# Processing NMR Data

When you are finished acquiring your NMR data, you will need to process the data into a spectrum and plot that spectrum on paper with a ppm axis, and possibly with integrals, peak lists, and other extras. You may want to expand interesting or complex regions of the spectrum on a separate plot so that the fine structure of peaks (splitting patterns, J values, etc.) can be analyzed. Each NMR instrument has its own software for data processing, and it can be daunting to try to learn all of the different commands and operations. The actual data processing task, however, is the same in all cases and the learning curve will be more efficient if we first deal with these tasks in general without discussing individual NMR programs. Starting with an FID (raw time domain data), we need to carry out the following operations:

1. Multiply the FID by a multiplier or window function.
2. Add zeroes to the end of the FID to increase digital resolution (optional).
3. Fourier Transform the time domain data to obtain a frequency domain spectrum.
4. Correct for Phase Errors by adjusting the Phase.
5. Find a reference standard peak and set its chemical shift to the reference value in ppm.
6. Expand the desired region of the full spectral window to be plotted.
7. Plot the spectrum.

In addition, there are several optional operations we might want to perform:

1. Measure the area under individual peaks by integration.
2. Plot the chemical shifts of peaks on the spectrum, or print a separate list.
3. List the acquisition and processing parameters on the spectrum or in a printout.
4. Expand and plot smaller regions of the spectral window.

To gain a better understanding of what is involved in these steps, we will start with a look at the raw time domain data (FID).

**What is NMR Data?** Raw time domain NMR data (the FID) consists of a list of numbers, usually negative and positive integers, as a function of time in equal time increments. The list is usually quite long, with as many as 16,000 or 32,000 entries. There are two types of data values (reflecting the two channels of the NMR receiver): real and imaginary. Data is arranged in the order real, imaginary, real, imaginary ..., with a pair of data values for each time point in the FID. This is the "raw" data of the NMR experiment. Bruker and Varian have slightly different ways of acquiring the real and imaginary data points as a function of time: Bruker alternately samples the data from the two channels and Varian samples both channels simultaneously. This difference has an effect on the way data is processed, and also control the way in which peaks outside the spectral window are folded or "aliased". The data is contained in a computer disk file (if you saved it on the spectrometer!) as a binary file containing a header (with some information about the spectrometer settings) and a list of numbers without the time values. For Varian Unity files, the FID is saved as a directory (**<filename>.fid**) which contains four files: **fid**, the actual binary data file; **log**, a history of the experiment with exact dates and times of acqusition and any error messages; **text**, a file you can create which describes the sample

and experiment; and **procpar**, a detailed list of the parameters used in the experiment. On the Bruker instruments you simply save the FID as a binary file.

| Time (μs) | Varian (simultaneous) | | Bruker (sequential) | |
|---|---|---|---|---|
| | Real | Imaginary | Real | Imaginary |
| 0 | 1. 23435 | 2. -2344 | 1. 23465 | |
| 80 | | | | 2. 93547 |
| 160 | 3. 836509 | 4. 3496 | 3. -3546 | |
| 240 | | | | 4. 34593 |
| 320 | 5. 45673 | 6. -234 | 5. 23486 | |
| 400 | | | | 6. -24367 |

Some experiments involve more than one FID: for example DEPT analysis performs a ^{13}C experiment four times with four different parameter settings; two-dimensional (2D) experiments involve collections of up to 750 similar FIDs. These can be combined in a single binary file. The FIDs are just listed one after the other in a single continuous list of data which Bruker calls a **serial** file. Varian treats these multiple FID files in the same way as single-FID files, but Bruker requires the extension **.ser** for all serial files.

