

MR Image Encoding

Introduction

Much of the success and flexibility of MRI is derived by its peculiar methodology; a set of techniques which has proved to be very flexible and informative for probing the properties of complex materials (such as the brain). MR imaging is fundamentally different from other types of imaging. Conventionally, when physicists refer to “imaging” they refer to a scattering experiment. Since our eyes “see” a rock by forming a 2 dimensional array of light intensity amplitudes being scattered off the rock, the instrumentational equivalent of this has come to be synonymous with imaging. In the canonical scattering experiment, “rays” are aimed at an object and then detected as they either scatter off of the object or penetrate through it. The rays can be deflected, lose or gain energy, deflect and then interfere with one another or even be converted from one type of ray to another. The “rays” in question are usually electromagnetic radiation (radio waves, microwaves infra red light, visible light, ultraviolet light, x-rays, or gamma rays) but can, of course, be almost anything including sound waves, water waves, or matter waves (particles). The particles used could be electrons (as in electron microscopy), or any of the zillions of other subatomic particles, or clumps of particles such as nuclei, atoms, molecules, or pieces of dirt.

A fundamental limitation of scattering experiments is that the resolution of the image cannot exceed the wavelength of the wave (electromagnetic or matter wave) used to probe the object. This is fundamentally derived from the uncertainty principle but has long been understood in classical optics. Physicists are fond of making such general statements, but, while true for scattering experiments, this imaging “law” does not hold for MRI. In MR, we use radio waves with a wavelength of several meters to image at a sub-millimeter resolution. Thus we exceed this “fundamental” imaging law by several orders of magnitude. How? The short answer is that we don’t determine the distribution of the body’s protons by bouncing stuff off of them; we determine it by asking them to report to us where they are. We query the body’s spins using a burst of radio waves (the excitation RF), and they report back some milliseconds later with a faint radio signal of their own. They declare their location and other even more valuable information encoded in the frequency and phase of the burst of RF energy they emit in response (the MR signal or echo).

MR Image Encoding

Review of the basic NMR experiment The previous lectures reviewed the basic equations of motions of the proton spin placed in a static, uniform magnetic field. The state of a given spin is either parallel or antiparallel to the static field. Since the parallel state is slightly energetically more favorable, a slim majority are found in this state. When the vector sum of all of the spins is considered, there is almost a complete cancellation between the aligned and anti-aligned spins; the net magnetization consists of the only the slight excess aligned with the magnetic field. Since the other spins always have a canceling partner, we do not detect them and only the small (~0.01%) excess aligned magnetization, M_0 , will be considered. Although quantum mechanics tells us that a measurement of the energy of a single spin system will result in only 2 answers (the energy of the aligned or anti-aligned state.) We know that the spin’s wavefunction can exist in a time dependent superposition state of these two energy eigen states. The spin’s state can also have a projection along the x or y axis. The ensemble average of these superposition states approximates the classical gyroscopic precession equation described when a suitably large group of non-interacting spins is considered.

Encoding basics

The information about how this magnetization is distributed in the body is derived by the frequency and phase of its precession during the detection phase of the experiment. After excitation, the detected MR signal processes with a frequency given by:

$$\omega = \gamma B(t,x,y,z) \quad \text{where } \gamma = 2\pi \cdot 42.577 \text{ MHz/Tesla for protons} \quad (1.1)$$

Here $B(t,x,y,z)$ is the total value of the magnetic field at location (x,y,z) at time t . The phase picked up in time τ after its initial excitation is given by:

$$\varphi(\tau) = \int_0^\tau \omega(t) dt = \gamma \int_0^\tau B(t,x,y,z) dt \quad (1.2)$$

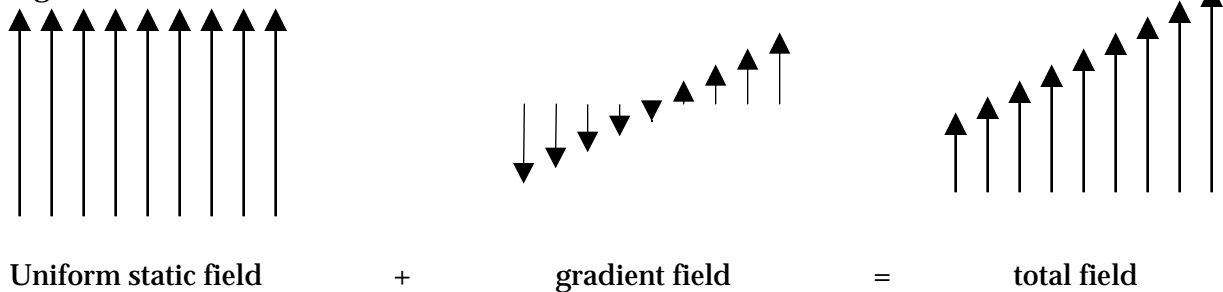
If B is uniform through the sample, then $\vec{B}(x,y,z) = B_0 \hat{z}$ and no interesting spatial information is learned from observing the frequency or phase of the spins precession. Since we can experimentally control $B(x,y,z)$, we can introduce a spatial dependence to phase and frequency by making B vary across the object. The easiest way to do this is to apply a gradient to the static magnetic field.

