Magnetic Resonance Spectroscopy Instrumentation

David Houlť

David Houlť Consulting Ltd., Winnipeg, Manitoba, Canada

An overview is given of the instrumentation needed for in vivo magnetic resonance spectroscopy. Starting with the main magnetic field, its strength, stability, and uniformity are discussed and the possibilities of drift and external disturbance noted. Problems with field homogeneity specific to in vivo operation are covered and the problem of collecting data from a region remote from the origin is analyzed. Next, shim and gradients coils are described and the consequences of rapid change of gradients upon field stability and environment examined. The development of shielded gradients and the advantages of constant-current power supplies are also explained. Turning to the radio-frequency instrumentation, power amplifiers are briefly described together with the use of PIN diodes for power steering and protection. Attention is then paid to the role of the pre-amplifier in preserving the signal-to-noise ratio and the concepts of noise figure and current blocking explained. The use of multiple receiving coils and pre-amplifiers is examined and attention briefly drawn to problems of electric field balance at ultra-high frequencies. Finally, the roles of the receiver and analogue-to-digital converter are dissected, including the trade-off between resolution and sampling frequency and the need for frequency locking when generating the transmitter pulse.

Keywords: instrumentation, magnet, shims, gradient coil, RF power, PIN diode, pre-amplifier, noise figure, mixer, ADC, undersampling, DAC

How to cite this article:
eMagRes, 2015, Vol 4: 475–488. DOI 10.1002/9780470034590.emrstm1450

Introduction

Instrumentation for performing biological nuclear magnetic resonance (NMR) spectroscopy (MRS) involves advanced technologies in the disciplines of cryogenics, electromagnetics, physics, mechanical engineering, radio-frequency (RF) electronics, digital interfacing, and computing. MR has challenged the best practitioners in each discipline and continues to do so in a way that is surprising for a technique that arose in the mid-twentieth century. In consequence, any author attempting to give an overview of that instrumentation can only give a hint as to the complexities that underlie much of it and so further reading material may be found at the end of the chapter. The main components of a biological MR system are shown in Figure 1 and we briefly examine their roles and how they determine the efficacy of almost any experiment. A basic knowledge of NMR is assumed (see ***The Basics***). Other chapters contain additional details where pertinent. There may occasionally be electrical engineering terms with which some readers are unfamiliar. Fortunately, engineers have thoroughly embraced the Internet, and definitions should easily be found.

The Magnetic Field

Field Characteristics

Stability. Since the earliest days of NMR, there have been three overarching concerns regarding the main magnetic field B_0 – its strength, stability, and uniformity. Even the most naïve approach to signal strength analysis indicates that the bigger the field strength is, the bigger will be the signal, so the urge to build stronger and better magnets has ever been present, for in the grand scheme of spectroscopic techniques, NMR is very insensitive. In 1964, however, the introduction of the superconducting magnet by the Varian company altered forever the practice of magnetic resonance, for this innovation at a stroke doubled the available field strength to 4.7 T. The magnet was made with many turns of copper wire containing strands of a niobium-tin alloy that sat in a cryostat or Dewar vessel holding liquid helium (4.2 K) and liquid nitrogen (77 K), and that accommodated 5 mm sample tubes. While it would take many more years for cryotechnology to deliver a magnet suitable for humans rather than test tubes, this advance was the precursor of nearly all future NMR systems. Once the magnet was energized, the dependence on a power supply was removed and this in turn removed the main source of field fluctuations, though they can still exist as a consequence of building vibration and rapid changes of field gradients.

It was soon discovered, however, that the fields of superconducting magnets were susceptible to changes of cryogen levels, the movement of nearby iron objects – elevators, buses, trains, etc. – and that imperfect superconductivity (usually in wire splices) also caused the field slowly to decay. Thus, a device that kept the field and Larmor frequency in unison over longer periods of time (seconds to days) was needed, and this was the 'field-frequency' lock.[1] Typically, the Larmor frequency of a secondary nucleus (usually deuterium, ^{2}H) was monitored by NMR, and if it departed from resonance, the field was corrected

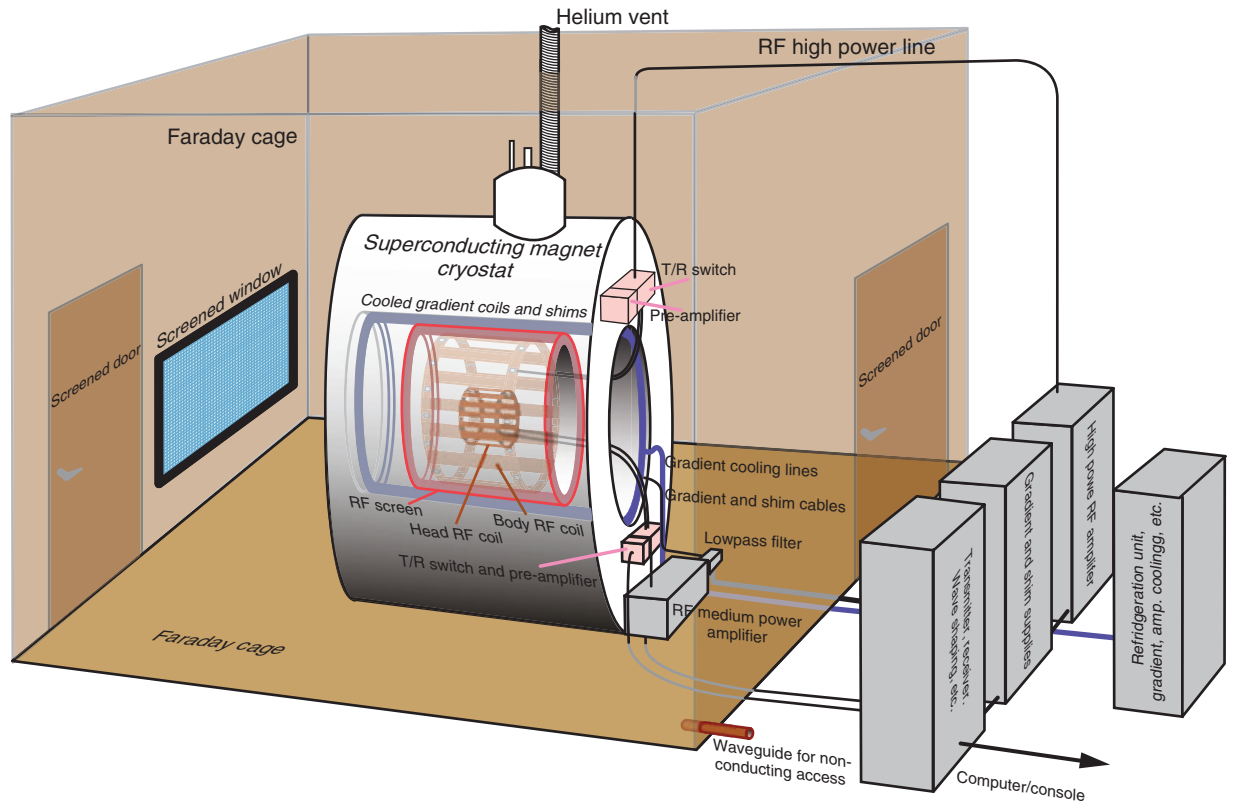

Figure 1. A conceptual drawing of a system for performing MR spectroscopy in animals and humans. The exact implementation naturally depends on the manufacturer and the work to be performed: some components in the figure may be redundant and/or additional components may be needed

accordingly. This process is not quite as simple as it sounds, for it also implies that the frequency source used for the deuterium NMR experiment must precisely track that used for the main experiment: the two sources must both be derived from a common, high-stability master oscillator. It too can drift in frequency, so a field-frequency lock actually ensures that the field and frequencies drift in tandem. (A typical fractional stability of a crystal oscillator in a temperature-controlled oven is of the order of 10^{-9} per hour.) Note that while the original implementation was to restore the field strength, an alternative is to alter the frequencies. However, excessive frequency change ($\gtrsim$0.5 MHz) demands a change of 'probe' tuning (a probe is a transmitting or receiving coil with capacitors attached – see ***Probe Design and Construction***).

If the main NMR signal has a sufficiently large spectral component (e.g. water), a variant on the field-frequency lock concept is occasional monitoring of the signal's frequency, followed by correction of field or frequency as necessary. Of course, if it is certain during the course of an experiment (usually a short one of a few minutes) that any field/frequency drift is negligible, locking/tracking can be dispensed with. In this regard, an advance in magnet technology has helped considerably.

The fringe fields surrounding large magnets are at best a nuisance and at worst a hazard, and so manufacturers have contrived to reduce these fields greatly, a practice commonly referred to as 'self-shielding'. However, if there is a magnetic disturbance at some distant point P, it can be shown that the resulting change in the magnet *current* is proportional to the fringe field at P. If this fringe field is small, then the effect on the magnet's current is small and the corresponding change in the field at the magnet's center is also small. Of course, the *total* change in field at the magnet center is the sum of the change due to the current and the field due directly to the disturbance itself, and after some early mistakes in this regard, magnets are generally designed so that the two effects are in opposition. Unfortunately, stability requirements are often greater for spectroscopy than for imaging, so when performing in vivo spectroscopic measurements on an imaging system, field stability should not be taken entirely for granted. If it is a problem, working at night sometimes helps, as external disturbances and building vibration tend to be less (see ***Probe Design and Construction***).

