CMOS RF Biosensor Utilizing Nuclear Magnetic Resonance

Nan Sun, *Student Member, IEEE*, Yong Liu, *Member, IEEE*, Hakho Lee, *Member, IEEE*, Ralph Weissleder, and Donhee Ham, *Member, IEEE*

***Abstract*—We report on a CMOS RF transceiver designed for detection of biological objects such as cancer marker proteins. Its main function is to manipulate and monitor RF dynamics of protons in water via nuclear magnetic resonance (NMR). Target objects alter the proton dynamics, which is the basis for our biosensing. The RF transceiver has a measured receiver noise figure of 0.7 dB. This high sensitivity enabled construction of an *entire* NMR system around the RF transceiver in a 2-kg portable platform, which is 60 times lighter, 40 times smaller, yet 60 times more mass sensitive than a state-of-the-art commercial benchtop system. Sensing 20 fmol and 1.4 ng of avidin (protein) in a 5 μL sample volume, our system represents a circuit designer's approach to pursue low-cost diagnostics in a portable platform.**

***Index Terms*—Biosensor, CMOS integrated circuit, low noise amplifier, nuclear magnetic resonance, RF transceiver.**

I. Introduction

SILICON radio frequency (RF) integrated circuits (ICs) have been at the center stage of various wireless chip developments over the past years. Here we report on a silicon RF IC designed for a different application, that is, sensing biological objects.

In a disease development, biomolecules characteristic to the disease, such as viruses and cancer marker proteins, emerge. The ability to sense these biomolecules would facilitate disease detection. Researchers from many areas of science and engineering are developing a variety of biosensors, aiming at increased sensitivity or low-cost diagnostics [1]. Our RF biosensor is a "circuit designer's way" to pursue low-cost diagnostics in a portable platform.

Fig. 1 shows our RF biosensor, central to whose operation is the silicon RF transceiver IC. The underpinning physical phenomenon of our sensing method is nuclear magnetic resonance (NMR), the resonant interaction between RF magnetic fields and protons in water, which is altered by the presence of target biological objects. This NMR-based biosensing was developed in 2002 [2] and has been used within a state-of-the-art commercial benchtop NMR system [3] which, however, is bulky and heavy (120 kg). The main contribution of our work is to drastically shrink the *entire* NMR system by developing the RF transceiver IC, hence enabling the NMR-based biosensing in the portable form of Fig. 1. Occupying only 2.5 liters and weighing only 2 kg, our system is 60 times lighter, 40 times smaller, yet 60 times more mass sensitive than the state-of-the-art benchtop NMR system: our system is actually the smallest complete NMR system ever built. As a biosensor, it detects 20 fmol, 1.4 ng of avidin (protein) in a 5 μL sample volume. This detection threshold can be even further improved, for it is currently limited not by our transceiver sensitivity, but by the specific bioassay used [4], [5]. Our system might offer a new way to pursue low-cost diagnostics in a portable platform.

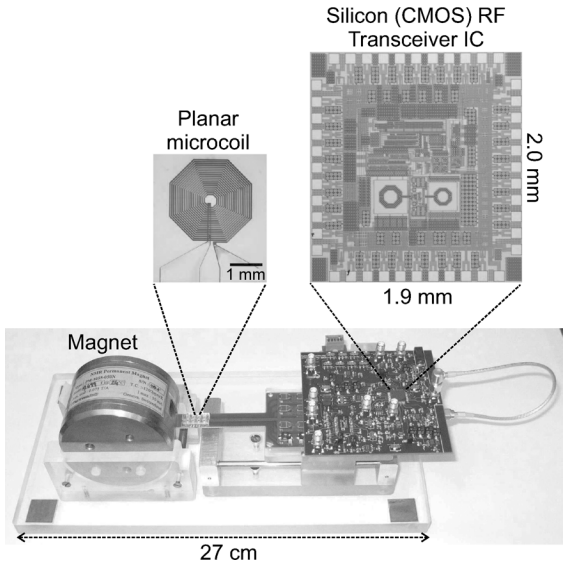


Fig. 1. CMOS RF biosensor (this work) utilizing NMR. The entire system weighs 2 kg, where the commercial magnet dominates the weight. Prior to this work, we built an intermediate miniature NMR biosensor [4] where we used the same magnet and in-house fabricated microcoil with discrete electronics.

After part of this work was briefly reported in the IEEE ISSCC [7], new key experiments ensued to complete this work. The goal of the present paper is to describe this entire work (design of, and experimentation with, the RF IC) in Sections IV–VII. We add two perspective sections, Sections III and VIII, to explain how our RF IC facilitated the dramatic miniaturization, and how this work is differentiated from existing miniaturization efforts. All of this would be best understood with some familiarity with NMR, so we start by reviewing the NMR basics.

II. Review of Nuclear Magnetic Resonance

Although NMR is a well-established subject [6], we provide this review to quickly introduce the basic concepts of NMR rel-

Manuscript received July 28, 2008; revised December 08, 2008. Current version published May 01, 2009.

N. Sun and D. Ham are with the School of Engineering and Applied Sciences, Harvard University, Cambridge, MA 02138 USA (e-mail: nansun@seas.harvard.edu; donhee@seas.harvard.edu).

Y. Liu was with Harvard University, Cambridge, MA 02138 USA, and is now with the IBM T. J. Watson Research Center, Yorktown Heights, NY 10598 USA.

H. Lee and R. Weissleder are with the Center for Systems Biology, Massachusetts General Hospital, Harvard Medical School, Boston, MA 02114 USA.