Field gradients

A linear gradient is the simplest form of variation of the static field; it is a linear increase in the static z field as a function of position. Since the static field is produced by current running through a large coil of wire (the magnet) and the gradient field is added to the uniform field by injecting current into an additional winding, we can easily switch between a uniform magnetic field and the gradient field. When an “x gradient” is applied the field as a function of position is:

$$\vec{B}(x,y,z) = B_0 \hat{z} + G_x x \hat{z} \quad \text{where } G_x \text{ is defined as: } G_x = \frac{dB_z}{dx} \quad (1.3)$$

Fig. 1.1



$$\vec{B}(x,y,z) = B_0 \hat{z} + G_x x \hat{z}$$

Since the constant B_0 part is uninteresting for image encoding, we often just consider the frequency offset from the reference frequency at the center of the magnet (isocenter):

$$\omega = \gamma G_x x \quad (1.4)$$

The applied gradient strength and is commonly measured in Gauss/cm (CGS units) or milliTesla/m (MKS units); 1G/cm = 10mT/m. State of the art body gradient coils can produce gradient strengths of 40mT/m. Thus, when such a gradient is on in a 1.5T magnet, the total field at $x = +10\text{cm}$ is 1.504T, only a small perturbation to the main static field.

The magnetic field gradient can be generalized to a vector: $\vec{G} = (\frac{dB_z}{dx} \hat{x}, \frac{dB_z}{dy} \hat{y}, \frac{dB_z}{dz} \hat{z})$ (1.5)

Therefore, in general the frequency and phase of the detected MR signal is:

$$\omega = \gamma \vec{G}(t) \cdot \vec{r} \quad (1.6)$$

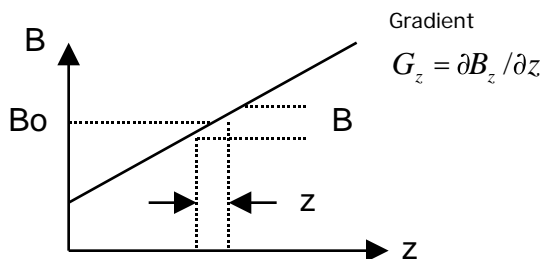
$$\varphi(\tau) = \int_0^\tau \omega(t) dt = \gamma \int_0^\tau \vec{G}(t) \cdot \vec{r} dt \quad (1.7)$$

Here we have considered the important case where the gradient might be changing with time and have therefore written $G(t)$. Typically we think of the frequency as being determined by whatever B field and thus gradient is present at that the time of the measurement. Of course, it takes multiple samples of the signal to calculate a frequency, so the gradient could be varying during the time of the frequency measurement. But, the simplest and most common MR imaging strategies do not vary the applied gradient during the measurement period. In contrast, the phase shift is the result of a time evolution of the signal during the entire time period between the excitation of the magnetization and the time of the measurement. Even the simplest MRI methods have considerable alteration of the gradients during this period, so we explicitly include the time dependence of the applied gradient in the equation for phase.

Slice selective excitation of a single plane through the body

In order to take 2 dimensional images which “cut” through the body, it is useful to excite only a 2D plane of spins. Then the signal arises only from this slice. The image is formed by encoding the two in-plane directions. The slice selection process is achieved by applying the RF pulse to tip the spins at the same time as a gradient. To excite a slice of spins in the xy plane, a gradient G_z in the z direction is used.