Homogeneity. Since the realization in 1951 that chemical spectra could be observed, a major concern has always been field homogeneity. A spectroscopic line originating from a chemical compound within a volume of interest ΔV has a natural linewidth δf determined by its molecular environment – typically in the range 3–300 Hz for biological samples. However, if the variation in B_0 field strength over ΔV is such that the variation in Larmor frequency is comparable to or greater than δf, the aggregate spectral line is broadened and its integral, a measure of the amount of compound, has reduced accuracy. Worse, adjacent spectral lines may not be resolved.

The implication in a 4.7 T field is that a magnetic field inhomogeneity (error) of better than ~0.1 ppm is needed for 10 Hz resolution. (The exact specification depends on how the field error over the sample volume is described: peak-to-peak, root mean square, etc. Beware of the hiding of mediocre performance behind statistics.) Traditionally, the overarching sources of inhomogeneity in the magnetic field have been unavoidable errors in magnet construction coupled with structural distortions owing to gravity and the magnetic forces between windings. Before the advent of superconducting magnets, initial attempts to correct the problems involved the use of small pieces of metal (shims) behind and on electromagnet pole faces, and so the term 'shimming the magnet' was born as a descriptor for improving the field homogeneity.

In 1958, Marcel Golay proposed a better method that was mathematically based. First, magnetic fields are traditionally assigned to the z direction in a Cartesian frame and the divergence of field associated with even a very large inhomogeneity (e.g., one part per thousand) produces only tiny fields B_x and B_y in the x and y directions. Now the MR system responds to the *magnitude* B_0 of the field, so

$$B_0 = \sqrt{B_x^2 + B_y^2 + B_z^2} = B_z\left(1 + \frac{B_x^2}{2B_z^2} + \frac{B_y^2}{2B_z^2} + \cdots\right) \cong B_z \tag{1}$$

and for all practical purposes, B_0 and B_z are synonymous. Second, the magnetic field in the air inside an empty magnet bore can be shown to obey Laplace's equation

$$\nabla^2 B_z \equiv \nabla^2 B_0 = 0 \tag{2}$$

Golay realized that the solutions of this equation formed an orthogonal basis for analyzing inhomogeneity: the field could be expressed as a sum of solutions. Given the circular symmetry of both electromagnet pole faces and superconducting magnet bores, the use of a cylindrical coordinate system would seem to be an obvious choice; however, a drawback is that the boundaries then determine the solutions – the latter change with sample size. Accordingly, Golay worked in spherical polar coordinates (r, θ, φ) and proposed that the field be analyzed in spherical harmonics U_{nm} and V_{nm}, namely

$$\begin{aligned} B_0 &= \sum_{n=0}^{\infty}\sum_{m=0}^{n} F_{nm}U_{nm} + G_{nm}V_{nm}; \\ U_{nm} &= r^n P_{nm}(\cos\theta)\cos m\varphi \\ V_{nm} &= r^n P_{nm}(\cos\theta)\sin m\varphi \end{aligned} \tag{3}$$

where F_{nm} and G_{nm} are constants, integer n is the order of the field, m is the degree ($m \le n$), and the P_{nm} are the associated Legendre polynomials in $\cos\theta$. In practice, the order n typically needs to run to no more than 4 or 5 to ensure an excellent description of the field. The form of the solution is unchanged with sample size, but the spherical harmonics are only orthogonal for a spherical sample *centered on the origin*. This methodology at first sight appears formidable; however, in practice, it is not, as n need not be large and the associated Legendre polynomials are then relatively simple. The harmonics are sometimes expressed in Cartesian coordinates (for example, $U_{22} = 3[x^2 - y^2]$), but this notation obscures the symmetries inherent in the spherical description.

Golay continued by proposing the winding of 'shim coils', each of which would produce a relatively pure spherically harmonic field of order n and degree m, up to some limiting value n_{max}. By appropriate adjustment of the currents in the coils, the field of equation (3) could then be 'shimmed' by cancelling the various spherical harmonics comprising the inhomogeneity. (As $P_{00} = 1$, the 'zeroth-order' field U_{00} is the main homogeneous field.) Golay coils originally comprised arcs and legs of current having specific coordinates but today, distributed designs (c.f. Figure 3) that aim to minimize power dissipation are also used. Because dissipation and the effects of winding errors increase dramatically with increasing order, it is rare to find coils with order $n > 4$.

To find the correct coil currents, there are broadly two methods: (i) map the inhomogeneous field by imaging methods or with the aid of a movable tiny NMR sample, analyze the map in spherical harmonics and knowing the characteristics of the shim coils, apply the appropriate currents; (ii) use a metric of the field homogeneity, such as the signal from a free induction decay or the height of a spectral line from a bulk sample, and maximize the metric by systematic adjustment of the shim coil currents. The latter method requires skill and experience if applied manually (as well as a signal with good sensitivity, which with biological samples implies ^{1}H), as it is possible to be trapped in local maxima that do not represent the best homogeneity attainable. It is therefore common nowadays to delegate such shimming to the computer, which can employ a global maximization algorithm (Nelder–Mead, simulated annealing, etc.) to reduce (but not eliminate) the risk of such entrapment.

Another method of shimming involves the use of optimally placed pieces of steel on the magnet bore wall or in the cryostat. While this so-called 'passive shimming' technique is quite potent, it is static and difficult to change. Golay's approach is therefore preferred for fine adjustable shimming, as it is flexible and can accommodate minor changes in the homogeneity, no matter the cause. It has, however, two serious drawbacks when we consider in vivo localized spectroscopy: (i) the spherical harmonics are orthogonal only over a spherical region centered on the origin, but the volume of interest in a person or animal is typically not spherical and not at the origin; (ii) Laplace's equation (2) is invalid if the magnetic susceptibility χ of the sample is dependent on spatial position $\mathbf{r}$ and thereby causes a significant spatial variation of field strength in comparison to the linewidth.

Consider first the question of position. Let the center (or origin) of the volume of interest be O′, some distance removed from the shim set's origin O. Neglecting variations in χ, the magnetic field in the vicinity of O′ can always be analyzed in spherical harmonics, but the further O′ is from O, the more the coefficients F'_{nm} and G'_{nm} will differ from F_{nm} and G_{nm}. Worse, the field from a shim coil, a relatively pure harmonic when measured about origin O, will in general be contaminated with all harmonics of orders and degrees less than n and m respectively when analyzed about origin O′: the further O′ is from O, the greater will be the contamination. The resulting

lack of orthogonality makes it virtually certain that local, rather than global, maxima will be attained with manual shimming; the risk thereof with algorithmic shimming is also increased. A simple one-dimensional example helps in clarifying this statement, as shown in Figure 2. Shimming manually with first the U_{10} (z) shim, the inhomogeneity is reduced over the region of interest to a minimum of 2.25 ppm, as shown in Figure 2b, blue line. (Note the large mean field change in going from Figure 2a to b, a prime warning sign of orthogonality problems.) Next, we turn to the U_{20} (z^2) shim, but it cannot improve the inhomogeneity, as is also true of the U_{30} (z^3) shim. We have been ensnared by a local maximum because in the region about $z = 3$, both z^2 and z^3 shims produce respectively dominant 9 z and 27 z variations. (The correct shim was actually $-1.0\ z^3$.) For a number of years, it was therefore conventional wisdom that one should only use the first-order shims for in vivo spectroscopy. However, if the coordinates of the center of the region of interest are known and the shims are computer controlled, the computer can reorthogonalize the shim system: the user adjusts, for example, z^3, but the computer is actually adjusting z and z^2 as well, for $(z-3)^3 = -27 + 27z - 9z^2 + z^3$. Even with algorithmic shimming, reorthogonalisation speeds the shimming process. Notwithstanding, it is always best to position the volume of interest as close to the origin as possible; if nothing else to ensure that shims do not encounter their limits of current.

When the cause of inhomogeneity is the sample itself, we tend to be limited in what we may accomplish. The relevant differential equation is of the form [c.f. equation (2)]

$$(1 - \chi[\mathbf{r}])\nabla^2\mathbf{B}[\mathbf{r}] + \mathbf{O}[\mathbf{r}] = 0 \qquad (4)$$

where the remainder **O** is complicated. This equation is generally insoluble. A constant **B** field of amplitude B_0 reduces the term in **B** in equation (4) to zero, but it may be shown that the remainder is then given by

$$\mathbf{O}[\mathbf{r}] = B_0\left[\frac{\partial^2\chi}{\partial x\partial z}, \frac{\partial^2\chi}{\partial y\partial z}, -\frac{\partial^2\chi}{\partial x^2} - \frac{\partial^2\chi}{\partial y^2}\right]$$

As this vector is not zero, we conclude that equation (4) cannot be satisfied by a homogeneous field and therefore, in general, a homogeneous field cannot be obtained. The volume susceptibility of organic material is usually of the order of -10^{-5}, whereas that of air is close to zero. It follows that when a region of interest includes an air-filled cavity, it can suffer considerable inhomogeneity. Even if the tissue is itself homogeneous but is *close* to an air-filled cavity (e.g., the sinuses), the field in the tissue may comprise orders higher than those for which shims are available. In conclusion, while magnet design is incomparably better than in the early days of NMR, in vivo spectroscopy presents challenges for which progress is difficult.