**The Fourier Transform.** The raw data or FID is a series of intensity values collected as a function of time; it is thus **time-domain** data. A single proton signal, for example, would give a simple sine wave in time with a particular frequency corresponding to the chemical shift of that proton. This signal dies out gradually as the protons recover from the pulse and relax. To convert this time-domain data into a spectrum we perform a mathematical calculation called the Fourier Transform, which essentially looks at the sine wave and analyzes it to determine the frequency. This frequency then appears as a peak in the spectrum, which is a plot in **frequency domain** of the same data. If there are many different types of protons with different chemical shifts, the FID will be a complex sum of a number of decaying sine waves with different frequencies. The FT extracts the information about each of the frequencies, their intensities, and even the rate at which they decay, which determines the linewidth of each peak in the spectrum (signals which decay quickly transform into broad peaks, while signals which last a long time transform into sharp peaks). The method of data collection (Bruker sequential vs. Varian simultaneous) will affect the type of Fourier transform calculation you must perform. This difference is invisible if you process your data on the instrument on which it was acquired, but if you transfer data to a separate workstation and use independent processing software (e.g., Felix), you will have to specify Bruker Fourier transform (**bft**) or complex Fourier transform (**ft**) for Bruker or Varian data, respectively.
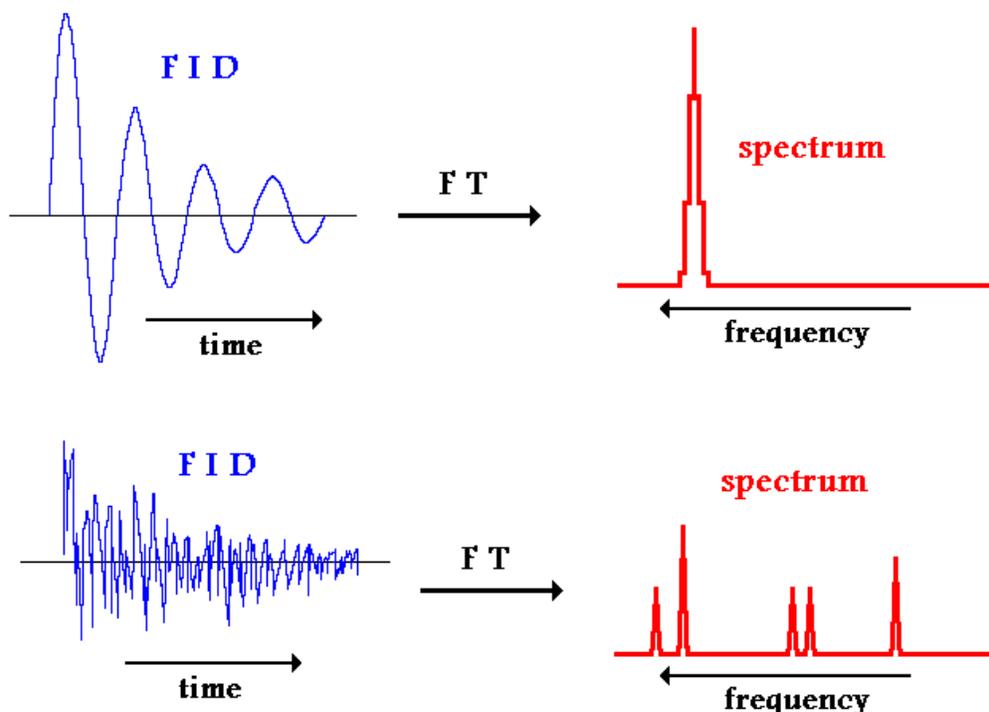
**Data Manipulation Before the Fourier Transform.**  Before performing the FT, there are two things we can do to enhance the quality of the spectrum.  First, the size of the dataset can be artificially increased by adding zeroes to the end of the list of FID data.  This process of **zero-filling** has no effect on the peak positions, intensities, or linewidths of the spectrum, but it does increase the digital resolution (fewer Hz per data point) in the spectrum.  This can be useful to give better definition of peak shapes for sharp peaks.  For example, you might have an FID which contains 4096 total data points (2048 pairs of real, imaginary).  If you transform it directly, you will have 2048 points in your spectrum.  If your spectral width (SW) was 6250 Hz when you acquired the data, your spectrum will have a digital resolution of 6250 / 2048 or 3.05 Hz per



**Low Digital Resolution**
**3.0 Hz / pt.**

**High Digital Resolution**
**0.3 Hz / pt.**
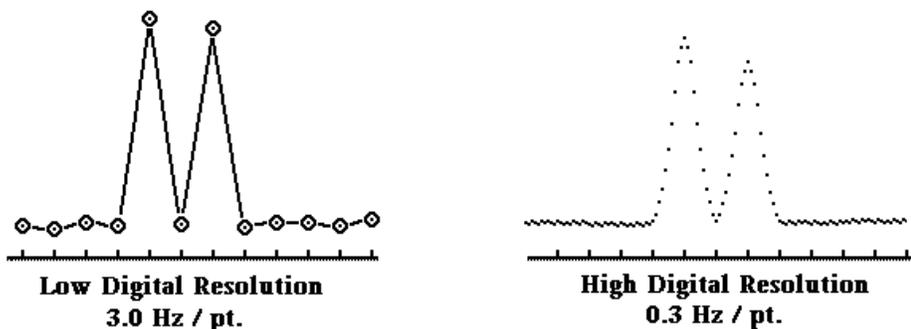
point.  A doublet with a splitting of 8 Hz would be described by only three points, so the measurement of the splitting would be very inaccurate due to the "graininess" of the spectrum.  If, on the other hand, you zero-fill the acquired data by adding 14366 pairs of zeroes to the data list before FT, you will get a spectrum with 16384 (16K) data points describing the full 6250 Hz

spectral window.  The digital resolution is much greater (6250 / 16384 = 0.381 Hz per point) and the same doublet would be described by 21 data points.  Zero-filling is accomplished by simply defining the data size before FT (Bruker **SI**, Varian **FN**) to a larger number than the acquired data (Bruker **TD**, Varian **NP**).  In the above example, you would set **TD** (**NP**) to 4096 and **SI** (**FN**) to 16384.

