely adequate.

Zero-Filling

Acquiring data for more time (longer **at**) will give better digital resolution simply because **np** is increased. But it will give worse S/N because the amount of noise in the later part of the FID is higher.

If we think about the line broadening function as emphasizing the beginning and de-emphasizing the end, we can take this to a further limit by adding zeroes to the end of the FID, which, of course, don't have any noise, but still increase the effective **np**:

(a)   (b)   (c)
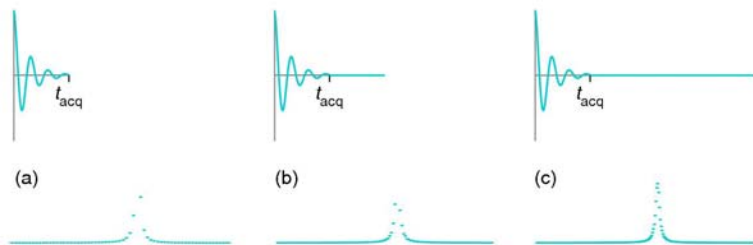
This is equivalent to using a Fourier series with more terms (more basis functions) to describe the signal more accurately. However, there is a limit to the improvement in resolution that can be achieved this way. Two-fold zero-filling is usually enough. (Remember, the FFT needs **np** to be a power of 2.)

Often, you will see "sinc wiggles" at the bottom of peaks. This is a result of a truncated FID. Zero-filling can't help you with that:
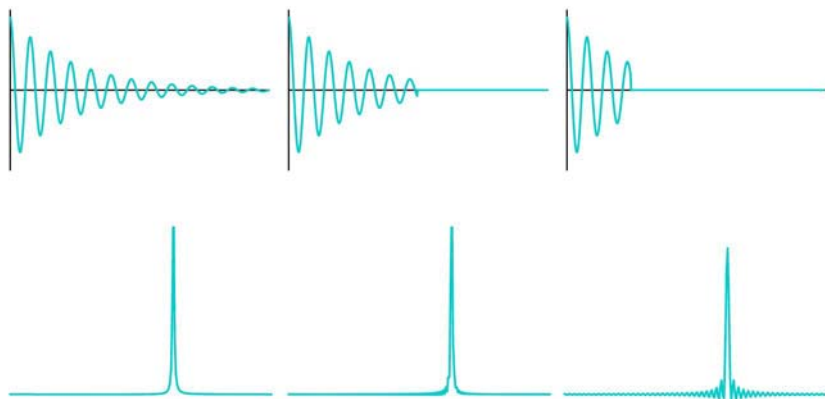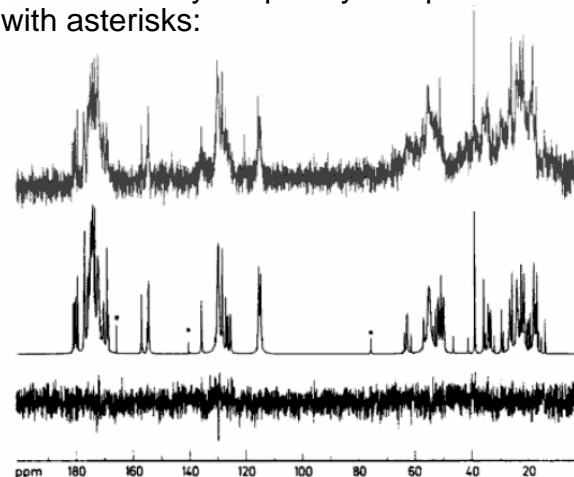

Linear Prediction

Q: How can we deal with truncation?

We know from the vector model that the FID should represent a bunch of oscillating vectors. This is in contrast to some other time series, like the performance of the stock market, where there is no reason to believe that a frequency description of it is valid. In **linear prediction**, we represent a data point S_n as a linear combination of the previous k points:

$$S_n = c_1 S_{n-1} + c_2 S_{n-2} + \cdots + c_k S_{n-k}$$

The value of k is the "order" of the prediction. This basically says that the future is like the past in the FID. In time series analysis terms, this is prediction based on autocorrelation, which is related to the Fourier transform by the Wiener-Khinchin theorem. (It says that the power spectrum is the Fourier transform of the autocorrelation.)

For 1D data sets, this is problematic because there are a very large number of frequency components, and so one needs a very large order of prediction k. Still, if you are very determined, this can be useful. Here is an image from the Roberts book (pg 112). The bottom is the FT of the residuals: there don't seem to be any frequency components left. Artifacts are marked with asterisks:

Linear Prediction

In 2D NMR spectra, it turns out that there are far fewer frequencies per component spectrum, so linear prediction can be used to great effect there. It can also be combined with zero-filling and apodization (although it wasn't for the example below; Claridge, pg 45).