Gradients

The use of transient magnetic field gradients, which temporarily cause field B_0 to vary linearly with distance x, y, or z, is fundamental to both imaging and localized spectroscopy and at first sight, their production appears to be simple: they

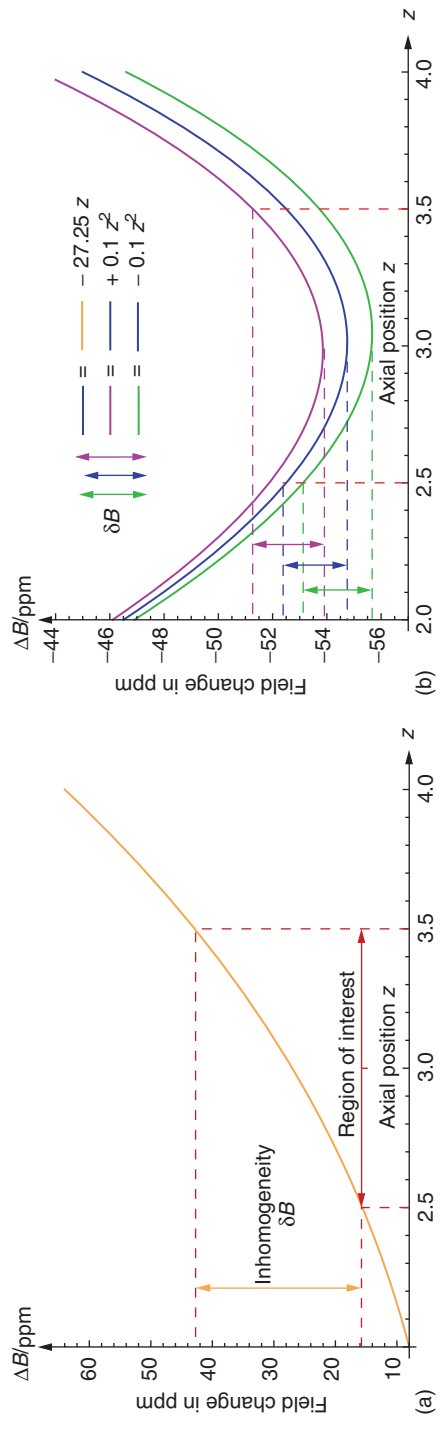

Figure 2. One-dimensional example of entrapment in a local maximum. In (a) a field variation over a region of interest remote from the origin is shown. In (b) the field variation obtained by optimally adjusting the z shim is shown (blue line). Counterintuitively, adjustment of the z^2 shim (green and magenta lines) degrades the homogeneity because the region of interest is not centered on the origin

merely require very powerful first-order Golay coils. Referring to equation (3), as $P_{10}(\cos\theta) = \cos\theta$ and $P_{11}(\cos\theta) = \sin\theta$, $U_{10} = z$, $U_{11} = x$ and $V_{11} = y$. However, numerous problems arise because the gradients are both powerful and transient. We require them for times very much less than T_2 and so they must turn on and off quickly; when present, they must be stable and accurately known; once off, the main field should have no memory of their former presence. These problems can roughly be divided into two categories: the effects of the gradients on the magnet and shim coils and *vice versa* and the design of suitable power supplies.

Any simple gradient coil has mutual inductance M with conductors in close proximity, and conductors abound in the magnet windings, the cryostat, shields, the shim coils, the other gradients, and the radio-frequency transmitting and receiving coils. When the current I through a gradient coil changes, an electromotive force $E = M\ \mathrm{d}I/\mathrm{d}t$ is induced in a nearby conductor and unless steps are taken to stop it, current flows: this immediately changes the main magnetic field. The nearest conductors such as the cryostat are not superconducting, so the so-called *eddy currents* eventually die away (typically exponentially), the times taken being dependent on the conductors' positions, resistivities, and geometry.

Whenever two conductors have currents flowing, there is a mechanical force between them (in general, both linear and torque) and if one of the currents changes, so does the force. Rapid change of current thus creates an impulse that can cause vibration of the magnet, of its cryostat (a bell carrying eddy currents) and of the gradient coil assembly, and quite apart from the ear-splitting noise this can make, *oscillatory* currents can be induced that once again change the main magnetic field. The depressing conclusion is that following a change of gradient, the main magnetic field B_0 can vary as

$$\delta B_0 = \sum_{n=0}^{\infty}\sum_{m=0}^{n}\sum_{p=0} F_{nmp}\left(\exp\left[i\omega_{nmp} - \frac{1}{\tau_{nmp}}\right]t\right)U_{nm} + G_{nmp}\left(\exp\left[i\omega_{nmp} - \frac{1}{\tau_{nmp}}\right]t\right)V_{nm} \quad (5)$$

where F_{nmp} and G_{nmp} are the amplitudes, the frequencies ω_{nmp} include zero, and the time constants τ_{nmp} can range from microseconds to seconds. Note that in general, the frequencies and time constants are not necessarily the same for different orders n and degrees m and that a main field offset U_{00} is included in the equation. Mathematically, this type of equation is formally known as a *complex mess*.

What tools can be brought to bear to reduce the F and G in equation (5) to acceptable sizes? Probably the most important tool is gradient shielding. Imagine a gradient coil inside an infinitely long superconducting cylinder. When the gradient is switched on, persistent eddy currents flow in the cylinder in a manner that prevents magnetic fields escaping. There is then no mutual inductance to exterior circuits. Of course, the superconducting eddy currents distort the interior gradient field, but we can redesign the coil to correct for this defect. We then have a gradient coil free of coupling to the outside world. Implementation of an actual superconducting cylinder presents major difficulties (what happens when the magnet is energized?), but we may mimic the persistent eddy currents by replacing them with current in wires of appropriate winding density as shown in Figure 3 (blue lines): where eddy currents are strong, the wires are close together; where eddy currents are weak, the wires are far apart. The scheme is imperfect as a continuous surface distribution of eddy currents cannot be exactly reproduced with discrete wires and further, the (nonconducting) cylindrical former upon which the wires are laid cannot be infinitely long. Nevertheless, the introduction of shielded gradients was an important advance and it reduced the coefficients in equation (5) substantially.

Other tools include gradient designs that have zero net torque, the use of a very thick copper magnet bore liner to shield the cryostat to very low frequencies (not quite a superconducting cylinder); the use of mechanical engineering techniques and building materials that reduce the ringing of acoustic resonances, the use of vacuum technology to contain acoustic noise and the shaping of gradient pulses to reduce high-frequency components, etc. Even with the application of all these tools, however, some of the coefficients in equation (5) may still be worrisome and we briefly examine them further below.

We must also consider the manifold interactions between the gradient coils used for imaging and localized spectroscopy, and shim coils. Fortunately, symmetry comes to our aid. Thus, the interactions between the gradient coils themselves are *nominally* zero, as are the interactions with higher degree ($m > 1$) shims. We may also eliminate the strong interactions with first-order shim coils by eliminating the shims! A small, variable direct current through the gradient coils then replaces them. There are potentially strong interactions between the z gradient and the z^3 shim ($n = 3$, $m = 0$) and between the z^2 shim and the magnet, but these should be removed by design. There then remain residual interactions and at this point, electronics can take over. The *generic* circuit diagram for both a shim and a gradient coil is the same, and it is shown in Figure 4a – only the component values differ. The coils, with currents I_{nm}, each have a self-inductance L_{nm}, a resistance r_{nm}, a parasitic capacitance C_{nm} that causes resonance at a high sonic or ultrasonic frequency, a parasitic impedance Z_{pnm} associated with external residual couplings (e.g., to the magnet), and, during gradient switching, a series emf, $E_{nm} = L_{nm}\ \mathrm{d}I_{nm}/\mathrm{d}t$ associated with self-inductance or a series emf, $E_{gnm} = M_{gnm}\ \mathrm{d}I_g/\mathrm{d}t$ associated with mutual inductance M_{gnm}, where the index g references the gradient coil whose current is being switched. Clearly, emf's E can temporarily change the desired flow of current, the change depending on the source impedance Z_{snm}. Thus, it is important to make impedance Z_{snm} as large as possible, and the tool we use is negative feedback to create a current source with very high dynamic output impedance.