Digital Object Identifier 10.1109/JSSC.2009.2017007

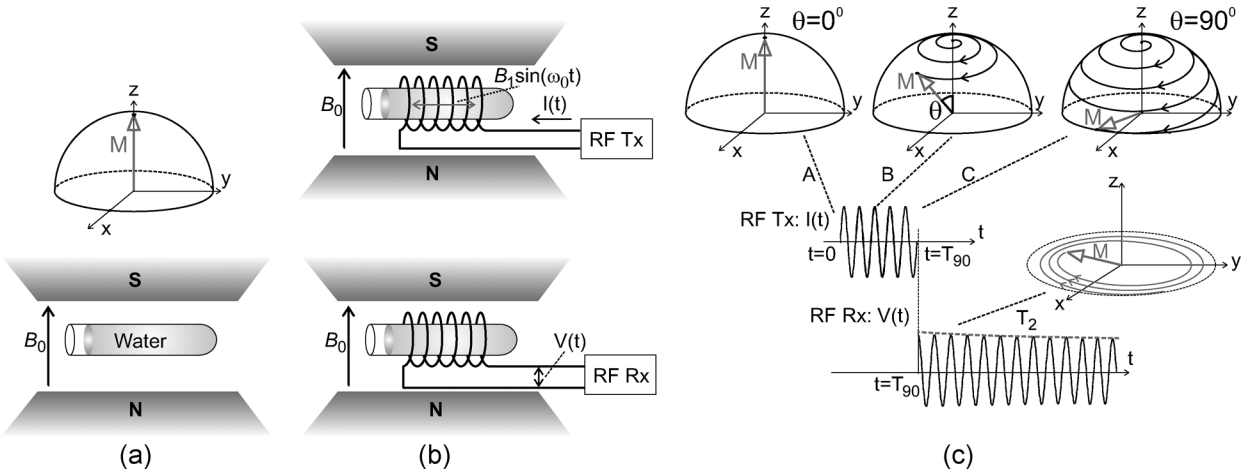


Fig. 2. (a) Water in a magnetic field B_0. (b) Excitation (top) and reception (bottom) phase in NMR. (c) Motion of magnetic moment M in NMR.

evant to this work, so that readers less familiar with NMR can rapidly access the main contribution portion of this paper.

A. Nuclear Magnetic Resonance

Atomic nuclei with spin act like tiny bar magnets. These nuclear magnets are the main actors in the NMR phenomenon. We explain NMR using protons, the simplest nuclei. Proton NMR is observed with as common a substance as water, for it contains numerous hydrogen atoms whose nuclei are protons.

Let water be placed in a static magnetic field B_0 [Fig. 2(a)]. After several seconds, the system reaches an equilibrium, where the proton magnets preferentially line up along the direction of B_0, just like a compass needle lines up along the Earth's field. Macroscopically, in the equilibrium, the water exhibits a magnetic moment M along the direction of B_0, which we take as the z-axis. This equilibrium state is the lowest energy state the system assumes.

Now if one wraps a coil around the water and sends an RF current into the coil, one can produce an RF magnetic field $B_1 \sin \omega_0 t$ across the water [Fig. 2(b), top], where B_1 and ω_0 are the amplitude and angular frequency of the RF field. If ω_0 is not the *Larmor frequency*, defined as

$$\omega_L = 2\pi f_L = \gamma B_0 \tag{1}$$

where γ is a proton-specific constant yielding $f_L = 21.3$ MHz for $B_0 = 0.5$ T, the proton magnets stay indifferent to the RF field and M remains along the z-axis. But for $\omega_0 = \omega_L$, proton magnets absorb energy from the RF field, raising the potential energy of M by misaligning M from B_0 with an increasing angle θ between them. In other words, M is tipped away from the z-axis. As shown at three time instants $\mathbf{A}$, $\mathbf{B}$, and $\mathbf{C}$ in Fig. 2(c), this downward motion of M is superposed with a much faster precession of M at the Larmor frequency ω_L about the z-axis. Overall, M exhibits a spiral downward motion, when absorbing energy from the RF field. This absorption of RF energy at ω_L by the proton magnets is termed nuclear magnetic resonance (NMR). During the energy absorption, the time taken for M to enter the xy-plane [$\theta = 90°$; $\mathbf{C}$, Fig. 2(c)] from its alignment along the z axis [$\theta = 0°$; $\mathbf{A}$] is

$$T_{90°} = \frac{\pi}{2\gamma B_1}. \tag{2}$$

Typically $B_0 \gg B_1$, which is why the downward motion of M (growth of θ) is much slower than its precession at $\omega_L = \gamma B_0$.

How does one figure out if ω_0 was tuned to ω_L and NMR took place? One can cease the RF transmission at $t = T_{90°}$, connect the coil to an RF receiver, and observe the coil output after $t = T_{90°}$ [Fig. 2(b), bottom]. If $\omega_0 \neq \omega_L$, NMR did not occur and M stayed along the z-axis. Thus, nothing registers at the receiver after $t = T_{90°}$. If $\omega_0 = \omega_L$, NMR took place during the RF field transmission $(0 < t < T_{90°})$ and M is in the xy-plane at $t = T_{90°}$ [Fig. 2(c)]. After $t = T_{90°}$, there is no more RF field for protons to absorb energy from, and hence, M will stay in the xy-plane, maintaining $\theta = 90°$ with a constant potential energy.[1] M, however, continues to precess with ω_L about the z axis in the xy plane. Therefore, the coil experiences a periodic magnetic flux variation that induces an RF voltage $V(t)$ of frequency ω_L across the coil. $V(t)$ registers at the RF receiver. Note that the RF excitation is stopped at $t = T_{90°}$ because M in the xy-plane induces the largest signal across the coil.