(a) nothing
(b) two-fold zero-fill
(c) linear prediction
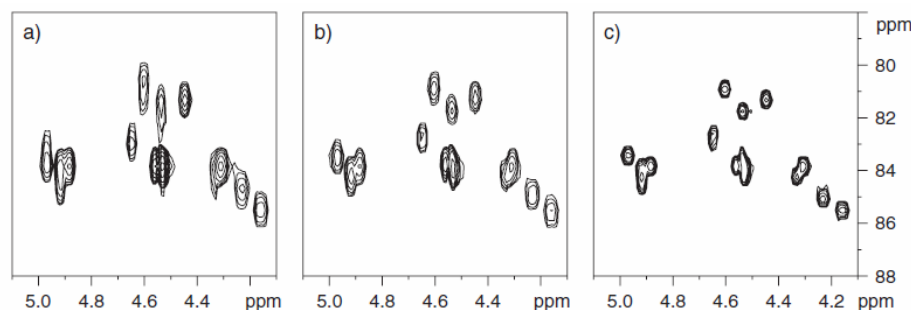
All of this has dealt with *forwards* linear prediction, which extends the FID as if you had collected more points. One can also perform *backwards* linear prediction, which can replace the first few points of the FID. These can become corrupted by the effect of the previous pulse or acoustic ringing, which is where mechanical vibrations in the probe give rise to signals. Here is a picture from Glenn Facey's NMR blog (Ottawa):

FID with acoustic ringing

Distorted NMR spectrum

FID with backward linear prediction

Undistorted NMR spectrum
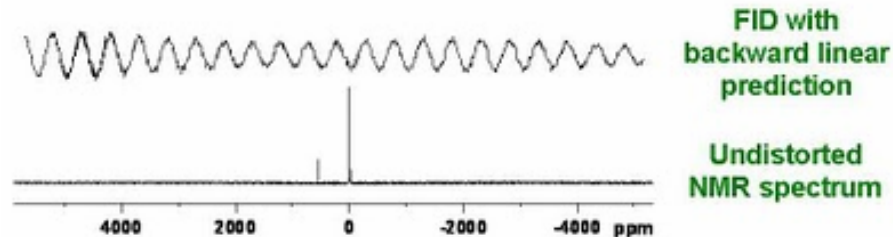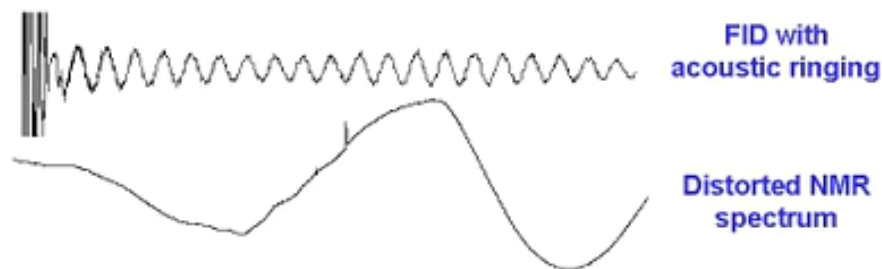
Phase Cycling

This is a complex topic, and here I only show the most basic phase cycle which is built into virtually every NMR experiment, even pulse-acquire. It is called **CYCLOPS**: cyclically ordered phase cycling. Why do we need it? Consider this spectrum (Claridge, pg ...)

Image

Transmitter frequency
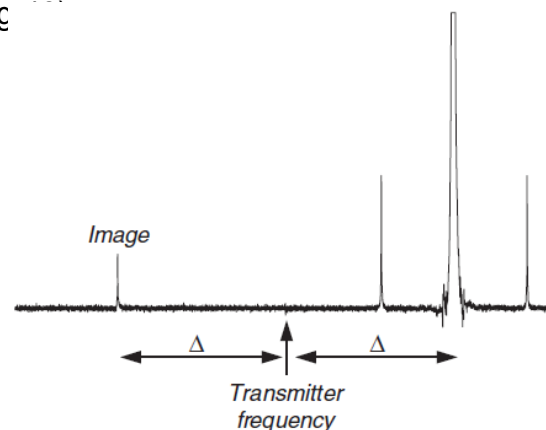
The artifact is a **quadrature image**, which often appears for very intense peaks as a reflection about the midpoint of the transmitter/receiver frequency.

Q: How do quadrature images arise?