The concept of a constant current source is encapsulated in Figure 4b. Obviously, to work correctly, the source power supplies must have voltages $\pm V_s$ somewhat greater than the extrema of E, and to avoid instability or oscillation, the feedback factor $\beta(\omega)$, a function of frequency, must be as large as possible over as large a bandwidth as possible but reduce to <1 at the frequency at which its phase has changed by 180° (the Bode stability criterion) – in practice, 135° is used to give

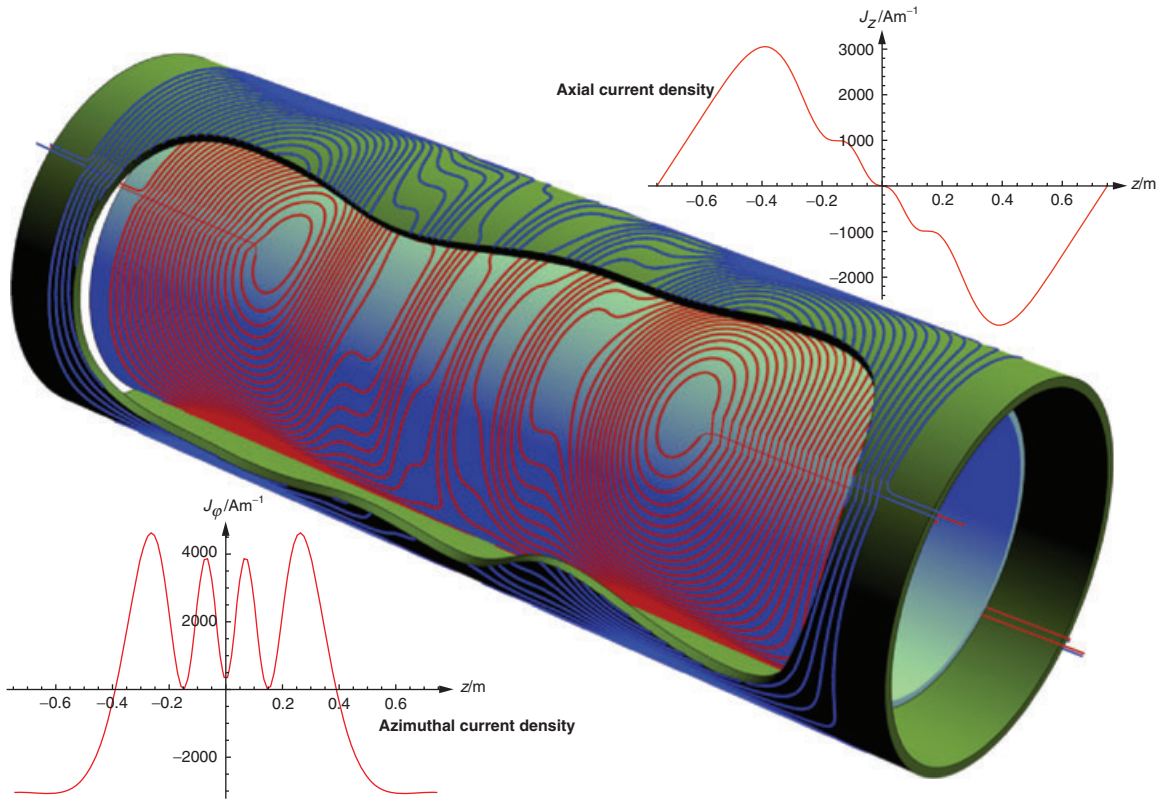

Figure 3. A shielded gradient coil assembly that uses distributed inner windings (red) to improve gradient purity and an outer, second set of windings (blue) to greatly reduce fields outside the assembly. Note that partial connection in parallel can reduce inductance. Similar distributed single-layer windings can also be used for shim coils

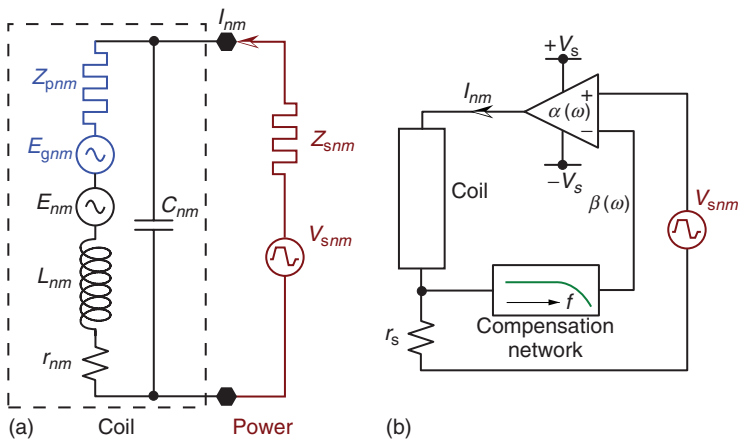

Figure 4. In (a), the circuit elements of a shim or gradient coil are shown with parasitic impedance and voltage due to exterior couplings (blue). The coil has a power supply with voltage V_{snm} and internal impedance Z_{snm} (red) but the current is not entirely independent of induced voltages E. To keep the coil current I_{nm} proportional to voltage V_{snm}, Z_{snm} must be very large, and so a constant current source is used, as shown in (b). For an amplifier with a very large gain α, the voltage V_{snm} at the positive input essentially equals that at the negative input. The latter is $I_{nm}r_s$, where r_s is a small sensing resistance, and therefore $I_{nm} \simeq V_{snm}/r_s$

a margin of safety. With increasing voltage, amplifiers rapidly increase in expense and the risk of arcing in the gradient coils magnifies, so it is important to maximize the gradient coil efficiency so that *for a constant desired gradient strength*, the inductance L is minimized. The time Δt in which a gradient current can be changed by ΔI then varies as $\Delta I\, L/V_s$. For shims, on the other hand, which carry current constantly, it may be more important to minimize resistance and thereby heat deposition in the magnet bore. (Design techniques for both cases exist.) If shim power supply voltages, determined by the size

of E_{gnm}, are excessive, it also may be necessary to introduce compensatory mutual inductance to cancel M_{gnm}. Minimization of L_{nm} has the added benefit of increasing the gradient coil self-resonant frequency $\omega_{nm} = (L_{nm}C_{nm})^{-1/2}$, which in turn allows the constant current source to function satisfactorily at higher frequencies.

With an ensemble of multiple coils and amplifiers with multiple interactions, there is always the risk of ensemble oscillation. Mercifully, this rarely occurs and it can usually be cured by a reduction of bandwidth. A huge additional benefit of negative feedback is that it ensures great linearity in the relationship between coil current I_{nm} and the drive voltage V_{snm} from the pulse programmer, provided the sensing resistor r_s in Figure 4b does not change in value as it heats. However, a linear relationship between voltage and current does not necessarily imply a time-independent linear relationship between voltage and gradient strength, as is hopefully clear by now.

When all the above tools have been used, there remain the residual field fluctuations of equation (5) and their effects on spectral line shape can sometimes be observed, so the user should not entirely forget about them. If they are important, a last resort is to calibrate the time dependence of those orders and degrees for which there are shim coils (the gradient coils themselves are included in this statement). Given that the fluctuations are proportional in amplitude to the gradient change, there then exists the possibility of cancelling them by injection of compensating currents into shims/gradients for which a dedicated micro-controller may be needed. These residuals are more important in spectroscopic applications than in imaging and can be responsible for oddly shaped spectral lines. Finally, there is one important component of equation (5) for which there is often no shim coil: U_{00}, a spatially invariant field. The latter commonly has a low-frequency, transient oscillatory component associated with acoustic ringing. If a U_{00} coil (a 'field-offset coil') is present, then it may be considered a shim coil and the above analysis pertains. If, however, it is absent (and magnet bore space is expensive), there remains the possibility of compensatory modulation of the receiver detection frequency instead; for example, by control of the signal-sampling rate. Typically, a user will only discover this piece of sophistry if an auxiliary receiver is attached to the MR instrument for some reason, in which case it can cause much head scratching!

In conclusion, it is obvious from equation (5) that residual transitory effects associated with gradient switching are generally smallest at the origin of the magnet/gradient/shim system. Thus, just as when considering homogeneity, the closer the region of interest can be to the origin, the more likely it is that the best results will be obtained.

Radio-Frequency Engineering

Core Components

We now consider a collection of RF electronic devices, grouped conceptually about the NMR sample, that form a natural ensemble (see Figure 1) with five principal tasks to perform: (i) The generation in a power amplifier of high-power (typically kW) precise RF pulses at about the Larmor frequency; (ii) The excitation of the NMR sample with that power by the creation, with a transmitting coil, of RF currents that generate a rotating magnetic field B_1; (iii) The detection of precessional magnetization in a receiving coil, predominantly by Faraday induction of a small RF voltage ξ (nV to mV); (iv) Amplification in a pre-amplifier of that voltage, with the addition of as little extra noise as possible (low 'noise figure'), so that it can be safely passed without the loss of signal-to-noise ratio to the receiver and computer; (v) Screening from external interference with a Faraday cage that also prevents the transmission of interference at the Larmor frequency to the outside world.

There are several possible permutations and combinations of the above devices: the Faraday cage or screen may be buried in the walls and doors of the magnet room or may be an integral part of the magnet; the transmitter and receiving coils may be one and the same or they may be separate; there may be multiple receiving coils in simultaneous use, each with its own pre-amplifier; there may even be multiple transmitter coils, each with its own power amplifier; excitation of more than one nuclear species at a time (e.g., ^{31}P and ^{1}H) may be possible, etc. The chosen arrangement depends entirely on the experiment and its demands (see ***Probe Design and Construction***). However, in all instances, there are fundamental design constraints that must be considered: the power amplifier must have good linearity so that the phase and amplitude variations of the B_1 transmitting field are as expected; the transmitting coil must be able to handle the applied power without arcing or burning out; RF power must not be wasted as it is expensive; if something goes wrong, it must be detected and the power amplifier quickly shut down; the transmitter power must not destroy the pre-amplifier; during signal reception, the power amplifier must be well and truly turned off so that it does not degrade the signal with extra noise; if multiple coils are used, they should not interact and degrade one another's performance; the first transistor in a pre-amplifier should be noise-matched to its receiving coil so that it adds minimal extra noise to the signal, etc. This is a large and difficult list and in the space available, we may only examine it briefly. Not surprisingly, any aspect of the above list may impinge on the others.