       A more common pre-FT massaging of data is the application of a **window function** or weighting function (Fig. 15).  The idea is to emphasize ("weight") certain parts of the FID at the expense of others.  For example, suppose that your FID signal dies out fairly rapidly so that there is nothing but noise after 0.2 seconds, even though you acquired data up to 1.0 seconds.  The noise from 0.2 to 1.0 sec. in your FID only increases the noise in your spectrum, and doesn't contribute to the peak height, so your signal-to-noise ratio is reduced.  One solution would be to simply set all the data after 0.2 sec. to zero, but this introduces a sharp discontinuity in the FID at 0.2 sec. which could introduce artifacts into the spectrum.  A smoother method is to multiply the FID by an exponential decay function which emphasizes the early data in the FID and de-emphasizes the later (mostly noise) data.  The "steepness" of this exponential multiplier can be varied so that it matches the natural decay of the signal.  The net effect on the spectrum is that the signal-to-noise ratio is increased and the peaks are somewhat broader, because faster decay of the signal leads to a broader peak.  If the line-broadening effect is not significant, the increase in S/N is usually worth the price, especially for carbon spectra where signal is always weak.  The steepness of the exponential multiplier is set by a parameter called line-broadening (Bruker and Varian **LB**) which is the additional line-width in Hertz which will result from this process.  I usually use an **LB** value of 0.2 Hz for proton spectra and 1.0 Hz for carbon spectra.  If you have a very weak signal carbon spectrum and just want to see if there are peaks, you can use an **LB** of 3.0 or 5.0 Hz.
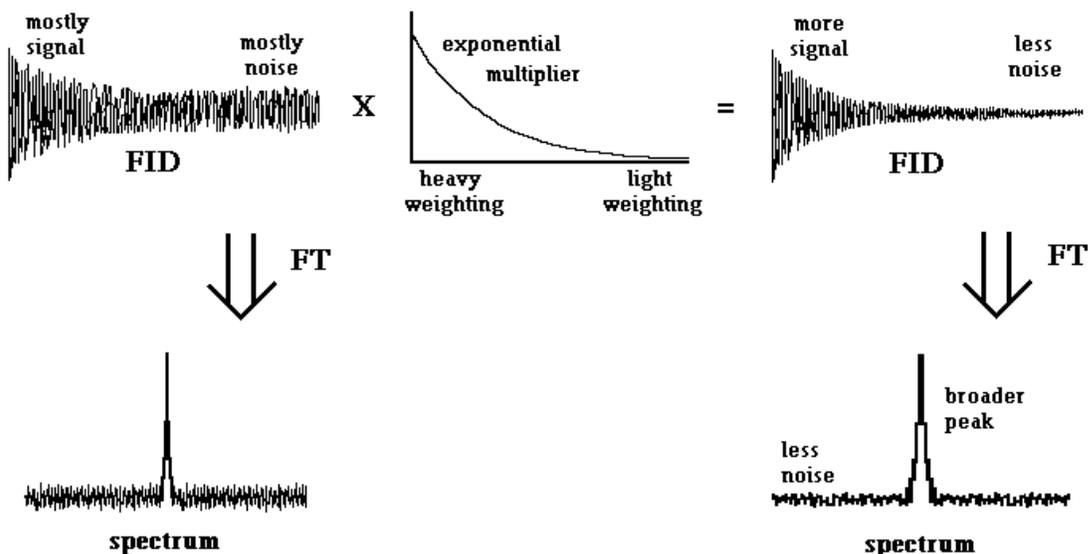


       Other window functions can be used for the opposite effect:  **resolution enhancement** (Figure 15).  By de-emphasizing the beginning of the FID and amplifying the later part, the natural decay of the FID signal is delayed and the peaks in the spectrum get sharper.  This is especially useful for measuring coupling constants.  Of course, there is no free lunch so you pay a

price in poorer signal-to-noise ratio, but with some samples you have more signal than you could ever want. Resolution enhancement also usually causes "ditches" on either side of a peak. Examples of window functions for resolution enhancement are sine-bell (just the function **sin(x)** for x = 0 to 180 degrees) and gaussian (the old bell curve). These window functions are very important for processing two-dimensional (2D) NMR data.
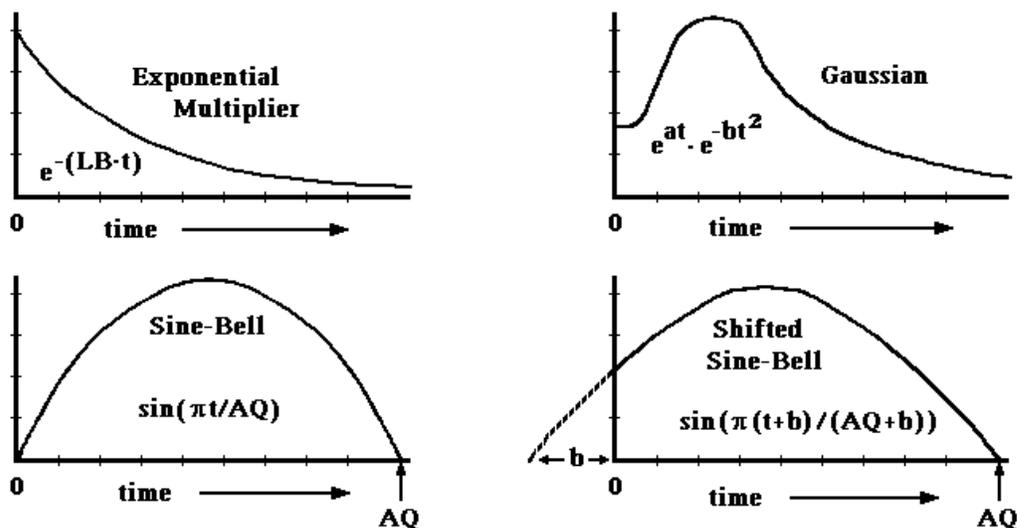


**Figure 15.** <u>Window (Multiplier) Functions.</u> The exponential multiplier emphasizes the early time points in the FID and de-emphasizes the later points, which are mostly noise. The effect is to increase the signal-to-noise ratio at the expense of degrading the resolution (i.e., broadening the peaks). Other window functions bring the FID to a small value or zero at the end of the acquisition time but emphasize the middle time points of the FID. The result is resolution enhancement (sharper peaks) at the expense of degrading the signal-to-noise ratio. Extreme resolution enhancement will usually lead to wiggles ("ditches") around a peak.