$V(t)$ is actually an exponentially damped sinusoid [Fig. 2(c)] with characteristic decay time called T_2. The damping arises because the precession of an individual proton magnet is randomly perturbed by tiny magnetic fields produced by nearby proton magnets (proton-proton interaction). Due to the perturbation, each proton precesses at ω_L but with its own phase noise. Consequently, precessions of a large number of individual protons, which were initially in phase, grow out of phase over time in the xy-plane. As phase coherence is lost, the vector sum M of individual proton magnetic moments exponentially decays, yielding the damped signal. For water, $T_2 \sim 1$ s and the quality

[1]As mentioned, after several seconds, M will lose energy to the environment, aligning along B_0 to reach equilibrium. The time considered here in the main text, however, is far shorter than the equilibrium restoration time.

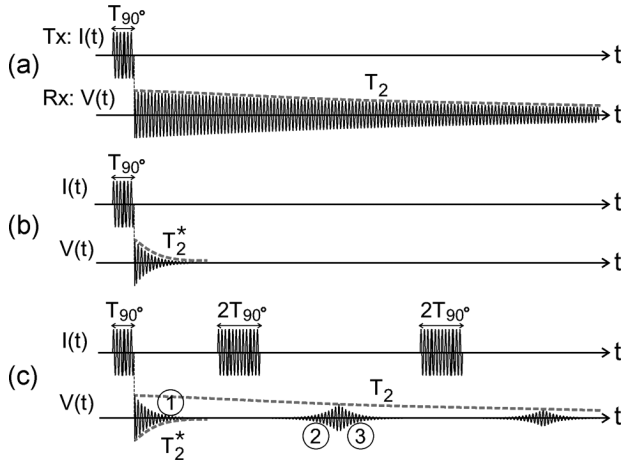


Fig. 3. (a) T_2 damping due to the basic proton-proton interaction. (b) T_2^* damping due to the static magnetic field inhomogeneity. (c) Echo technique.

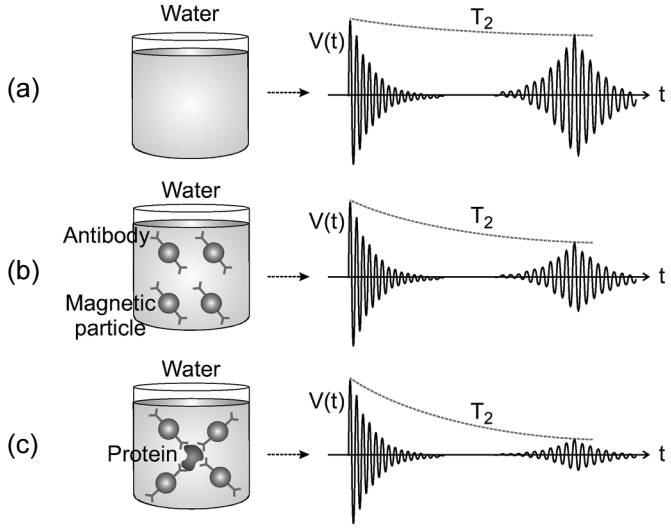


Fig. 4. NMR-based biosensing [2].

factor $Q = \omega_L T_2$ of the damped sinusoid with $f_L = 21$ MHz is about 10^8.

Through the NMR experiments described above, one obtains the values for ω_L and T_2, which carry information on the compositions and properties of the substances in study. In this way, NMR has found applications in fields such as chemical spectroscopy, medical imaging (MRI), and biosensing.

B. Echo Technique

We have so far assumed a uniform static magnetic field B_0 across the water. In this case, all protons precess at the same frequency $\omega_L = \gamma B_0$ with phase noise caused by proton-proton interactions, yielding the T_2-damped sinusoid $V(t)$ at the receiver [Fig. 3(a)]. In reality, any magnet suffers a spatial variation of the field magnitude around the intended value B_0. When such field inhomogeneity is present, $V(t)$ is damped much faster with decay time T_2^*, overriding the T_2 damping [Fig. 3(b)]. The faster damping arises because protons at different positions precess at different Larmor frequencies due to the field inhomogeneity, making them rapidly grow out of phase. To estimate T_2^*; assume that the field inhomogeneity is 50 parts per million (ppm) around $B_0 = 0.5$ T. If we denote this fractional variation as Δ, then the precession frequencies range from $\gamma(B_0 - \Delta B_0/2)$ to $\gamma(B_0 + \Delta B_0/2)$, and hence, the damped sinusoid of Fig. 3(b) has a bandwidth of $\gamma \Delta B_0$ around γB_0. Thus, $Q = \omega_L T_2^* \sim (\gamma B_0)/(\gamma \Delta B_0) = 2 \times 10^4$, or $T_2^* \sim 200\ \mu$s $\ll T_2 \sim 1$ s: even with 50 ppm inhomogeneity, the T_2^*-damping completely masks the T_2-damping.

Since field inhomogeneity is deterministic, one can effectively remove the T_2^* damping. Such a technique is important, as T_2, which carries useful information, would be immeasurable otherwise. Fig. 3(c) illustrates one widely used technique [6]. The T_{90° RF pulse is followed by repeated RF transmissions, each of duration $2T_{90^\circ}$. After the initial T_{90° pulse, the proton magnets brought down to the xy-plane precess at different Larmor frequencies: those at a stronger field position precess faster and ahead of those at a weaker field position, resulting in the T_2^* damping [①, Fig. 3(c)]. Now the $2T_{90^\circ}$ pulse performs the same function as the T_{90° pulse, but twice, changing θ by 180°. Thus, after the $2T_{90^\circ}$ pulse, the proton magnets return to the xy-plane, but with the faster precessing protons now lagging the slower ones (this inversion in relative positions may be difficult to grasp without detailed explanation, but let us accept it and move on, as the goal here is to become familiar with basic characteristics of the resulting signal, $V(t)$). The former will catch up the latter, regaining phase coherence, hence showing the growing sinusoid [②, Fig. 3(c)]. After all protons acquire the same phase, the faster precessing protons again get ahead of the slower ones, once again resulting in dephasing and the T_2^* damping [③, Fig. 3(c)]. As the $2T_{90^\circ}$ pulse repeats, multiple such "echoes" register at the receiver. Since T_2 damping arising from random proton-proton interaction cannot be removed, the envelope of the echoes slowly decays with T_2, allowing its measurement despite field inhomogeneity.