Power Amplifiers. The manufacture of RF power amplifiers is a highly specialized trade. In the past, it was common for huge vacuum tubes (valves) to be employed that could generate the necessary kilowatts, and there is still a place for them as they are amazingly robust and can handle considerable abuse when reflected RF power is deposited in them rather than in the load (the transmitter coil). However, they suffer from a serious drawback – they do not function in a magnetic field greater than roughly 50 mT and must therefore be kept well away from the magnet. This creates the problem of how to transport the RF power from the remote transmitter to the interior of the magnet without radiation of radio waves, and to do this a coaxial cable is used; the characteristic impedance is typically 50 Ω. Thus, power amplifiers are normally designed to deliver their maximum power into a load of 50 Ω. Note that this does *not* imply that the output impedance of the device is 50 Ω, quite the contrary. (See below for a discussion of how to employ this fact gainfully.) While such power matching is

efficient, a design that implements it is often constrained by maximum safe working voltages. Thus, matching is sacrificed for greater available output power. It should be noted that efficiency is also sacrificed when long coaxial cables are needed, as they have resistance.

Power transistors (typically metal oxide field effect transistors or MOSFETs), by contrast, *can* function in strong fields, but typically only deliver hundreds of watts. In recent years, there has been a move to place multiple MOSFET amplifiers near the magnet and combine their outputs to obtain the desired power, either in Wilkinson or quadrature hybrid couplers. This has required innovative power line design, as a standard procedure to prevent RF power escaping from an amplifier is to use ferrite-based power line filters. Ferrites, of course, saturate in a magnetic field and cease to have a large differential permeability. Even more than tubes, MOSFET power amplifiers sacrifice power matching for greater maximum power output, and they are also more likely to breakdown or burn out when confronted with loads that are not 50 Ω. There *are* design methods that help protect amplifiers against this possibility: for example, ferrite lumped-element circulators can divert reflected power to a 50 Ω load. In this instance, the ferrite needs a saturating magnetic field, which can conveniently be provided by the MR magnet.

There is a trade-off in amplifier design between power efficiency and linearity, and a common compromise is so-called 'class-AB' push–pull amplification. A class-AB amplifier normally has two power transistors, one for the positive portion of a cycle of RF signal and the other for the negative portion. The transition, or crossover, from one transistor to the other is arranged to be smooth so as to reduce crossover distortion, but the latter cannot be removed entirely, and so the amplifier generates harmonics of the Larmor frequency. A more serious generator of harmonics, however, is the two transistors' 'running out of steam' at high power as they approach saturation. This flattens the top and bottom of a cycle and introduces odd-order harmonics as well as reducing the output at the fundamental frequency. The consequence is that the graph of output voltage versus input voltage falls away from a straight line at higher values, and this nonlinearity can compromise the efficacy of shaped or selective pulses. More subtly, the phase of the RF output may also change with increasing power. While RF negative feedback techniques (e.g., Cartesian feedback[2]) exist to correct these defects, they are difficult to use and so a simpler approach is the use of predistortion: once the amplifier's characteristics are known, the microprocessor that shapes the amplifier's input waveform can compensate for nonlinearity. The main drawback to this approach is that the amplifier's characteristics change during warm-up and may also drift over a period of months. Recalibration is usually possible if poor pulse performance is suspected.

At the highest magnetic fields (7–11 T), the use of multiple transmission coils is a topic under active research at the time of writing as, in principle, sample power deposition for ^{1}H can be reduced and a more homogeneous B_1 field obtained. However, the coils tend to have considerable mutual impedance, and so driving one coil causes appreciable current to flow in its nearest neighbors, thereby corrupting the desired B_1 field. The fact that the output impedances of RF power amplifiers are not 50 Ω can help decouple coils to a limited extent (∼ −11 to −16 dB), as discussed in detail below for the corresponding case of receiver coil interaction, and Cartesian feedback can create constant current sources (c.f. gradient coil power supplies) resulting in excellent isolation (∼40 dB). However, as mentioned earlier, the latter technique is both expensive and difficult, so current approaches tend to focus on calibrating the interactions and then orthogonalizing the interaction matrix (c.f. the discussion on translated shims above).

Taming RF Power. The issues of protecting the pre-amplifier from transmitter power and noise are intimately connected in that both rely on the same philosophy – the use of PIN diodes, controlled by the passage of direct current, that act as RF switches. For primary protection purposes, these devices have almost entirely replaced ordinary high-speed diodes connected in anti-parallel ('crossed diodes'), as they do not have a threshold for operation and generate less distortion. PIN diodes (*p*ositively doped, *i*ntrinsic and *n*egatively doped semiconductors) are capable of storing charge Q_I in their intrinsic layer when a direct current I_{DC} flows. The charge can then be extracted over a short time by a reverse voltage and replenished by a forward voltage; in other words, the diode can conduct RF current. As charge Q_I is the integral of current, it follows that the higher the frequency (and hence the shorter the period), the more RF current can be drawn. If the current limit is exceeded, the diode is starved of charge, its RF resistance increases and it may burn out. Conversely, PIN diodes have a minimum recommended frequency; at much lower frequencies, they behave as ordinary diodes. As with all electronic components, the correct diode has to be chosen for the job at hand. The effective 'on' resistance is typically $\sim 0.01/I_{DC}$ Ω plus a parasitic inductance (∼1–2 nH), where I_{DC} is in the range 10–100 mA; the 'off' impedance is usually dominated by the diode's parallel capacitance of typically ∼1 pF. In advanced designs, both inductance and capacitance may be tuned out.

PIN diodes are normally operated in the 'on' mode in the presence of RF power because when off, a reverse voltage V_R large enough to prevent significant RF conduction must be applied. Voltage V_R may need to be as a large as the RF signal amplitude, depending on frequency, and switching large direct voltages rapidly is then cumbersome and costly. Figure 5 shows the genesis of series and parallel switches that isolate the RF power from the direct current source and *vice versa*. These switches do *not* use the ferrites typically found in manufacturers' literature for PIN diode switches,[3] and so are suitable for use in the fringe field of the magnet. They rely on T-section filters; when allowing RF current at the Larmor frequency to pass, they can be shown to maintain 50 Ω continuity and so are compatible with coaxial cable. As an example of use, albeit not recommended due to excessive line length, the circuits could be joined and a probe's cable also connected to the joint, as shown in the figure. The reader can doubtless think of many other configurations and modifications. With higher powers and frequencies, switches may need to be concatenated to achieve the desired protection or noise rejection, but the relevant specifications for any switch are insertion loss and isolation, both usually measured in decibel. To achieve excellent results,

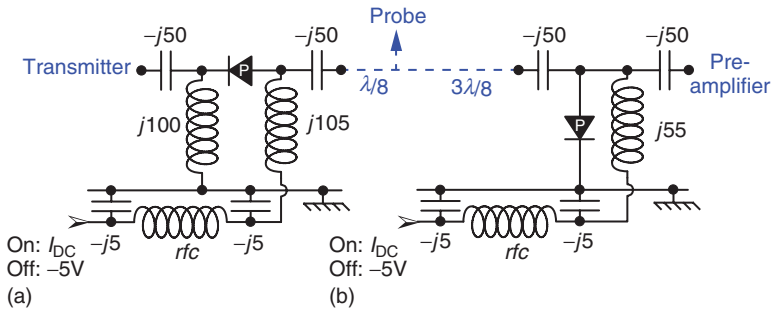

Figure 5. Elementary building blocks for PIN diode switches. When on (direct current I_{DC} flowing), circuit (a) is designed to pass high power while preserving 50 Ω continuity; circuit (b) is designed to short-circuit power to protect a device to its right, but to preserve 50 Ω continuity when off. Conversion of reactance to high (or low) impedance can be accomplished with appropriate filters or lines of the correct length, as shown in blue. With such additions, a variety of switches for the various MR permutations can be constructed

all lines, capacitors, and inductors must have the least loss possible and, in general, the fewer components the better. Note that the filters in the direct current supply lines may ring at a low frequency if the current/voltage source does not provide damping. Finally, a potential problem with the abandonment of ferrites with their broadband capability is that switches may not protect the pre-amplifier against harmonics created by the power amplifier. This possibility should always be borne in mind if pre-amplifiers burn out for no apparent reason. It should, however, have been guarded against with filtering in the power amplifier.

The Faraday Cage. A Faraday cage is a conductive closed box, usually made of copper foil or sheet, that encases at a minimum the probe and sample, and maximally, the magnet, patient table, power amplifier, and pre-amplifier, as shown in Figure 1. The underlying principle is that radio waves, electric fields, and high-frequency magnetic fields cannot pass through the copper. Any magnetic resonance system operates at radio frequencies; in other words, at frequencies that could be used if we so desired to transmit and receive radio waves. However, we emphatically do not want to do so, hence the cage. The problem with this concept is that we wish to pass cables and maybe light, perfusion lines, and other exotica through the cage, thereby potentially destroying its efficacy. When a cage ceases to function properly, as manifest by unexplained spectral phenomena such as localized noise, unexpected and variable spectral lines, etc., or complaints of interference from neighbors, it is almost certain that the integrity of the screen has been breached by the passage of something in an unapproved manner; for example, a cable. That cable then acts as an antenna outside the screen and carries power from the local FM radio transmitter inside, and *vice versa*. If such leakage is suspected, it can sometimes be detected with a portable radio inside the cage. (Beware magnetic batteries!) Absolutely no signal should be detectable at any wavelength or frequency.