Bruker uses the command **EM** (exponential multiplication) to implement the exponential window function, so a typical processing sequence on the Bruker is **EM** followed by **FT** or simply **EF** (EF = EM + FT). Varian uses the general command **wft** (weighted Fourier Transform) and allows you to set any of a number of weighting functions (**lb** for exponential multiplication, **sb** for sine-bell, **gf** for gaussian function, etc.). Executing **wft** applies the window function to your FID and then transforms it.

**Phasing.** After you Fourier-transform your FID you get a frequency-domain spectrum with peaks, but the shape of the peaks may not be what you expected. Some peaks may be upside-down, while others may have a "dispersive" (half up - half down) lineshape (Fig. 16). This is because the spectrum has to be "phased" after FT to give the correct "absorptive" peak shape. Phasing corrects certain unavoidable instrumental errors involved in acquiring the FID. Recall that the raw NMR data (FID) consists of two numbers for each data point: a real and imaginary value. After the Fourier transform, there are also two numbers for each frequency point: real and imaginary. In a perfect world, the real spectrum would be in pure absorptive mode (normal peak shape) and the imaginary spectrum would be in pure dispersive (up / down) mode. In reality, each spectrum is a mixture of absorptive and dispersive modes, and the
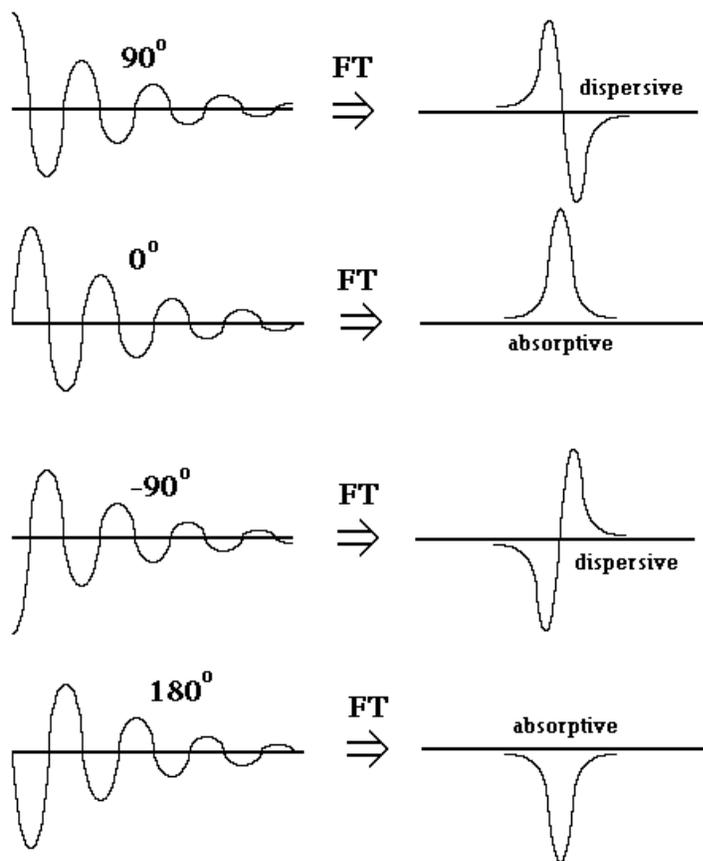
proportions of each can vary



**Figure 16.** <u>Phase.</u> Where the FID signal starts in terms of the sine function determines the phase of the transformed spectrum peak. If the FID starts at the (positive) value of the sine function at 90°, the transformed spectrum will give a dispersive line shape, with equal peak heights above and below the baseline. If the FID starts at the zero value of the sine function at 0°, the peak will have the normal absorptive phase. FIDs which start at -90° and 180° will give peaks with upside-down dispersive and absorptive phase, respectively. Imperfections in the RF electronics and the variability of samples makes it impossible to start the FID at 0° for all acquisitions, so phase correction is necessary in data processing to get absorptive line shape.

with chemical shift (usually in a linear fashion). To correct for this, we calculate a linear combination of the real spectrum and the imaginary spectrum, and use this for our "phased" spectrum (Fig. 17). For the mathematically inclined, the actual linear combination is:

$$\text{Phased spectrum} = (\text{Real spectrum})*\cos(\theta) + (\text{Imaginary spectrum})*\sin(\theta)$$

and the problem of phasing boils down to finding the correct phase correction angle $\theta$. Well, actually it's a little more complicated because the phase correction $\theta$ is usually a linear function of the chemical shift ($\delta$). To define the line:

$$\theta(\delta) = (\mathbf{m} * \delta) + \mathbf{b}$$

requires that you determine two parameters: the intercept **b** (called the zero-order phase correction) and the slope **m** (called the first-order phase correction). All phasing routines are based on optimizing these two numbers. The process always starts with choosing a large peak at one end of the spectrum as the "pivot" peak: this is the peak which is defined as $\delta = 0$ for the linear correction. The phase of this peak is optimized by varying the intercept (**b**) value. Then another peak is chosen at the other end of the spectrum and its phase is optimized by adjusting the slope (**m**) parameter. The key to this method is that changing the slope has no effect on the