C. NMR-Based Biosensing

Consider detecting certain proteins (e.g., receptors of circulating cancer cells, or soluble cancer biomarker proteins such as VEGF, PSA, CEA, and AFP) in a blood sample. Nanoscale magnetic particles [8] coated with antibodies that specifically bind to the target proteins are suspended in the sample (Fig. 4). If the proteins exist, the magnetic particles self-assemble into clusters around the proteins [2]. This self-assembly can be detected by performing proton NMR: a significant portion of blood is water, so we may regard the blood sample as water containing many protons for our purpose.

Fig. 4 illustrates the echo signal $V(t)$ from precessing *protons in water* in three different cases. In Fig. 4(a), water with no magnetic particles yields T_2 due to the basic proton-proton interactions. Fig. 4(b) shows the case of water containing mono-dispersed magnetic particles in the absence of target proteins. The magnetic particles incessantly move around due to inevitable Brownian motion and produce fluctuating local magnetic fields, perturbing precessions of proton magnets and causing more phase noise than is expected from the basic proton-proton interactions. Thus, phase coherence is lost more rapidly, reducing T_2. In Fig. 4(c), due to the presence of target

proteins, the magnetic particles self-assemble into larger magnets, which are even more efficient in increasing phase noise in the proton magnets' precession, further accelerating dephasing, yielding the smallest T_2. Thus, by monitoring T_2, one can detect target objects [2].

III. CMOS RF Transceiver IC Design: Challenges

The NMR-based biosensing requires an NMR system that can faithfully measure T_2. To this end, the 120 kg commercial benchtop system [3] [Section I] has been used. The main focus of this work is the drastic miniaturization of the NMR system, enabling the NMR-based biosensing in the portable form of Fig. 1. Here we describe challenges that lie in such miniaturization, and how we met them through our RF IC design.

As seen in Fig. 2, any NMR system consists of a magnet, a coil, and an RF transceiver. By far the largest component is the magnet (e.g., the 120 kg weight of the benchtop system is largely due to a large magnet). Therefore, any effort to shrink an NMR system should start with using a small magnet. An NMR system with a small magnet, however, suffers from a reduced signal $V(t)$ at the input of the receiver.

To see this, we note that $V(t)$, induced in the coil due to the periodic variation of magnetic flux caused by the precession of M, is proportional to the rate of change of flux, ω_L, and the total magnetization $M \times [\text{sample volume}]$: the volume factors in here because M is measured per unit volume. Since $M \propto B_0$ at room temperature[2] and $\omega_L = \gamma B_0$, we have

$$V(t) \propto B_0^2 \times [\text{sample volume}]. \tag{3}$$

This equation explains why a small magnet yields a small $V(t)$. Firstly, in general a small magnet tends to produce a weak field B_0, and $V(t)$ is reduced with quadratic dependence on B_0. Secondly, a smaller magnet exhibits a more pronounced field inhomogeneity. In the presence of field inhomogeneity, protons at different positions of the sample have different Larmor frequencies, and thus, they are excited at different frequencies. Although the T_{90° RF pulse has other frequency components than $\omega_L = \gamma B_0$ due to its finite duration of T_{90°, it still has a limited bandwidth ($\sim 1/T_{90^\circ}$). Therefore, the T_{90° pulse can efficiently excite only protons in the region where their Larmor frequencies fall within the bandwidth of the T_{90° pulse. If the field inhomogeneity is increased with a smaller magnet, the spatial extent the T_{90° pulse can excite is reduced, effectively reducing the sample volume in (3) thus decreasing $V(t)$, regardless of the physical sample size. One may increase the bandwidth of the excitation pulse by decreasing the signal duration T_{90°, but this requires a larger B_1 [see (2)] and thus a larger transmitted signal power, which is electronically limited.

In summary, with a small magnet, $V(t)$ is substantially reduced and its detection requires a highly sensitive RF transceiver. For facile NMR experiments avoiding the stringent requirement for the RF transceiver, people have used large magnets, leading to the existing bulky NMR instrumentation.

[2]This well-established relation is derived from the Boltzmann's law, which we will use without showing the derivation.

To achieve the drastic miniaturization, we took the opposite approach: we use a very small magnet the size of an adult fist (Fig. 1), and to detect the resulting, substantially reduced signal $V(t)$ (amplitude: 130 nV; available power: 0.5 fW), we design a highly sensitive RF transceiver. By opting to use a very small magnet for miniaturization, we shifted the entire system design burden onto the RF transceiver design, which we successfully carried out. Furthermore, we integrated the RF transceiver onto a CMOS IC. As a very small, low cost magnet is used, the circuit integration makes sense: in a conventional system where a large, expensive magnet dominates the system size and cost, integration of the RF transceiver would hardly reduce either the system size or the cost.

Our design also uses a planar microcoil on a glass substrate (Fig. 1), having in mind mass production of the coil via standard microfabrication techniques. As compared to commonly used solenoidal coils, the planar coil's geometry does not allow effective induction of $V(t)$ from precessing protons, which was partly the reason for the weak signal amplitude of 130 nV. Once again, we overcame this undesirable effect by the design of the highly sensitive RF transceiver.

IV. CMOS RF Transceiver IC Design: Architecture

Fig. 5 shows the architecture of our NMR RF transceiver IC, which we describe now. As our magnet has $B_0 = 0.5$ T, the RF design is based at the Larmor frequency of 21.3 MHz.