To pass RF coaxial cables through the screen is usually quite simple. It is the cable's outer shield that acts as an antenna and picks up radio waves and so it is the *shield* that must not pass through a hole in the cage. Rather, the cage is rigorously connected to the shield, usually with the aid of a metal panel on which is mounted a 'bulkhead connector' – a double connector for cables that is designed for just this purpose. Once exposed, for example, at the probe, the cable's center conductor can potentially radiate interference but it is supposed to originate in a device, such as the power amplifier, that is itself shielded. It may be helpful to imagine the cage and every external device with a penetrating RF cable as having a conducting outer surface of complicated but *unbroken* topology.

Passing *unshielded* cables such as those for shims and gradients is a little more difficult. Each cable must pass through a lowpass filter that severely attenuates radio frequencies but does not affect switching times. The filter comprises π-sections whose arms are inductors and whose legs are capacitors directly connected to the screen. Each filter is usually in its own shielding box. Rarely, a defective filter may be the cause of a failure of shield integrity. For passing nonconducting lines such as fiber-optic cable through a screen, a waveguide is used. This is a copper tube, of length about 10 times its diameter D, that passes through the screen and is soldered in place with no gaps round the circumference. Radio waves of wavelength greater than $1.7D$ are rapidly attenuated and effectively cannot pass.[4] A classic error is to pass a conductor or cable through a waveguide. In this context, remember that tubes carrying saline are conductors. To pass light into a room for projection of images, a waveguide can be used, or if this is too restrictive, a window that is essentially an aquarium filled with saturated salt solution. The challenge is to connect the conducting solution to the screen without leakage and corrosion. Yet another technique is to use glass over which a fine copper mesh has been placed. The wires are, of course, connected to the screen. However, this method, which is also used for patient viewing in MRI suites, can reduce screening efficacy a little.

The Pre-amplifier. During signal reception, the receiving coil has an induced NMR signal voltage ξ, together with an induced random noise voltage caused by the Brownian motion of the sample's electrolytes. The random motion of the coil's electrons also generates a little noise, and neglecting temperature differences, a measure of the total noise N is the effective coil resistance r_c. Noise N is usually $\sim 0.1\,\text{nV}/\sqrt{\text{Hz}}$, so in a bandwidth of 1 MHz, it is $\sim 0.1\,\mu\text{V}$, a rather small voltage. Now signal and noise are ultimately passed through a receiver to an analogue-to-digital converter (ADC) so that they may be

processed in a computer, and we do not want the signal to be degraded in any way before the conversion. Degradation takes two principal forms: the addition of extra noise and distortion or signal clipping that may give anomalous results such as small spurious spectral lines and dips beneath large lines. The pre-amplifier is the first piece of active electronics encountered by ξ and N, and its job is to amplify them by a factor α so that αN is much greater than any subsequent additional receiver noise N_{rec}. If α is too big, however, a particularly large signal may overload the receiver, causing distortion. Thus, a compromise usually sets α in the range 20–30 (26–30 dB).

In the process of signal reception, the pre-amplifier itself must add minimal extra noise, and it is worth examining how this is accomplished. A simple noise-equivalent circuit for a transistor is shown in Figure 6. It comprises voltage and current noise sources, N_a and I_a, respectively, which, to a good approximation, are uncorrelated with one another and with the noise N from the receiving coil. The transistor also has an input impedance Z_a that is considered noiseless – any noise is subsumed in the noise sources. Now assume that our received signal and noise are somehow transformed in size losslessly such that their source resistance r_c, the receiving coil resistance, becomes a new resistance R_c, as shown. (How this is performed lies in the domain of receiving coil or probe design – see ***Probe Design and Construction***; suffice it to say that the aid of low-loss capacitors is vital.) The signal and noise, by conservation of energy, are increased by a factor of $\sqrt{R_c/r_c}$ to ξ' and N'. Remembering that uncorrelated random functions add quadratically, the total noise N_{tot} at the input to the transistor is given by

$$N_{tot}^2 = \beta^2\left(N^2\frac{R_c}{r_c} + N_a^2 + I_a^2R_c^2\right); \quad \beta = \frac{Z_a}{Z_a + R_c} \tag{6}$$

whereas the signal at this point is

$$\xi' = \beta\xi\sqrt{\frac{R_c}{r_c}} \tag{7}$$

Let us now maximize the signal-to-noise ratio $\Psi = \xi'/N_{tot}$ by varying the transformation. Setting the differential of Ψ with respect to R_c equal to zero, we obtain for the optimal value of R_c

$$R_{opt} = \frac{N_a}{I_a} \tag{8}$$

This is an important equation. It states that a primary condition for obtaining the best noise performance from a transistor is that the signal source must have its source resistance r_c transformed to a resistance R_{opt}. Equally importantly, note that R_{opt} has nothing to do with the input impedance Z_a – they are generally completely different in value. R_{opt} is under the control of the transistor manufacturer, whose datasheet must state either R_{opt} or E_a and I_a *at the frequency of interest* for it to be useful. Representative values of R_{opt} in the region of 100 MHz are 1000 Ω for a field effect transistor and 100 Ω for a bipolar transistor, but the datasheet or some other reliable source should always be consulted in this regard. Input impedances Z_a are typically much greater.

Noise Figure. When we look at how much extra noise the pre-amplifier has added, a measure is the ratio

$$F' = \frac{N_{tot}}{N'} = \left(\frac{N'^2 + 2N_a^2}{N'^2}\right)^{1/2} \tag{9}$$

where we have used equation (6) and substituted for I_a from equation (8). The quantity F' is known as the *noise factor* of the transistor (the square is also used), but it is more common for the noise figure in decibels, $F = 20\log_{10}F'$, to be quoted. An excellent noise figure is 0.3 dB, which translates to $F' = 1.035$. While a transistor manufacturer may quote such a figure, in practice, it is extremely difficult to maintain when switches and other circuits involving lines, inductors, and capacitors are part of the transistor's signal source. In particular, inductors with quality factors $Q = \omega L/R$ greater than 200 are difficult to construct and the introduction of their resistance R generates extra noise N_R, as given by Nyquist's famous formula

$$N_R = \sqrt{4kTR\Delta f} \tag{10}$$

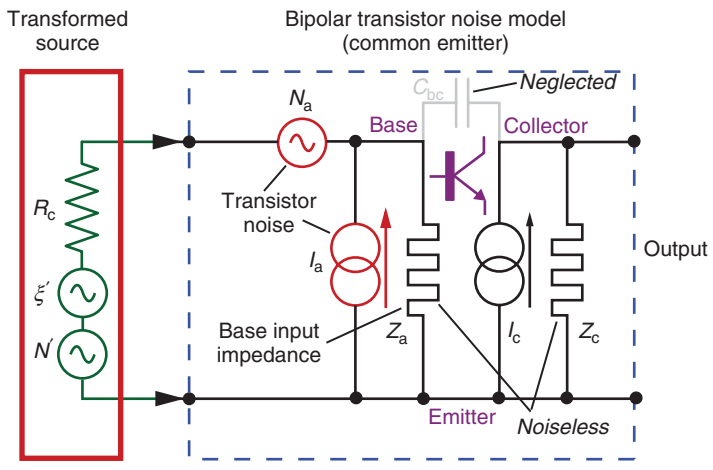

Figure 6. A simple transistor-noise model with noise matching. By appropriately transforming the source resistance, signal, and noise to new values, the noise added by the transistor may be minimized

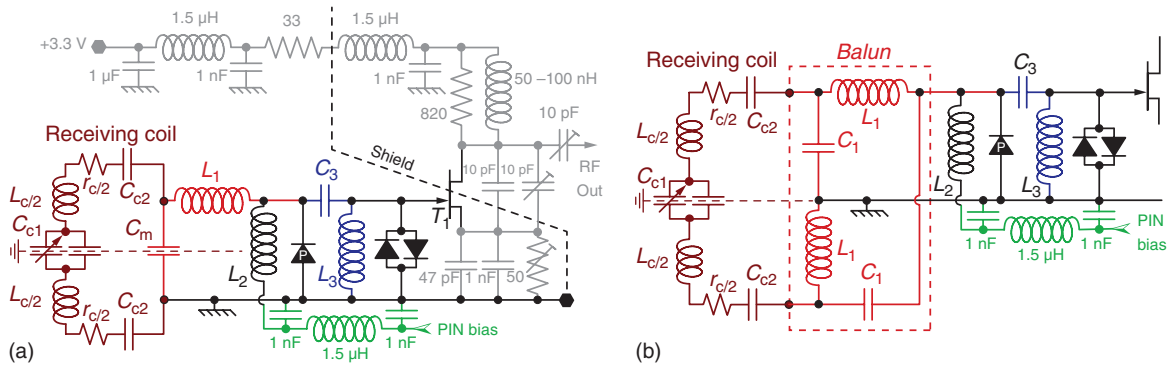

Figure 7. (a) A generic schematic of a low-noise FET pre-amplifier suitable for 100–400 MHz is shown. (b) Replacing coil-matching capacitor C_m with a balun creates an amplifier with a differential input. Component values depend on the frequency and the receiving coil's characteristics. Suitable gallium arsenide (GaAs) transistors are available from a variety of manufacturers but are only usually characterized for the low-GHz-frequency range

Here, k is Boltzmann's constant, T the resistance's absolute temperature, and Δf the bandwidth in hertz of the measuring equipment. The noise figure is usually defined for a standard source temperature of 290 K. Equation (10) also applies, of course, to r_c and R_c.