41

"pivot" peak because we chose it to have $\delta = 0$.  The actual zero of the chemical shift scale is not
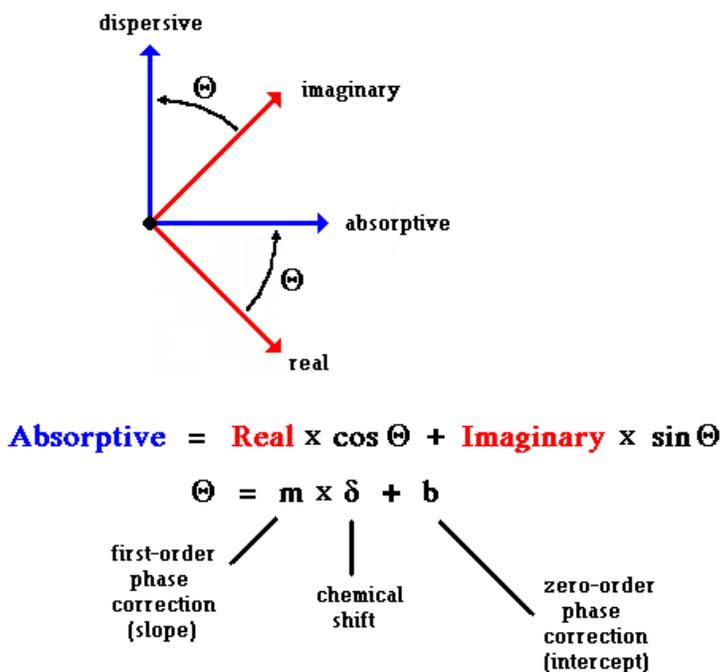


**Figure 17.**  Phase Correction.  The real and imaginary spectra, which result from Fourier transformation of the real and imaginary FIDs, are usually rotated by an arbitrary angle from the pure dispersive and absorptive lineshapes due to unavoidable imperfections in the recording of the FID.  Phase correction is the process of rotating the spectrum by an angle $\Theta$ so that the real part corresponds to pure absorptive lineshape and the imaginary part (which is discarded) corresponds to pure dispersive lineshape.  In fact, this correction angle usually varies with chemical shift in a linear fashion, so that the correction consists of chemical shift independent correction (zero-order correction, b) and a chemical shift dependent correction (first-order, m).

important - this is just for the purpose of the phase calculation.  With both parameters set, the line is defined and all peaks in between should also be correctly phased (Fig. 18).

There is one situation where this will not be the case:  it is possible to set **m** to a large value such that the second (non-pivot) peak is given a phase correction which is $360^{\circ}$ too large (Fig. 19).  This will not affect the shape of the peak, but will introduce a "phase twist" to all the peaks in between.  For example, a peak exactly between the pivot peak and the second peak in chemical shift will have a phase error of $180^{\circ}$ and thus will appear upside-down.  Even greater phase twists ($720^{\circ}$, $1080^{\circ}$, etc.) can be applied if you're not careful.  If you get into to this situation, it is best to adjust the first-order phase correction by looking at a peak which is close to the pivot peak, and then moving to peaks farther and farther away.



**Fine Tuning the Phase Correction**
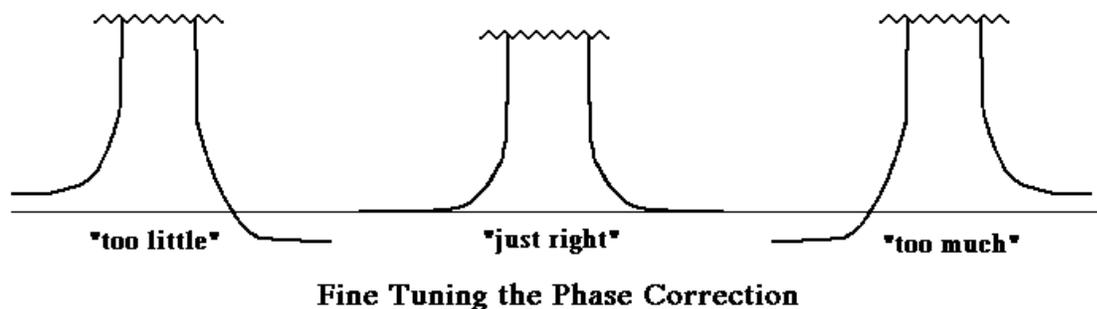
When you get close to the correct phase parameter, focus your attention on the baseline (or noise line) on either side of the peak in question.  This should be at the same (vertical) level on each side of the peak.  Sometimes it's useful to increase the vertical scale before phasing so