A. Receiver Path

Let us first focus on the receiver path in Fig. 5. The signal appearing at node ② is a weak echo signal whose available power is 0.5 fW. The bandwidth of the echo signal is determined by the faster T_2^*-damping due to the field inhomogeneity, instead of the slower T_2-damping due to the proton-proton interaction. As discussed in Section II-B, the bandwidth corresponding to the T_2^*-damping is $\gamma \Delta B_0$ where Δ signifies the fractional field inhomogeneity. As we have $\Delta = 50$ ppm, the echo signal in the frequency domain (Fig. 5) is centered at the Larmor frequency 21.3 MHz, with a bandwidth of $\gamma \Delta B_0/(2\pi) \approx 1.1$ kHz.

The echo signal is amplified by a low noise amplifier, and then frequency down-converted using mixers and a local oscillator. Choosing the proper frequency for the local oscillator is a critical task in dealing with the weak echo signal. Let us denote the oscillator frequency as $\omega_L + \delta$. If δ is set to zero, the down-converted signal at node ③ is swamped by $1/f$ noise. On the other hand, if δ is too large, rejecting noise outside the signal bandwidth of 1.1 kHz after the frequency down-conversion is difficult, as it requires a high quality factor for the bandpass filter. Since the signal is very weak, rejecting out-of-band noise in order not to corrupt the signal-to-noise ratio (SNR) is a critical task. We select $\delta/(2\pi) = 3$ kHz, with which the down-converted signal is not centered at dc, but very close to dc (Fig. 5), facilitating out-of-band noise rejection using a bandpass filter with a moderate quality factor.

As in any heterodyning receiver, image noise would add 3 dB to the receiver noise figure, unless suppressed. We incorporate an image rejection algorithm in the digital domain after the

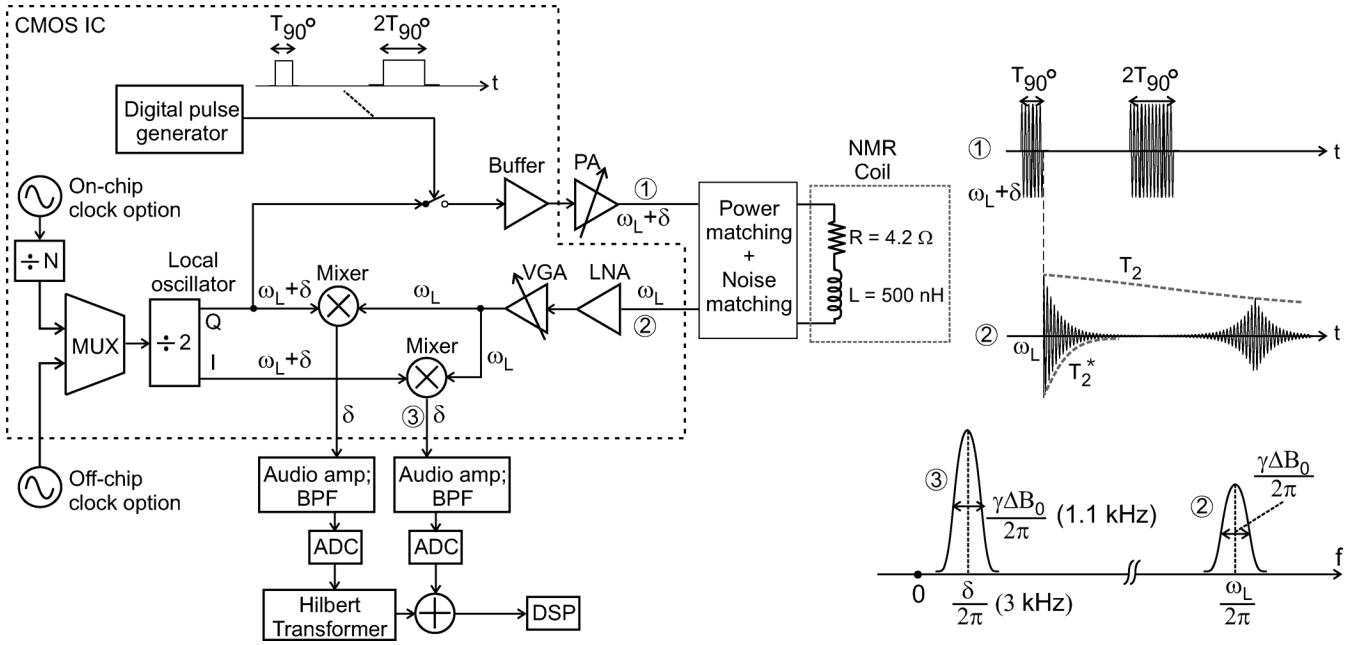


Fig. 5. The architecture of our NMR RF transceiver IC.

down-converted signal is digitized using analog-to-digital converters (ADCs) (Fig. 5). A Hilbert transformer shifts the phase of the signal down-converted by the quadrature-phase local oscillator by $-90°$. Addition of the Hilbert transformer output to the signal down-converted by the in-phase local oscillator nulls image noise.

B. Transmitter Path

Consider the transmitter path in Fig. 5. To produce the RF magnetic field in the coil for excitation of proton magnets, we use the same local oscillator used for the frequency down-conversion in the receiver. The frequency of the local oscillator, $\omega_L + \delta$, is not the Larmor frequency ω_L, but this arrangement is still effective in causing NMR, as we will explain shortly. The RF signal produced by the local oscillator is gated by the digital pulse generator, in order to generate the T_{90° pulse followed by repeated $2T_{90^\circ}$ pulses [Fig. 3(c)], which is to obtain the echo signal at the receiver.