In a system that uses 50 Ω cable, it is inconvenient to have a pre-amplifier that gives its best noise performance from some other impedance R_{opt}. It is therefore common practice to hide inside a commercial pre-amplifier a filter, comprising a high Q-factor inductor and capacitors, that *at the Larmor* frequency transforms 50 Ω to R_{opt} with as little loss as possible. Crossed diodes to 'ground' may be included as part of one of the capacitances so as to give the transistor a modicum of protection from destructive voltages (see Figure 7). Alternatively, a transistor (or several transistors in parallel) having $R_{opt} = 50\,\Omega$ may be chosen. The pre-amplifier is then a broadband device. In both cases, however, just because there is an appellation '50 Ω pre-amplifier', it does not mean that the device's input impedance is 50 Ω. In general, it is not; rather it is Z_a reverse-transformed by any filter.

It is particularly easy to measure the noise figure of a '50 Ω pre-amplifier' if it is attached to an MR console that can measure root-mean-square (r.m.s.) noise. (To be accurate, it is the noise figure of the entire system that is measured, which is arguably even more useful.) All that is needed is liquid nitrogen, a short cable terminated in a *well-shielded* 50 Ω RF resistor at room temperature T and another terminated similarly in 50 Ω but at 77 K. (In general, the two resistors will not have the same values at the same temperature. A network analyzer may be used to check the values.) Let the noise measured on the console with the room-temperature resistor be N_T while that measured with the liquid-nitrogen resistor is N_{77}. Then if α is now the overall gain from the pre-amplifier to the measured noise, from equations (6) and (10)

$$N_T^2 = \alpha^2\beta^2\left(N_S^2\frac{T}{290} + 2N_a^2\right);\quad N_{77}^2 = \alpha^2\beta^2\left(N_S^2\frac{77}{290} + 2N_a^2\right) \tag{11}$$

where N_S is the noise from a standard-temperature resistor.

From the definition of noise factor in equation (9),

$$F'^2 = 1 + \frac{2N_a^2}{N_S^2} \tag{12}$$

and solving equations (11) and (12) for the noise figure F at 290 K, we obtain

$$F = 10\ \log_{10}\left(\frac{213 - [290 - T]\,N_{77}^2/N_T^2}{290[1 - N_{77}^2/N_T^2]}\right) \tag{13}$$

For example, if the measured noise at 23 °C was 1.22 units and that at 77 K was 0.712 units, the noise figure of the system would be 0.51 dB.

Pre-amplifier Input Impedance Effects. It has been emphasized earlier that the pre-amplifier input impedance Z_a is generally very different from the optimum source resistance R_{opt} and also that the effective resistance r_c of a receiving coil has to be transformed with as little loss as possible to resistance R_{opt}. High-Q capacitors, inductors, PIN diode switches, and low-loss coaxial lines may all be involved, in sum or in part as needed, to effect this forward transformation, the exact details depending on the experiment. However, the corollary is that Z_a suffers a reverse transformation, producing in the process a new impedance Z_b *in series* with the receiving coil. By introducing an excess number of variables in the forward transformation (usually at the expense of a little sensitivity), Z_b can be manipulated within broad limits while maintaining the transformation $r_c \to R_{opt}$. Uses of these excess degrees of freedom include speeding the ringdown of a low-frequency probe after a pulse and blocking the flow of current in the probe during signal reception while maintaining sensitivity. Such blocking, obtained by maximizing Z_b so that $Z_b \gg r_c$, is useful for: suppressing radiation damping, an effect discovered over 40 years ago by scientists at the Perkin-Elmer Corporation but never published; minimizing change of signal strength when a conducting sample changes size (e.g., a perfused heart);[5] and reducing induced voltages when there is more than one coil and there is mutual induction M_{pq} between them,[6] a situation that is increasingly common as B_0 field strengths increase. Such intercoil coupling can correlate signals and noise and reduce the efficacy of an array of coils. (Comparison with the use of

constant current supplies for shim and gradient coils is apt.) The voltage V_q induced in coil q by signal voltage ξ_p in coil p is then approximately $j\omega M_{pq}\xi_p/(Z_b + r_c)$ rather than the much larger $j\omega M_{pq}\xi_p/r_c$. (It is assumed here, as is usually the case, that the receiving coil's inductance is approximately cancelled by capacitance.) The ratio of these two voltages, or synonymously that of the currents in coil p, depends mainly on the ratio of R_{opt} to the real part of Z_a as mediated by the transformation circuitry, and is typically in the range -17 to -25 dB: -40 dB would be ideal.

When multiple receiving coils are used, the intermediary of 50 Ω cable is usually dispensed with and a pre-amplifier placed directly on each coil. Consider the circuit of Figure 7a, which combines low-noise pre-amplification, PIN diode protection from large induced voltages during transmission, and both PIN diode and transistor current blocking. Ignoring inductor losses for the moment, the primary characteristic of the circuit is that inductor L_1 (red) has the same reactance X_1 as the coil matching capacitance C_m. Thus, when the PIN diode is turned on and has very low impedance, L_1 and C_m parallel-resonate, thereby creating a high blocking impedance in series with the receiver coil. When the diode is off, it follows that a high blocking impedance Z_b can be maintained if capacitor C_3 and inductor L_3 (blue) also resonate, albeit in series (the transistor's input impedance is assumed to be very large). In both cases, the value of (large) biasing inductor L_2 has little effect as it feeds a low impedance. It can then be shown that the condition

$$L_1 \cong L_3\sqrt{\frac{r_c}{R_{opt}}} \qquad (14)$$

must be obeyed for good noise performance.

To find the value of L_3, and indeed to optimize the circuit, we must consider resistive loss and the Q-factors of the inductors. The author favors toroids wound on small Teflon formers as Q-factors well over 100 can be obtained and they have negligible coupling. Once an average Q-factor is known, the circuit values are optimized by numerical minimization of a metric that includes insertion loss and the reciprocals of the two blocking factors. In general, insertion loss and current blocking trade against each other, but with care, it is possible to obtain an insertion loss of <-0.1 dB and blockings approaching -30 dB. It is usually found that equation (14) is obeyed but that L_3 and C_3 may not exactly resonate. In passing, note the use of (nonmagnetic) crossed diodes for extra transistor protection – all other components must be nonmagnetic too. Other design techniques include the use of self-rectification to power the PIN diode from the induced coil voltage during transmission, and the provision of power and/or PIN diode bias down the same cable used to send the signal to the receiver.

Finally, we note an esoteric and poorly understood fact about current blocking that becomes increasingly important at higher Larmor frequencies. When the length of conductor in a receiving coil begins to approach a wavelength, stray or parasitic capacitances to the surroundings (including other coils) can no longer be ignored and so symmetric operation, both physically and electrically, is employed in an attempt to minimize their effects. Now current blocking often effectively replaces a matching capacitance such as C_m in Figure 7a with large impedance Z_b that is resistive when maximized. Then, symmetry is only maintained if the mid-point of Z_b is on the coil axis of symmetry (the brown dashed line) *and that axis coincides with zero potential* ('ground') for the pre-amplifier. This is clearly not the case in the figure, but Figure 7b shows a circuit that fulfils this goal. The balun, in which C_1 and L_1 resonate, effectively replaces the matching capacitor C_m of Figure 7a. Once again, when the PIN diode is on, or if L_3 and C_3 are resonant, a high blocking impedance is placed in series with the receiving coil. The value of L_1 is again given by equation (14) and numerical optimization should still be employed to obtain insertion loss and blocking comparable with that of Figure 7a. Further discussion of the fascinating topic of balance is beyond the scope of the chapter.

The Receiver

Dynamic Range and Filtering. The signal and noise from the pre-amplifier pass out of the Faraday cage and into the receiver, whose job it is to prepare them for digitization and thence passage to the computer. The largest signals encountered in spectroscopy are from ^{1}H, and high-resolution proton spectroscopists are well accustomed to having to saturate the water solvent signal, as it can overload the receiver. However, biological ^{1}H signals are often from fat as well as water (see ***The Basics***), and while in the *frequency* domain, the fat spectrum is much smaller than that of water, thanks to greater linewidths, in the *time* domain, the initial height of the rapidly decaying fat FID may also overload the receiver, albeit only for a short time. It follows that the receiver must be designed to handle large signals (say ~0.2 V), unless it can be guaranteed that ^{1}H experiments will never be performed. Of course, the *smallest* observable frequency-domain signal may be buried in the noise in the time domain. The r.m.s. noise from the pre-amplifier is typically $\sim 20\sqrt{\Delta f}$ nV, where the bandwidth Δf may be many megahertz.