42

that these baseline differences are amplified.  Varian has an automatic phase routine (command **aph**) which will often get you in the ballpark or even take the place of manual phasing.
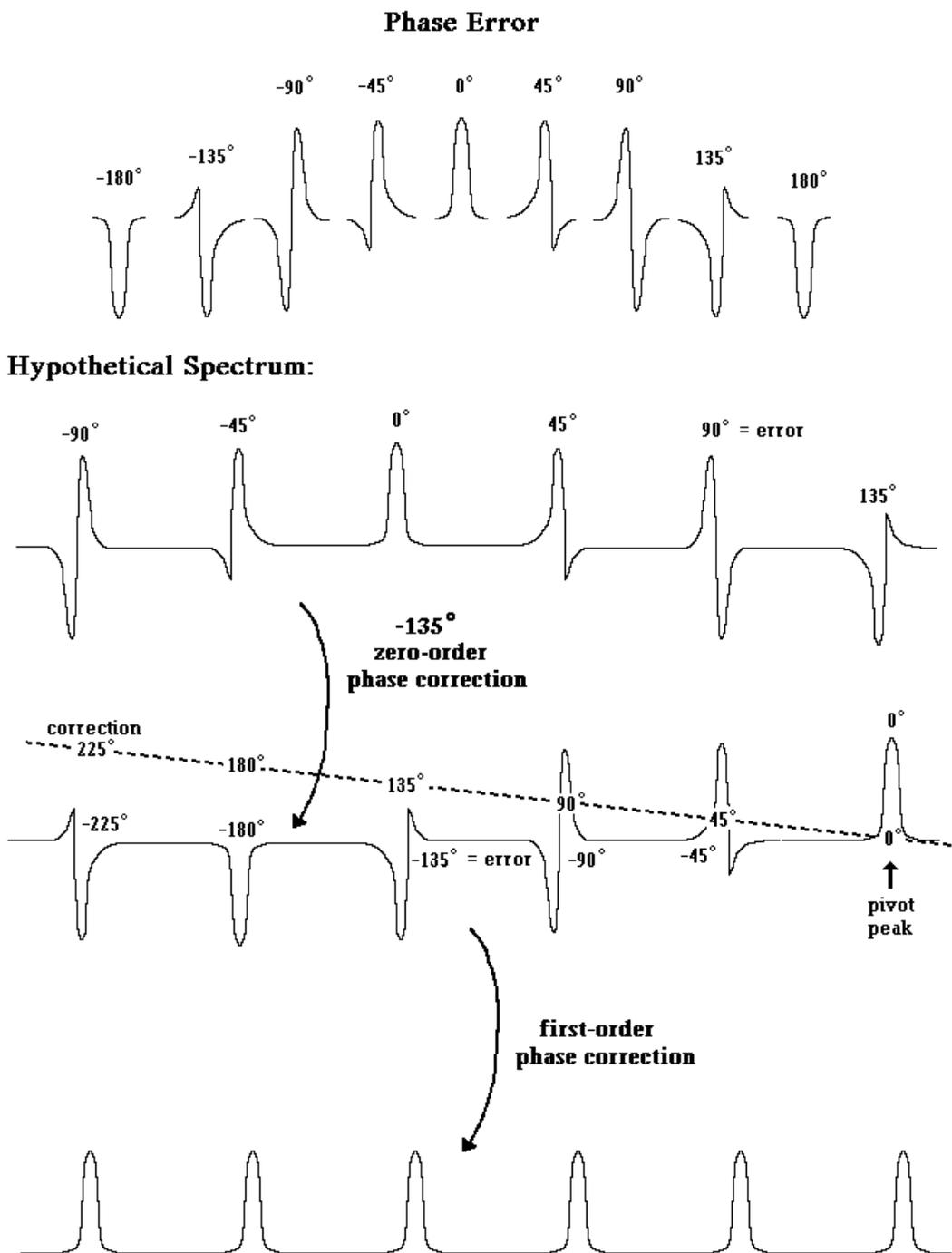


**Figure 18.**  Phase-Correcting a Hypothetical Spectrum.  The shape of peaks with phase errors ranging from -180° to +180° is shown at the top.  A hypothetical spectrum has a range of phase errors which vary linearly with chemical shift.  Selection of a pivot peak at the upfield end and application of a -135° zero-order phase correction to correct this peak subtracts 135° from all of the phase errors, since the correction is applied equally across the spectrum. Then the slope of a line which goes through zero at the pivot peak is adjusted to correct the phase of all the other

peaks. This first-order phase correction will not affect the phase of the pivot peak. After both corrections are applied, all of the peaks are in pure absorption mode.
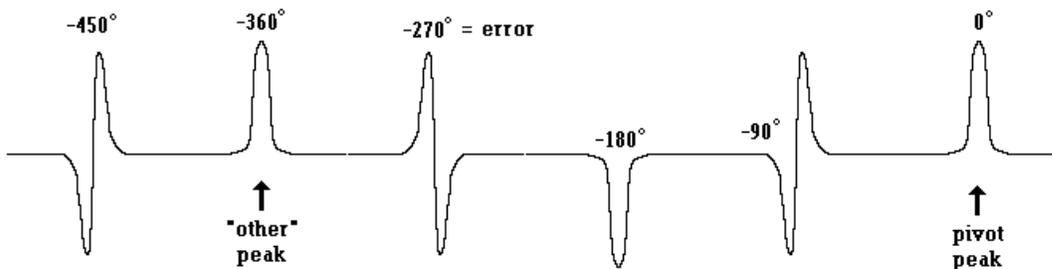


**Figure 19.** <u>Phase Twist.</u> An opposite first-order phase correction changes the phase error of the second peak from -180$^o$ to -360$^o$, with all the peaks in between this peak and the pivot peak having phase errors intermediate between -360$^o$ and 0$^o$. The solution to this problem is to use a peak closer to the pivot peak to make a coarse adjustment of the first-order phase correction, and then fine-tune this correction using peaks further away from the pivot peak.

**Setting the Reference.** This is a simple procedure whereby a reference peak (e.g., TMS in organic solvents) is selected with a cursor (Bruker uses a triangle, Varian a vertical line) and given a chemical shift value. Without this reference, the chemical shift scale of your spectrum will be meaningless.