How can the T_{90° pulse with the off-tuned frequency $\omega_L + \delta$ excite proton magnets in the sample? On one hand, the bandwidth of the T_{90° pulse may be taken as 30% (15 kHz) of $2/T_{90^\circ}$ (shaded region in Fig. 6) centered about $(\omega_L + \delta)/(2\pi)$. On the other hand, across our 5 μL sample, the static field inhomogeneity Δ is 50 ppm, and thus, the Larmor (excitation) frequencies of various positions of the sample range within the bandwidth of $\gamma \Delta B_0/(2\pi) = 1.1$ kHz (hashed region in Fig. 6) centered about $\omega_L/(2\pi) = \gamma B_0/(2\pi)$. Combining, the entire sample is excited, as the range of the Larmor frequencies of the sample lies well within the bandwidth of the T_{90° pulse. Of course, this scheme is possible because our sample is very small and the corresponding field inhomogeneity (and the resultant Larmor frequency variation) is relatively small.

We could use a separate oscillator for the transmitter, but in such a scenario, we would have to tune the frequencies

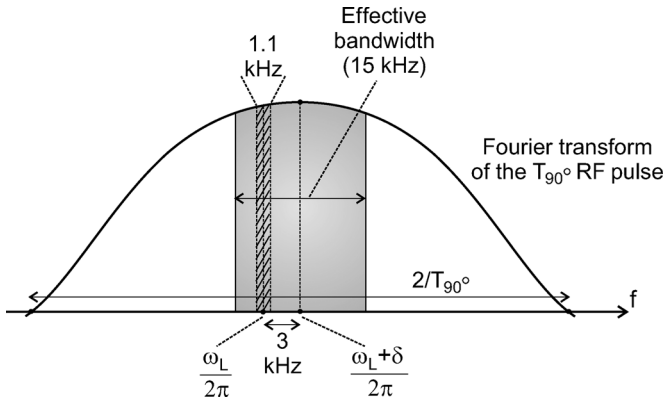


Fig. 6. Fourier transform of the T_{90° pulse with $\omega_L + \delta$ excitation frequency.

of both transmitter and receiver oscillators separately in the process of finding the frequency for NMR. In contrast, in our strategy sharing the same oscillator between the transmitter and receiver, we can just tune this single oscillator to find the right frequency for NMR, until the maximum signal appears at the output of the bandpass filter of Fig. 5.

V. CMOS RF Transceiver IC Design: Front-End Design

The planar microcoil (Fig. 1) is modeled as an inductor, $L = 500$ nH, in series with parasitic resistance, $R = 4.2\ \Omega$ (Fig. 5). Its quality is $Q = \omega_L L/R = 16$. Although the coil is inherently an inductor while a radio antenna assumes a 50 Ω impedance, the coil is analogous to the antenna in that it emits power and picks up signals. Just like in radios, the matching of the transceiver to the coil and the low noise design of the receiver amplifier are crucial, especially in our case where $V(t)$ is diminutive. Here we describe the front-end design.

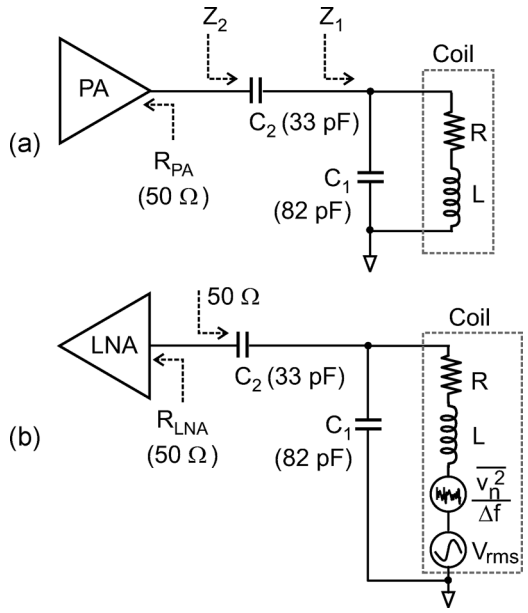


Fig. 7. 50 Ω impedance matching. (a) PA-coil. (b) LNA-coil.

A. Power Matching

The RF excitation pulses are sent to the coil through a power amplifier (PA). For maximum power transfer, the PA whose output impedance is $R_{\rm PA}$ (real number, e.g., 50 Ω) is to be impedance matched to the coil. Two capacitors C_1 and C_2 arranged next to the coil in Fig. 7(a) form an impedance matching network widely used in NMR system design. With suitably chosen C_1 and C_2, the real part of Z_1 becomes $R_{\rm PA}$ while Z_1's inductive reactance is resonated out by C_2, leading to $Z_2 = R_{\rm PA}$.

B. Noise Matching via Passive Amplification

In the receiving stage of NMR experiments, the coil is connected to a low noise amplifier (LNA) to amplify the signal $V(t)$ across the coil induced by precessing proton magnets. To minimize the noise figure, the LNA should be connected to the coil cautiously. In the widely practiced discrete design of the NMR RF transceiver, an off-the-shelf discrete LNA with a 50 Ω input impedance is impedance matched to the coil using the same $C_1 - C_2$ network discussed above. See Fig. 7(b), where $V_{\rm rms}$ in the coil's model is the rms value of $V(t)$ and $\overline{v_n^2}/\Delta f$ is the power spectral density of the thermal noise generated by the coil's resistance R. This predominantly used strategy will work satisfactorily, as long as the LNA's noise figure is small at the source impedance of 50 Ω.

Here we will seek a different LNA-coil connection option to obtain minimum noise figure, without demanding *a priori* either the 50 Ω input impedance for the LNA or the impedance matching condition. To this end, consider a general passive network between the coil and the LNA [Fig. 8(a)]. The passive network has a voltage transfer function $H(\omega)$, which includes the effect of LNA's input impedance, $Z_{\rm LNA}$. Define $\alpha \equiv |H(\omega_L)|$. If noise in the passive network is ignored, or equivalently, if the passive network has negligible loss (this crucial assumption will be justified shortly), the passive network will "amplify" both the voltage signal and noise of the coil by the same factor α, while maintaining the SNR: at the output of the passive network, the signal is $\alpha V_{\rm rms}$ and the noise power spectral density is $\alpha^2\overline{v_n^2}/\Delta f$.