Fig. 1.2



As before the resonance frequency of the spins during the z gradient is:

$$\omega(z) = \omega_0 + \gamma G_z z$$

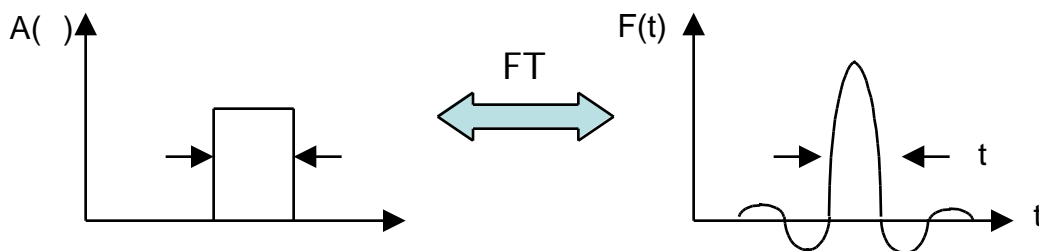
Note that at $z=0$ (isocenter) the frequency $\omega = \omega_0$. If the frequency of the excitation RF pulse is $\omega = \omega_0 + \Delta\omega$, then the z location of the excited spins (and thus the slice plane) will be:

$$z = \Delta\omega / \gamma G_z$$

If the excitation pulse contains a range of frequencies $\Delta\omega$, then the slice will be centered at the above z position but also excite spins included in the slab $z \pm \gamma G_z \Delta z / 2$, where $\Delta z = \Delta\omega / \gamma G_z$. Recall from the time-bandwidth theorem of the FT that it is impossible to form a pulse of finite time duration that does not include a range of frequencies.

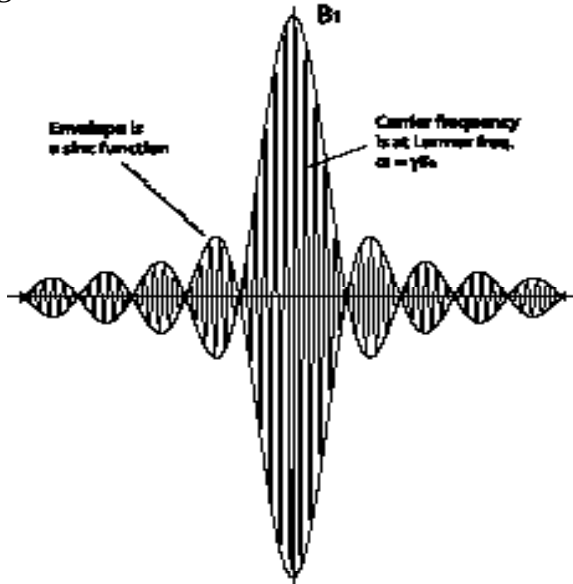
The exact nature of the range of frequencies incorporated in the RF pulse is important for determining the shape of the slice profile. We control the frequencies present in the RF pulse by intelligently choosing the shape of the RF pulse envelope (its amplitude as a function of time). Since a clean “square” slice profile is desired, we seek a time shape with a square frequency spectrum. Recall:

Fig. 1.3



Thus a sinc function with lots of side lobes provides a well defined slice profile. The amplitude is chosen so that the desired flip angle is achieved: $\theta = \gamma \int_0 B_1(t) dt$

Fig. 1.4



RF excitation pulse (sinc envelope)

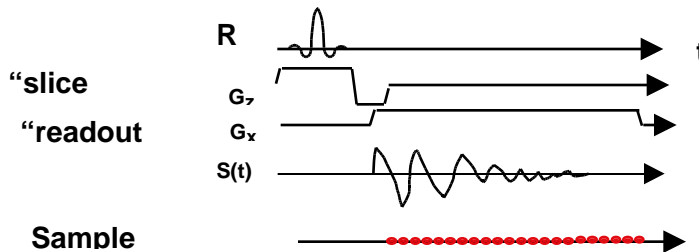
In general the analysis of the pulse shape using the Fourier transform is only a good approximation for small flip angles. Thus, a more complicated analysis is needed for larger flip angles, especially slice selective inversion pulses (180°) (see Pauly et al.).

An additional complication is that during excitation in the presence of a gradient, the spins are precessing with a frequency which depends on their location in the slice profile. Since the excitation process takes a non-zero amount of time (usually between 1 and 5ms), the spins will end up with a phase shift which is a function of position in the slice direction. If this phase is allowed to remain, the experiment will start off with the signal partially dephased. The loss due to the partial cancellation can be recovered by simply reversing the sign of the gradient after the excitation pulse for just the right amount of time to undo the dephasing of the excitation gradient. In practice the rephase lobe is $\sim 50\%$ of the time-area of the slice selection lobe.