**Peak Lists.** You will often want a printed list of chemical shifts for all the major peaks of your spectrum. First you have to set a threshold intensity (Bruker minimum intensity **MI**, Varian threshold **th**) below which a peak is not included in the list. If you set the threshold too low, you will get a very long list which includes many noise intensities; if you set it too high, you will miss real peaks. Peak lists can either be displayed on the screen next to each peak (Varian only), plotted on the spectrum next to each peak, or printed out as a list on a printer. The last possibility is especially useful for determining coupling constants, since you can get a list with both ppm and Hertz values for each peak; subtraction gives the J values in Hertz.

**Baseline Correction.** The "baseline" is the average of the noise part of your spectrum. Ideally, this would be a straight, horizontal line representing zero intensity. In the real world it can drift, roll, and wiggle like a drunken sailor. These errors general result from erroneous data which is collected at the very beginning of the FID, when the electronics is still recovering from the shock of the exciting RF pulse. This becomes a problem when you try to measure peak areas (see below: integration). Varian makes it very easy to correct baselines. First you need to indicate where you have peaks and where you have noise in your spectrum; this is done by indicating integral regions as part of the integration process. Once this is done, you simply execute the command **bc(1)** and the baseline is corrected by a mathematical fitting of the noise portions of the spectrum to a smooth function, which is then subtracted from the whole spectrum including peaks. Bruker has a manual adjustment of polynomial correction parameters, but it is essentially useless. The workstation processing package "Felix" has a variety of very sophisticated baseline correction methods.

**Integration.** To get quantitation of peak areas (numbers of protons) you need to plot an integral. In the old days before Fourier transform NMR, the plotter was set to integral

mode and the pen was swept through the peak as the pen level rose with the integrated intensity. For this reason, integrals are still presented as lines which start at the left side of a peak and rise vertically as they pass through the peak. In FT NMR there is often a problem with baseline "wiggle" and this will lead to inaccurate integration of proton peaks (carbon peaks are essentially never integrated because their peak areas are determined more by differences in relaxation rates than by differences in the number of carbons). To get good integrals, you may need to correct the baseline first (see above). Integration generally involves adjusting the height of the integrals, indicating the start and end points of each peak integral, correcting the drift so that noise regions give a horizontal line in the integral, and normalizing the peak areas so that numbers of protons can be read directly. All processing software allows you to plot the integral areas directly on the spectrum next to each integral, or to print out a list of integral values. Details of how these steps are accomplished are specific to the software being used and will not be dealt with here.

**Plotting.** A hardcopy of your spectrum can be obtained using the Hewlett-Packard (HP) plotters. These are very expensive devices which roll the paper back and forth (X-axis) as the pen moves up and down (Y-axis). On some plotters, different colors can be used for spectrum, axis, integrals, and parameters. There are a number of things you can include in your plot:

     Spectrum
     Integrals
     Integral Areas (numerical values)
     Scale (x-axis in ppm)
     Peak position labels (in ppm)
     Parameters
     Title
     Text Describing Sample, Experiment (Varian only)

Plots can be made on "normal" paper (8.5" X 11") or "large" paper (11" X 17"). Multiple spectra can also be plotted on the same paper, either side-by-side or one above the other. The plotting procedures for Varian, Bruker and the SGI workstation software (Felix) are quite different so these details will not be covered here.