If α is large enough to put $\alpha^2\overline{v_n^2}/\Delta f$ far beyond the LNA's input referred noise $\overline{v_{a,n}^2}/\Delta f$ [Fig. 8(b), left], the SNR is not degraded appreciably by the LNA, and the noise figure is small. In contrast, if α is not large enough and $\alpha^2\overline{v_n^2}/\Delta f$ is comparable to, or smaller than, $\overline{v_{a,n}^2}/\Delta f$ [Fig. 8(b), right], the SNR is degraded considerably by the LNA, resulting in a high noise figure. Therefore, noise figure minimization entails two tasks: maximization of α and minimization of the LNA's input referred noise, $\overline{v_{a,n}^2}/\Delta f$. We defer the discussion of the second task till Section V-D, and here discuss how to maximize α.

We can immediately note that $Z_{\rm LNA}$ should be infinite to maximize α, as any finite $Z_{\rm LNA}$ will only lower the voltage at the output of a given passive network. Now with $Z_{\rm LNA} = \infty$, we design a passive network that returns the maximum α. Various topologies may be worked out, and here we use the $C_1 - C_2$ network: Fig. 8(c). The aim is to choose C_1 and C_2 that maximize α. C_2 should be infinite (short circuit), as any finite impedance in the C_2 path will lower the voltage at the passive network's output. With $C_2 = \infty$, we choose C_1 that maximizes α, which is given as a function of C_1:

$$\alpha(C_1) = \frac{1}{\sqrt{(1-\omega_L R C_1 Q)^2 + (\omega_L R C_1)^2}} \tag{4}$$

where $Q = 16$ is the coil's Q. $\alpha(C_1)$ becomes maximum

$$\alpha_{max} = \sqrt{Q^2+1} \approx Q \tag{5}$$

when

$$C_1 = \frac{1}{\omega_L R} \cdot \frac{Q}{Q^2+1} \approx \frac{1}{\omega_L R Q} = \frac{1}{\omega_L^2 L}, \tag{6}$$

that is, when C_1 resonates with L at ω_L.

We have found that for a fixed LNA input referred noise, a shunt capacitor C_1 that resonates with the coil inductor at ω_L maximally amplifies the input signal and noise with a gain of Q, minimizing the noise figure. We assumed no noise (no loss) in the capacitor. This is to be justified, as any significant noise directly translates to noise figure, defeating the purpose of our technique. The assumption is indeed valid, for the quality factor of the capacitor is on the order of several hundreds, while the coil's Q is 16. In this connection, the LC resonator in the minimum noise figure configuration may be understood as a passive preamplifier with a gain of 16 but with almost zero noise figure, which is unattainable with active amplifiers. Forming the LC resonator as a preamplifier naturally exploits the intrinsic inductor L already available in the coil.

Note that the minimum noise figure configuration does not involve impedance matching: the impedance looking into the passive network is Q^2R, while $Z_{\rm LNA} = \infty$ [Fig. 8(c)]. Not the power, but the voltage is maximally delivered to the LNA, in achieving the minimum noise figure.

Numerical examples may help appreciate this technique. The maximum amplitude of $V(t)$ is 130 nV, thus $V_{\rm rms} = 92$ nV. The noise generated in the coil is

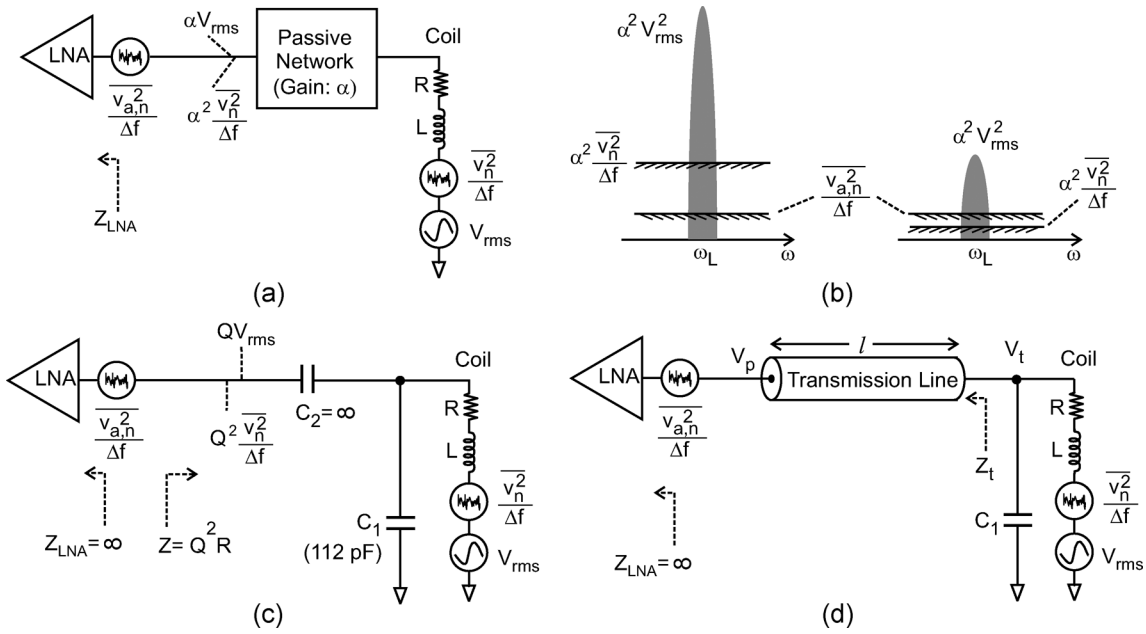


Fig. 8. (a) General passive network between a coil and an LNA. (b) Passively amplified signal and coil noise, and LNA's input referred noise in frequency domain for a large (left) & small (right) α. (c) Maximum passive amplification. (d) Maximum passive amplification in the presence of a transmission line.