Physical picture of what a gradient pulse does to the spins

Since the frequency of the spins precession in the presence of a gradient field is a linear function of frequency and thus a linear function of position, it's easy to imagine that we can learn where a given group of spins is located (in one direction at least) by measuring their frequency. All we need to do is deconstruct the time domain NMR signal (in the presence of a gradient field) into a histogram of the frequencies present in the signal (the spectrum). We do this by taking the Fourier transform of the observed signal. For obvious reasons, this method is called frequency encoding. The particular use of a gradient during acquisition of the signal is called the "readout gradient" and the direction of the gradient (and thus the encoded direction) is called the "readout direction". A diagram of the experiment might look like:

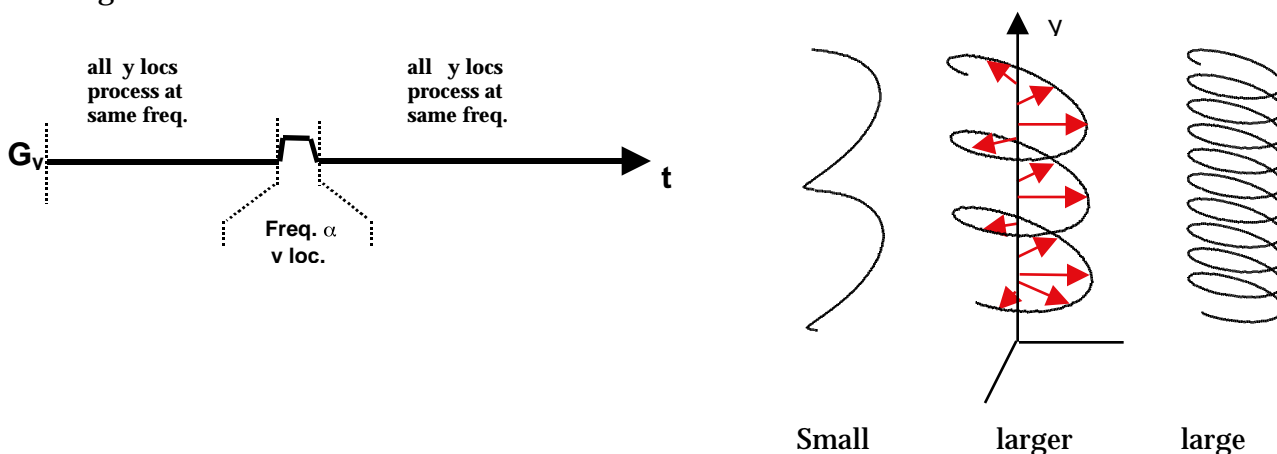
Fig. 1.5



Similar information is contained in the phase of the signal. The phase of the MR signal from a voxel full of spins depends on their position and the entire history of the gradients applied from excitation (definition of zero phase) to measurement. Although less intuitively obvious how to utilize the phase information, several important aspects of MR image encoding are best understood by examining spatial pattern in phase after application of a gradient. Consider what happens after a y gradient

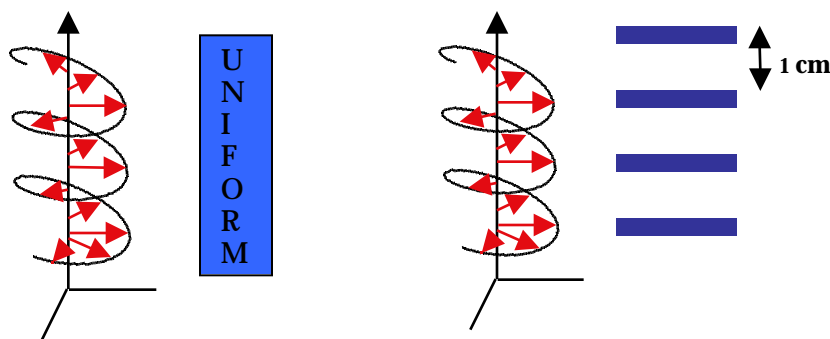
(G_y) is turned on for a brief period prior to sampling the MR signal. The MR signal is sampled after the gradient is turned off, so during the sampling there is no spread of frequencies due to the gradient; all the spins in the head are precessing at $\omega = \gamma B_0$. During the gradient pulse itself, spins at different y locations precess at different frequencies and over the time period Δt a spin at location y will gain a phase shift of $\phi = \gamma G_y y \Delta t$ compared to the reference spins at $y=0$. The relative phase shift is linear with y . If you represent the magnetization vectors in space they will form a helix along the y axis. The larger the gradient G_y or the longer it is left on (Δt), the tighter the helix of magnetization will be wound. After the gradient is turned off, the helix remains since the spins return to an identical precession frequency.