The ratio in the time-domain of the largest signal to the noise is the signal dynamic range S_{DR}, and to a limited extent, this is under engineering control because it involves the noise with its square root dependence on bandwidth. However, the ADC plays a key role in this regard. If buried signals are not to be lost, the converter must sample the noise adequately and it may be shown that its resolution (1 bit) should be smaller than approximately twice the r.m.s. noise voltage. In contrast, the *maximum* signal the ADC can handle is determined by its total number of bits B. Remembering that it must sample both positive and negative voltages, the ADC's dynamic range A_{DR} is then $\sim 2^B$. At audio frequencies, it is possible to obtain ADC's with very high resolution: B can be 18 or even 24 bits. However, with increasing frequency, the resolution diminishes, until, in the region of 300 MHz, the *effective* number of bits may only be 8 or 9 unless a fortune is paid. Note the use of the adjective *effective*: an ADC may be advertised as having B bits resolution, but with increasing sampling frequency, the resolution diminishes. To combat this problem, undersampling is sometimes used; in other words, a signal at the Larmor frequency f is sampled at a much lower frequency f_S. Provided the sampling aperture is much less than the Larmor period, this strategy combats the loss of bits quite well, but may actually

be counterproductive in improving dynamic range. The reason is that the bandwidth of the noise must be less than $f_S/2$ (the Nyquist sampling theorem) if noise is not to be aliased and sensitivity sacrificed. However, loss of bandwidth diminishes noise and thereby increases dynamic range, so more bits are needed!

Clearly, the trade-off between undersampling and dynamic range depends on the technology of analogue-to-digital conversion, which, as with all digital electronics, advances rapidly. However, at the time of writing, to be able to handle the MR experiment's dynamic range, it is common for frequency changing to be employed, wherein the frequency of the MR signal is reduced in a mixer. An example of how this can be done is shown in Figure 8a for a Larmor frequency of $f_0 = 300$ MHz. There, in a mixer and letting $\omega = 2\pi f$, the signal $A\cos(\omega_0 t)e^{-t/T_2}$ from the pre-amplifier is multiplied with a 279 MHz signal $\cos(\omega_1 t)$ from a frequency synthesizer. By simple trigonometry, sum and difference frequencies are generated and the signal $A/2\cos([\omega_0 - \omega_1]t)e^{-t/T_2}$ at 21 MHz is extracted with a 21 MHz filter. The latter restricts the bandwidth of the noise to 2 MHz and so gives a possible signal dynamic range of ~70 dB. The signal is then undersampled with an ADC at a rate of 4 MS/s (million samples per second, $f_s = 4$ MHz); several suitable and inexpensive 14-bit devices exist. As shown in Figure 8b, the resulting digitized signal clearly has a frequency of 1 MHz and in a simple approach, following Fourier transformation, the MR spectrum can be extracted from the small region it occupies in the midst of the 2 MHz spectral range. Other approaches include before Fourier transformation, digital reduction of the sampling frequency with signal processing hardware that maintains S/N ratio, a process known as *decimation*. Note that the figure also shows generation of a 1 MHz signal for the pulse programmer. This ensures that in a repetition of an experiment, the digitized MR signal always has the same phase.

Pulse Generation

Figure 8a also shows four signals that aid in the creation of the transmitter pulse, all of which are locked to a highly stable, master 1 MHz oscillator in a frequency synthesizer. Let the 279 MHz signals be $I_1 = \cos(\omega_1 t)$ and $Q_1 = \sin(\omega_1 t)$ and the 21 MHz signals have components $I_2 = \cos(\omega_2 t)$ and $Q_2 = \sin(\omega_2 t)$. While the generation of various pulse *shapes* and *phases* at given times lies in the domains of the computer, a micro-controller and digital-to-analogue converters (DACs), those pulses have to modulate the transmitter signal at the Larmor frequency and the four signals aid in that process. Let a complex pulse be expressed in the form $P(t) = R(t) + iS(t)$ where R and S are the outputs of two DACs. We combine these signals in so-called 'single-sideband mixers' such that the output is

$$\begin{aligned} V(t) &= R[\cos(\omega_2 t)\cos(\omega_1 t) - \sin(\omega_2 t)\sin(\omega_1 t)] \\ &\quad - S[\cos(\omega_2 t)\sin(\omega_1 t) + \sin(\omega_2 t)\cos(\omega_1 t)] \\ &= R\cos(\omega_2 + \omega_1)t - S\sin(\omega_2 + \omega_1)t \\ &\equiv \mathrm{Re}[P(t)\exp(i\omega_0 t)] \end{aligned} \tag{15}$$

The RF pulse $V(t)$ is then passed to the power amplifier. Part, or even all, of equation (15) may be constructed digitally.

Conclusion

Once the signal is safely in a computer, the burden of data handling shifts to the computer engineer and mathematician; the analysis of signal, the presentation of results, the storage of data, and the generation of pulses are all interesting topics but outside the scope of this brief survey. There are numerous esoteric design issues that have not been covered, particularly concerning the generation of the low-level transmitter signal and its phase relationships to the received signal sampling frequency and any mixing frequency. Those relationships must be highly stable with very small random variation (low phase noise) so that there is reproducibility between successive free induction decays or echoes. Nevertheless, the issues briefly discussed in the past few pages hopefully serve as an introduction to a fascinating, diverse, and complex discipline that continues to evolve and that is worthy of far deeper study by

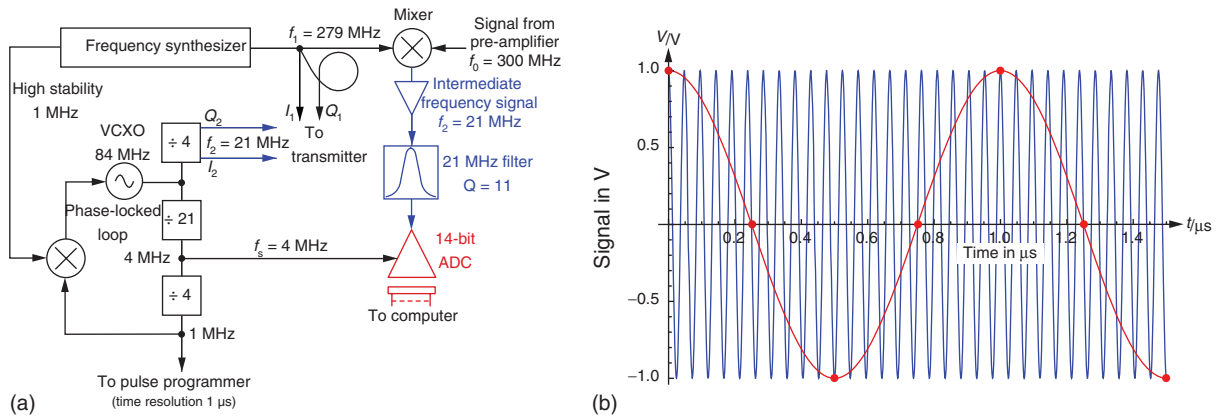

Figure 8. An example of frequency changing and undersampling for improvement of digital dynamic range. In (a) an intermediate frequency of 21 MHz is used but the signal is sampled at 4 MHz. The sampling (red points) is shown in (b) and the sampled signal clearly has a frequency of 1 MHz. After Fourier transformation, the MR spectrum is in the middle of the 2 MHz frequency range

spectroscopists if they wish to use their expensive equipment optimally. The following reading list may help.

Biographical Sketch

David Hoult. b. 1945. B.A. 1968, D. Phil. 1973, Oxford. Involved in biological magnetic resonance since 1969, most recently at NRC Canada. Has published more than 80 papers, has lectured extensively internationally, is coauthor of a book on MR technology, and has been on editorial boards of six journals. Recipient of numerous awards, including Gold Medal of ISMRM.

Related Articles

Resistive and Permanent Magnets for Whole Body MRI; Cryogenic Magnets for Whole Body Magnetic Resonance Systems; Weaver, Harry E.: Historical Comments on the Early Years of Superconducting NMR; Shimming of Superconducting Magnets; Shimming for High-Resolution NMR Spectroscopy; Hurd, Ralph E.: A Brief History of Gradients in NMR; Whole Body MRI: Local and Inserted Gradient Coils; Safety and Sensory Aspects of Main and Gradient Fields in MRI; Dadok, Josef: High-Field NMR Instrumentation; Multifrequency Coils for Whole Body Studies; Impedance Matching and Baluns; Receiver Design for MR; Radiofrequency Power Amplifiers for NMR and MRI

References

1. D. I. Hoult, R. E. Richards, and P. Styles, *J. Magn. Reson.*, 1978, **39**, 351.
2. D. I. Hoult, D. Foreman, G. Kolansky, and D. Kripiakevich, *MAGMA*, 2008, **21**, 15.
3. Microsemi Corporation, The PIN Diode Circuit Designer's Handbook, Microsemi Corporation: Watertown, MA, 1998.
4. D. M. Pozar, Microwave Engineering, Wiley: Hoboken, NJ, 2005.
5. D. I. Hoult and R. Deslauriers, *Magn. Reson. Med.*, 1990, **16**, 41.
6. P. B. Roemer, W. A. Edelstein, C. E. Hayes, S. P. Sousa, and O. M. Mueller, *Magn. Reson. Med.*, 1990, **16**, 192.

Further Reading

C.-N. Chen and D. I. Hoult, Biomedical Magnetic Resonance Technology, Adam Hilger: Bristol, 1989. This book is out of date but still useful.

D. L. Eggleston, Basic Electronics for Scientists and Engineers, Cambridge University Press: Cambridge, 2011. An introductory book at the undergraduate level.

P. Horowitz and W. Hill, The Art of Electronics, 2nd edn, Cambridge University Press: Cambridge, 1989.

P. Horowitz and W. Hill, The Art of Electronics, 3rd edn: , 2015. The definitive but rather overwhelming text. The third edition is eagerly awaited.