The "x detector" and "y detector" may not be exactly 90° out of phase. Thus, the x channel may have some unwanted y-component and vice versa. This is called "receiver imbalance." In the limit of a complete imbalance, where one detector is turned off, one gets full aliasing, as positive and negative frequency discrimination is lost.

Phase Cycling

(One can also have a DC offset error, where the mean of the received signal is non-zero, which leads to an erroneous zero-frequency peak.)

In the CYCLOPS scheme, the phase of the transmitter and the receiver are incremented in 90° increments to form a four-step phase cycle (**see Jacobsen, pg 211**):

scan 1: 90_x pulse; receiver phase: x
scan 2: 90_y pulse; reciever phase: y
scan 3: 90_{-x} pulse; receiver phase: -x
scan 4: 90_{-y} pulse; receiver phase: -y

To see how this works, consider what happens when I apply a 90_{-x} pulse in scan 3. We've already seen this: it rotates the equilibrium +z vector 90° counterlockwise. (*Convention note:* From now on, *all* of our rotations and pulses will be counter-clockwise by default.)
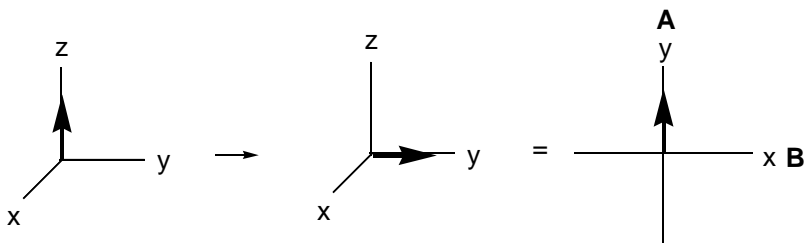

A and B denote the two quadrature channels. Thus, A will receive a cosine signal, while B will receive a sine signal. Remember, with quadrature detection, we want to form the quadrature signal as (ignoring the magnitude):

$$S(t) = M_x + i\, M_y = \cos(\Omega t) + i \sin(\Omega t) = \exp(i\Omega t)$$

This puts cosine in the real part and sine in the imaginary part, giving a purely absoptive Re[FT] and a purely dispersive Im[FT].

However, we cannot do this directly here:

$A = M_y = +\cos(\Omega t)$
$B = M_x = -\sin(\Omega t)$
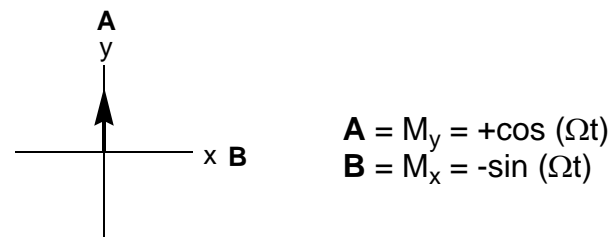
Now, we'll have a problem, because while **A** is cosine, **B** is negative sine. Thus, we must invert the signal from **B**. This is equivalent to moving **B** to the -x axis, which is what the receiver phase means:

pure absorption
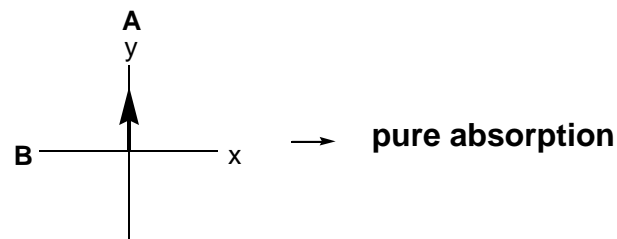
For the whole cycle, we have:

Scan	Pulse phase	Vector starts on	M_x	M_y
1	$+x'$	$-y'$	$\sin(2\pi\Delta vt)$	$-1.1\cos(2\pi\Delta vt)$
2	$+y'$	$+x'$	$\cos(2\pi\Delta vt)$	$1.1\sin(2\pi\Delta vt)$
3	$-x'$	$+y'$	$-\sin(2\pi\Delta vt)$	$1.1\cos(2\pi\Delta vt)$
4	$-y'$	$-x'$	$-\cos(2\pi\Delta vt)$	$-1.1\sin(2\pi\Delta vt)$

In this chart, we assume the **A** / M_y channel is out of balance by 10%. Thus, we can form the FID as follows:

real FID = -M_y(1) + M_x(2) + M_y(3) - M_x(4) = 4.2 cos(Ωt)
imaginary FID = - M_x(1) + M_y(2) - M_x(3) - M_x(4) = 4.2 sin(Ωt)

By collecting the positive cosines and sines from each scan, the good signals progressively add up while the bad signals cancel out. Now, the spectrum is balanced, even though the receivers are unbalanced. Thus, **nt** should be a multiple of 4.