Fig. 1.6



Consider the effect of winding a helix of magnetization by briefly turning on a gradient pulse. At first glance it appears that if the object (a bottle of water for example) extends over more than one cycle of the helix, then there would be complete cancellation of the magnetization vectors and no observed signal. For every voxel in a position such that the magnetization had a phase of ϕ , one can find a voxel at a different location with a phase of $-\phi$. This is in fact the case for a uniform sample. It becomes more interesting if the sample has some well-defined spatial periodicity. If the gradient strength and timing are chosen so that the helix has the same periodicity as the sample (1cm is the example below), then there is no cancellation of the magnetization vectors and the MR signal is as large as it would be if no gradient pulse were applied.

Fig. 1.7



Uniform sample produces nosignal Periodic sample produces full signal

Another way of looking at it is that applying a gradient pulse before measurement provides a measurement of a single spatial frequency component. In the case above, the gradient amplitude and duration are set to select the signal from anatomy with a 1cm periodicity. With a single time-point measurement, we have determined the amplitude and phase which the 1cm^{-1} spatial frequency of the object contributes to the whole. If you desire an image with 1mm spatial

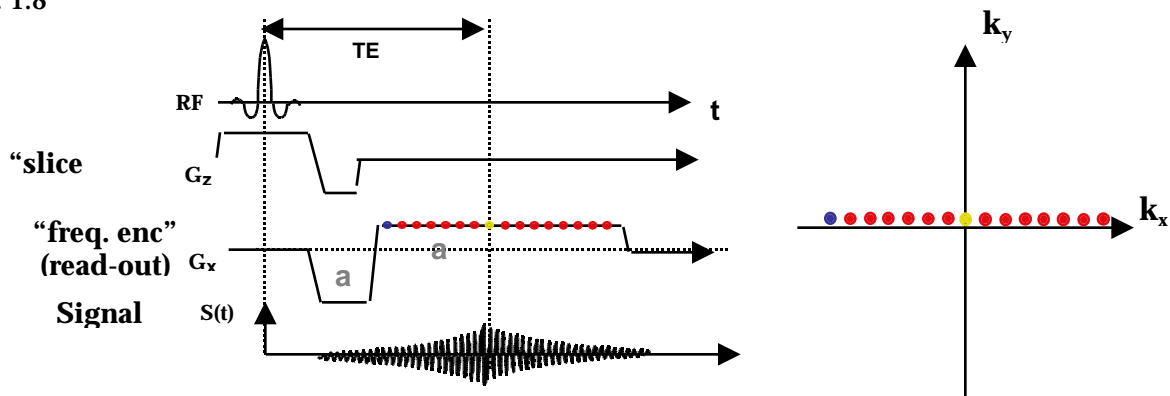
resolution, then you must acquire spatial frequencies up to 1mm^{-1} . Clearly with enough measurements of different spatial frequency components in the 2 orthogonal directions, we could reconstruct the object with a Fourier transform from spatial frequency space to object space. Blipping on a gradient to wind a helix and then sample the MR signal for a measure of a spatial frequency component is referred to as phase encoding.

Spin warp imaging (the bread and butter MR encoding method)

Reconsidering frequency encoding it is clear that a similar helix is wound, the only difference is that you continuously sample the spatial frequencies as the helix gets tighter and tighter. There is no real need to turn the gradient off; its more efficient if you leave it on. Thus in one readout period we typically sample 256 points all with different helicities and thus 256 spatial frequency components of the object in the readout direction. In conventional MR imaging, one readout measurement (of 256 or 512 kspace samples) is taken per excitation. Thus in a single excitation, the spatial frequencies of the readout direction are fully sampled. Again, if you want an image with 1mm spatial resolution, you must make sure $G_x t$, is large enough so that the maximum spatial frequency sampled is at least 1mm^{-1} . Of course there are two ways to achieve this, G_x can be big or the sampling t can extend for a long time. As we will discuss, there are reasons not to sample too long.