**Archiving (Saving) Your Data.** If you had the foresight to save your data on the NMR instrument's hard disk, you will find that this data is periodically "purged" as the disk gets full. Why save it forever? Someday you will be writing up a paper or thesis and will ask the inevitable question, "what was the coupling constant for that triplet at 3.5 ppm?". Since you probably didn't anticipate this question when you plotted your spectrum, you will need to get the data back and re-process it. At that point you will thank yourself profusely for having the foresight to archive your data. NMR data is generally archived in the raw (FID) form so that you have the maximum flexibility in processing it. Data can be transferred from the Bruker AM-250 console to a PC which is on the internet. From the PC, it can be saved on standard $3^{1/2}$" X $3^{1/2}$" DOS format floppy disks (1.44 MByte capacity) or on tape, or it can be sent by **ftp** (file transfer protocol) over the internet to any other computer (Fig. 20). The most likely destination is the main NMR facility's Silicon Graphics workstation ("UANMR"). Bruker gives data file sizes in terms of the number of data points, so an FID with 16,384 data points will appear in Bruker's
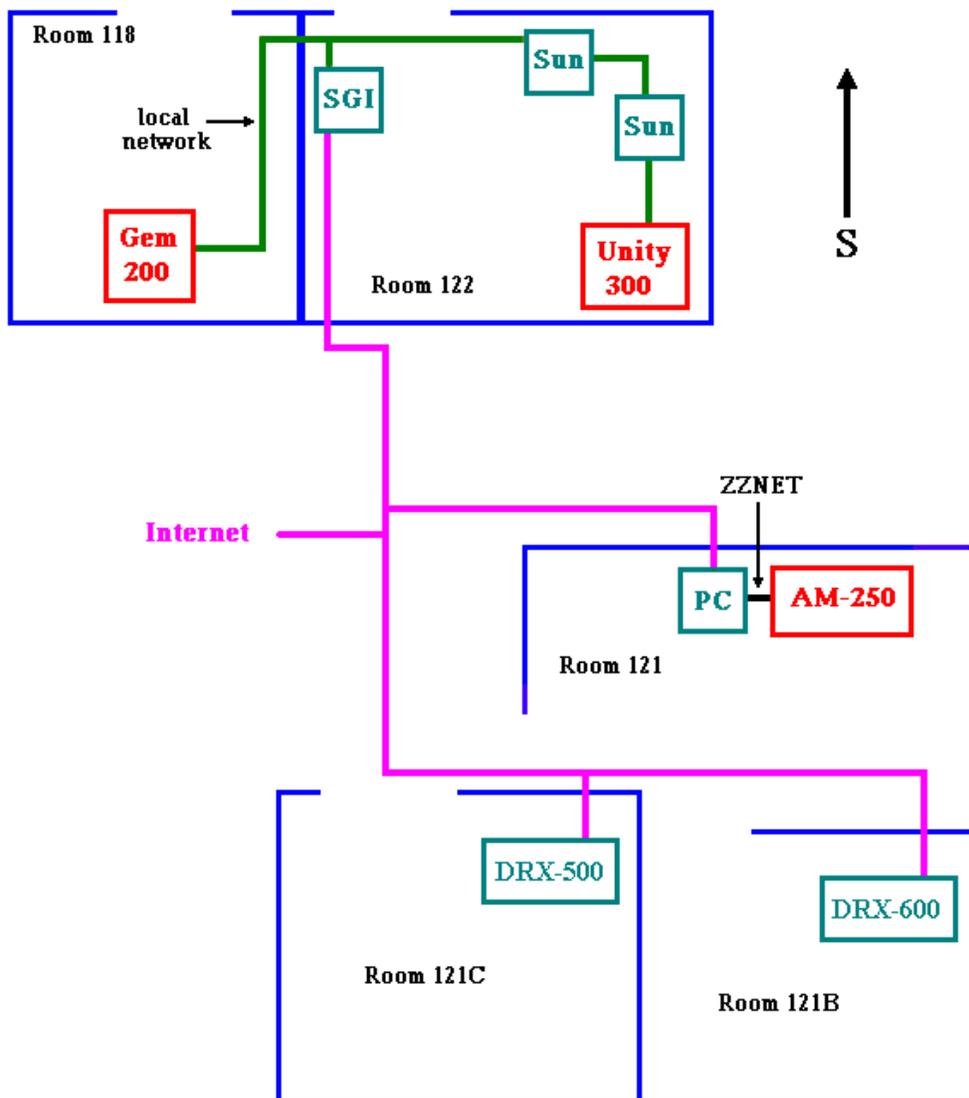
**Figure 20.** The NMR Facility Computer Network. A local ethernet network connects the Sun workstations in the Unity room (Bloch, Ernst, and Unity300) to the Gemini spectrometer (Gem200) and the facility SGI workstation (UANMR), which is also on the Internet. Tape drives (8-mm) are connected to Ernst and Bloch, the Sun workstations. The hard disk units /felix and /datax are mounted simultaneously on Ernst, Bloch and Unity300. Data transfer from Gem200 to the Sun stations uses a special transfer protocol. The Campus-Wide NMR Server (SGI Indy workstation "UANMR") can be operated remotely over the Internet using a PC as an X terminal. The Bruker AM-250 instrument is connected to a PC using the ZZNET transfer software, and this PC is on the Internet. The SGI O2 (Unix) computers which operate the Bruker DRX-500 and DRX-600 spectrometers are directly on the internet.

as a 16K file. But because each data point uses 3 Bytes (AM) or 4 Bytes (DRX) of data, the actual file size is a little more than 48 kBytes on the AM (the "little more" is for a file header) or 64 kBytes on the DRX (no file header). When you transfer it from the AM-250 to the PC (DOS) it will thus appear to be a little over three times bigger.

The Bruker DRX-500 and DRX-600 disks are mounted on UANMR, so that you can

process your data in room 122 immediately after it is acquired, using the same XWinNMR software which operates on the spectrometers.  Thus UANMR (the SGI Indy workstation) serves as a satellite data processing workstation for the Bruker DRX spectrometers in the same way that the Sun workstations Ernst and Bloch serve as satellite stations for the Varian Unity-300.  One advantage is that data can be processed from your own PC using X-windows software.  The actual program (XWinNMR) is running on UANMR and your PC functions as a remote graphics terminal.  You can even set it up to print spectra on the printer attached to your PC.

Data acquired on the Varian Unity-300 can be directly saved to 8 mm tape or transferred to UANMR.  Varian Gemini data can be transferred to any of the Sun workstations, or you can save the Gemini data on a $5^{1/4}$" floppy disk and read it on a PC using special software called V_Helper.  Data on the Sun workstations can be processed using VNMR (Varian's UNIX-based software package), or archived on the large (8 mm) tapes.  You also have the option of processing either Bruker or Varian data using the Felix software package on the Campus-Wide NMR Server (UANMR), an SGI Indy workstation which can be operated remotely from any high-end PC on the network.  Felix is in many ways superior to either Bruker's or Varian's data processing software, especially for 2D data analysis.  To process data using Felix, you convert your data files from Bruker or Varian format to Felix format using the programs **vnmr2felix** or **drx2felix**.  Accounts on UANMR with with XWinNMR and/or Felix are free.

The best way to archive your NMR data is to "burn" it onto a CD-ROM using the CD writer attached to Avogadro, a PC in the Computer Graphics Facility (CGF).  The data can be transferred from UANMR to Avogadro using the FTP program.  Since the Unity-300, DRX-500 and DRX-600 data disks are all mounted on UANMR, you only need to transfer from UANMR to Avogadro.  Data on the AM-250 is archived by the NMR facility and put onto CD's which you can check out.