Gradient echo In practice frequency encoding is sufficiently efficient that it is useful to acquire some redundant information in by sampling both a negatively wound helix and a positively wound helix. The negative sense is obtained by simply applying a negative gradient for some period of time prior to sampling. So a large negative gradient winds a helix of the negative sense, then reverse the direction of the gradient and start sampling the signal. The initial samples still have the large negatively wound helix which the positive gradient unwinds over time. When the helix is completely unwound you are sampling the zero spatial frequency component of the sample. For a large, uniform phantom, this is the only point in kspace which has significant signal. Since MR physicists spend so much time imaging bottles of water (and since the head is to first approximation like a bottle of water), the time when the $k_x = 0$ point is sampled is given the special name of the “TE” (time to echo) of the acquisition. In most objects the signal is quite small when sampling the high spatial frequencies, builds up to a maximum at $t = \text{TE}$, and then gets smaller again for the opposite signed high spatial frequencies. Thus this basic experiment is referred to as a “gradient echo”. Note that the gradient echo occurs when the area ($G_x t$) of the positive lobe equals the area of the negative “prewind” lobe. The experiment (pulse sequence) for frequency encoding can thus be diagrammed as follows:

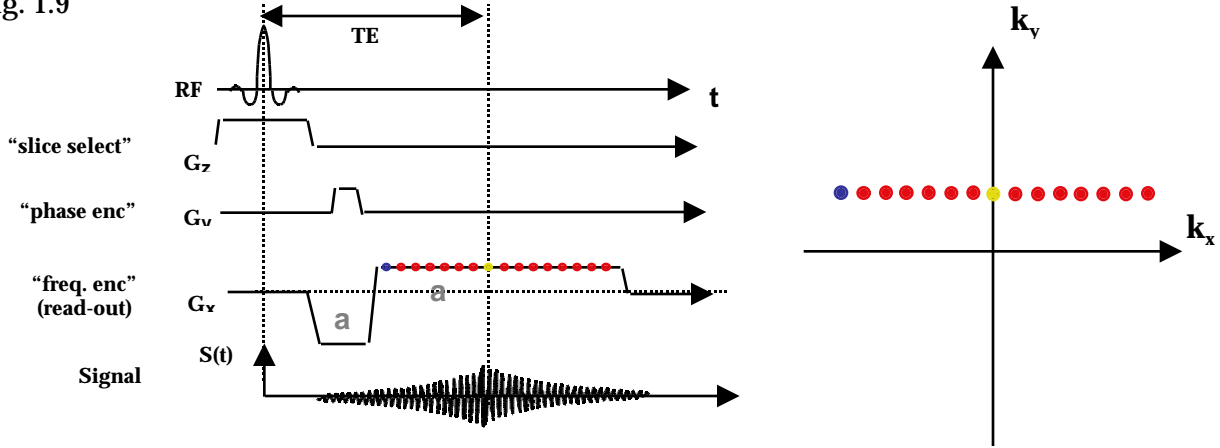
Fig. 1.8



All that is left to do is add in encoding for the y direction. To reconstruct an image we need to sample all combinations of the objects x and y spatial frequencies. The simplest (and commonest) way to do this is simply repeat the above frequency encoding experiment once for every offset in k_y . This means winding a helix in the y direction and then performing the readout procedure of winding and unwinding the x helix. The y helix can be quickly wound and then left in place for the duration of the readout procedure by turning on a brief G_y gradient

after excitation but before initiating the readout. Thus we use a different phase encode gradient before each readout experiment. If we desire a 128 x 256 image matrix, we would typically perform 128 excitations each with a different area of the phase encode blip gradient of area $G_y \cdot \tau$. Since it is inconvenient to change the timing from excitation to excitation and only the area of the blip matters, typically the amplitude of the phase encode gradient is stepped from negative to positive values. Thus the diagram for one excitation looks like:

Fig. 1.9



The NMR Imaging Equation (mathematical picture of what happens to the spins).

The goal of the 2D MR imaging experiment is to determine the distribution of the proton spins in the plane of the excited slice. The desired spin density function which we hope to display as a grayscale image is defined to be $\rho(x,y)$. Following excitation, the magnetization is precessing at its characteristic frequency $\omega_0 = \gamma B_0$. We then proceed to encode the x and y directions as in Figure 1.9 above.

Frequency + phase encoding To encode the x direction we will use frequency encoding, we apply an x gradient and then record the signal as a function of time, $S(t)$ while the gradient is on. Thus we are recording the frequency and phase evolutions that occur as a function of x during the presence of a constant “readout” gradient field. The phase encode gradient consists of a y gradient turned on for a brief period of time τ .

The MR signal comes from the RF detector which surrounds the entire head. Thus the detected signal is just the summation of the signals from all the spins within the head. Thanks to the gradients of Fig. 1.9, the phase of the signal of a given spin depends, on its location. If the signal is sampled at time t after turning the x gradient on, the phase induced on the spins at location x by the readout gradient alone will be $\phi(t) = \omega_0 t + \gamma G_x x t$. The phase induced on spins at location y by the phase encode blip alone will be $\phi(t) = \omega_0 t + \gamma G_y y \tau$. Thus the total phase shift for a voxel at location (x,y) is:

$$\phi(t) = \omega_0 t + \gamma G_x x t + \gamma G_y y \tau$$

Since the signal emitted by a small voxel at location x,y is proportional to the number of spins at that location, and it will have the phase given above, the signal $S(x,y,t)$ from a voxel at (x,y) is proportional to $\rho(x,y) \exp(i \omega_0 t + i \gamma G_x x t)$. The RF coil equally sums contributions from all locations, so the signal detected by the coil is the sum of this over the head:

$$S(t) = \int_{object} \rho(x,y) e^{i\omega_0 t + i\gamma G_x x t + i\gamma G_y y \tau} dx dy$$

The phase factor $\exp(i \omega_0 t)$ is a simple modulation factor representing the Larmor precession of the spins. This carrier frequency can be moved outside the integral and is, in fact, simply thrown away in the detection hardware and we are left with:

$$S(t) = \int_{\text{object}} \rho(x, y) e^{i\gamma G_x t + i\gamma G_y y \tau} dx dy$$

Since the integral is over only the spatial dimensions, it is useful to define $k_x = (-\gamma G_x t)$ and $k_y = (-\gamma G_y \tau)$ so that:

$$S(t) = \int_{\text{object}} \rho(x, y) e^{-ik_x x - ik_y y} dx dy$$

If you consider x and k_x as the Fourier conjugate variables (instead of x and t as before) this of course looks a lot like a 2D Fourier integral. Therefore, to solve for $\rho(x, y)$ all we need to do is take an inverse FT of the detected signal.

$$\rho(x, y) = FT^{-1} [S(k_x, k_y)] = \int_{\text{k-space}} S(k_x, k_y) e^{ik_x x + ik_y y} dk_x dk_y$$

The innocuous change of variables to k is so useful that we have already worked out an entire physical intuition about “ k ”. The value of k_{xy} directly determines the tightness of the helix of magnetization wound and thus to the spatial periodicity of the sample that contributes to that time-point of the signal. This is consistent with the standard interpretation of the FT where k refers to a spatial frequency. Similar to referring to the time signal as “time domain” and the FT of the time signal as “frequency domain” (or spectrum), we will refer to the normal spatial dimensions spanned by (x, y) as “real space” (or image space) and the FT conjugate space spanned by (k_x, k_y) as “ k -space” or “spatial frequency space”. The units of k are cm^{-1} . Note also that while we are sampling the signal at different time-points (during the readout gradient) the change of variables ($k = -\gamma G t$) forces us to think of each sample as being a sample at a different k point in k -space; exactly the same conclusion that our intuitive picture with the helix lead us to. So from now on we will think of sampling the signal not as a function of time, but as a function of k (the only difference between k and t is the scale factor γG). In MRI we measure the amplitude and phases of the points in the k -space matrix and then calculate the image with the FT.

In the phase encode direction where $k_y = -\gamma G_y \tau$, the pulse sequence of Figure 1.9 shows that we only get one measurement of the signal at a given y spatial frequency per excitation. Therefore, in conventional imaging we must fill in the k -space matrix $S(k_x, k_y)$ one row at a time. A typical image matrix is 128 points in the phase encode direction by 256 in the readout direction. In the timing diagram of Fig. XX, the readout length is typically $\sim 5\text{ms}$ long. For some types of image contrast we can repeat the excitation as fast as every 10ms. In this case the a single image is encoded in 1.28 seconds. For many applications it is desirable to wait $\sim 1\text{ s}$ between excitations. In this case the image takes 128 seconds to encode.

Some k -space facts.

Each point in k -space that we sample is represented by a complex number (magnitude and phase of the signal for that sample point.) If we want a 256×256 matrix image with 1mm spatial resolution we typically sample a 256×256 k -space matrix of k values from $k_{\min} = -1/1\text{mm}$ to $k_{\max} = 1/1\text{mm}$. Thus the resolution sets the maximum k value to be sampled. As property of the FT is that if the object is real (good assumption!) then half the information in the k -space matrix is redundant; it is related by the complex conjugate: $S(k) = S(-k)^*$. So, if we desire we can only sample half of the full k -space matrix.

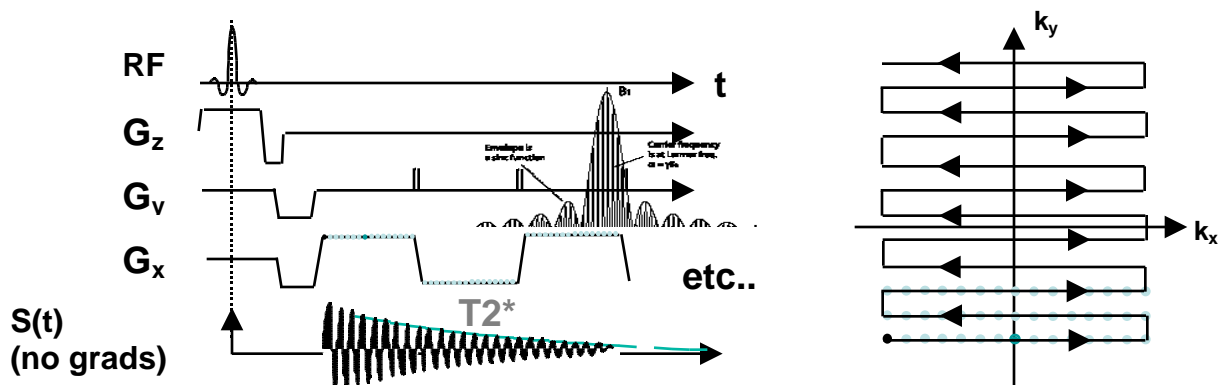
The resolution of the image is the single most important image parameter since that tells us the maximum value of k that must be sampled. Since $k = -\gamma G t$, the resolution determines the maximum gradient that must be used and the maximum amount of time that it must be left on. These are two very important practical considerations. The image matrix is also important since it tells us the number of samples we must take. The image Field of View (FOV) is trivially related to the two: $\text{FOV} = \text{resolution} \times \text{matrix}$. In k -space, the Δk between the samples is $1/\text{FOV}$.

Echoplanar Imaging

In the conventional spin warp imaging sequence of Fig. 1.9 above, a single line of points in k space is collected with each excitation. This has the advantage that the phase of the signal is reset prior to collecting each line of k space by the excitation process (which always starts with the equilibrium magnetization pointing along the z axis). This is of enormous benefit since there are several practical problems which cause us to lose control over the phase of the signal. Of course, phase errors tend to build up over time so periodically “resetting” the phase can be a big help. The principle drawback of conventional imaging is one of speed. Even if the excitations are placed very close together (say 100ms) the total time to encode 64 lines of k space is over half a second. At this rate, virtually nothing has a chance to return to equilibrium. And, while this sounds pretty fast, it would take us almost 20 seconds to cover the head with 30 slices. This is longer than the hemodynamic changes accompanying neuronal activation. Even worse, any motion that occurs during the ~ 1 s it takes to encode an individual images will cause artifacts that are bigger than the activation related changes we are seeking.

By simply repeating the gradient echo part of the conventional spin warp sequence, and thus not waiting for the next excitation we can considerably speed things up. The result is the technique named “echoplanar imaging” (EPI) diagramed in Fig. 1.10. EPI requires state of the art imaging hardware since the sequence requires the sign of the readout gradient to be quickly changed. The EPI sequence can be played out in 40ms. Thus we can get a full 64×64 image of the head (with 3mm resolution) in well under a tenth of a second. Imaging the entire head (with 30 slices) then takes less than 3 seconds, just less than the hemodynamic response time. Finally, it is truly a “snapshot” image since no motion occurs during the 40ms of encoding.

Fig. 1.10



References:

Waldo, Hinshaw and Lent, "an introduction to NMR imaging from the block equation to the imaging equation." Proceedings of the IEEE Vol 71 no. 3, March 1983, p.338-350.

Pauly, LeRoux Nishimura and Macovski, "Parameter relations for the Shinnar-LeRoux selective excitation pulse design algorithm. IEEE Transactions on Med. Imaging. Vo 10. No. 1, March 1991, p 53-65.

Kimmich R. "NMR Tomography Diffusometry Relaxometry", Springer-Verlag, Berlin, 1997.

Haacke EM, Brown RW, Thompson MR, Venkatesan R, "Magnetic Resonance Imaging, Physical Principles and Sequence Design." Wiley-liss, New York, 1999.