TABLE I
LNA NOISE FIGURES FOR VARIOUS VALUES OF α

α	16	1	3.4
$\alpha^2 V_{rms}^2$	$(1471\ \text{nV})^2$	$(92\ \text{nV})^2$	$(313\ \text{nV})^2$
$\alpha^2 \overline{v_n^2}$	$(139\ \text{nV})^2$	$(8.8\ \text{nV})^2$	$(29.7\ \text{nV})^2$
$\overline{v_{a,n}^2}$	$(43\ \text{nV})^2$	$(43\ \text{nV})^2$	$(43\ \text{nV})^2$
Noise figure	0.4 dB	14 dB	4.9 dB

$\overline{v_n^2}/\Delta f = 4kTR = (0.26\ \text{nV}/\sqrt{\text{Hz}})^2$, and the input referred noise of our LNA is $\overline{v_{a,n}^2}/\Delta f = (1.3\ \text{nV}/\sqrt{\text{Hz}})^2$ [Section V-D]. From these values, we created Table I where the passively amplified rms signal power $\alpha^2 V_{rms}^2$, passively amplified coil noise over the signal bandwidth ($BW = 1.1$ kHz), $\alpha^2 \overline{v_n^2} = \alpha^2 \overline{v_n^2}/\Delta f \times BW$, and LNA's input referred noise over the bandwidth, $\overline{v_{a,n}^2} = \overline{v_{a,n}^2}/\Delta f \times BW$, are given for three values of α. For each α, it also shows the noise figure, which was calculated using $NF = 10\log[(\alpha^2 \overline{v_n^2} + \overline{v_{a,n}^2})/(\alpha^2 \overline{v_n^2})]$.

From Table I, we note:

- $\alpha = Q = 16$: In this optimum case [Fig. 8(c)], the passively amplified coil noise dominates the LNA's input referred noise, yielding a noise figure of only 0.4 dB.
- $\alpha = 1$: This is the case where the coil is directly connected to the LNA with $Z_{\text{LNA}} = \infty$ with no passive network. The LNA's input referred noise dominates the passively amplified coil noise. Consequently, a large noise figure of 14 dB results.
- $\alpha = 3.4$: If one uses the configuration of Fig. 7(b) with C_1 and C_2 yielding 50 Ω but with an LNA with $Z_{\text{LNA}} = \infty$, $\alpha = 3.4$. This is a scenario where the $C_1 - C_2$ network already designed for the impedance matching for the PA is directly connected to the LNA with $Z_{\text{LNA}} = \infty$. The noise figure is 4.9 dB.

As seen, the noise figure varies substantially with α, as it determines the relative difference between the passively amplified coil noise and LNA's input referred noise. The point of our technique is to maximize the passively amplified coil noise so that it is far larger than LNA's input referred noise.

C. Generalization of Passive Amplification

As NMR transceivers have been realized mostly at the discrete level, the 50 Ω matching with abundantly available 50 Ω off-the-shelf LNAs [Fig. 7(b)] has been the dominant choice for the LNA-coil connection. Regarding the resonance noise matching with the infinite impedance LNA [Fig. 8(c)], we suspected that this simple yet powerful technique might have been used during the 60 years of NMR, but there are very limited publications on the technique, and it is difficult to know the actual degree of its awareness and usage. It appears that the technique has been rarely conceived because custom design of an LNA (to produce non-50-Ω impedance) is seldom exercised in the conventional discrete NMR system design. We found only two papers [9], [10] reporting the technique, where it was empirically demonstrated but with no explanation of its underlying machinery. We also found a patent [11] that discusses it, but wrongly disqualifies it as the optimum option. In this connection, our foregoing elucidation of the resonance noise matching may help its further recognition and appreciation, and clarify its concept.

The reader may be concerned that the resonance noise matching may not work in the case the high impedance LNA is connected to a coil through a 50 Ω coaxial cable [Fig. 8(d)], due to impedance mismatches at both ends of the line (this is a legitimate concern in that even at 21 MHz, a line length of, say 20 cm, introduces a wave effect). This concern may be another reason for the dominance of the 50 Ω matching: when a 50 Ω cable is used, "everything 50 Ω" offers the most convenient design choice. Contrary to the concern, however, the resonance

noise matching works well with a more general passive network containing a transmission line.

To show this, let us work out the resonance noise matching in Fig. 8(d), based on our articulation of it from the previous subsection. The parameters of the transmission line are; characteristic impedance: Z_0, wave velocity: v, length: l, wave number: $\beta = \omega/v$. We seek to find C_1 that maximizes α for a fixed l. Assume $\beta l \ll 1$, although we can solve the problem generally. V_p and V_t in Fig. 8(d) are related through $V_p = V_t(\omega)/\cos(\beta l) \approx V_t$. The impedance Z_t is $Z_t = -jZ_0\cot(\beta l) \approx -jZ_0/\beta l = vZ_0/(j\omega l)$ since $Z_{\rm LNA} = \infty$. Therefore, the transmission line behaves like a capacitor $C_t = l(vZ_0)^{-1}$ at the coil end. Thus, to maximize α, L should resonate with $C_1 \| C_t$ at ω_L:

$$C_1 = \frac{1}{L\omega_L^2} - \frac{l}{vZ_0} \tag{7}$$

The resulting maximum α is again $Q = 16$. This design is valid, as the line has a negligible measured loss (less than 0.05 dB for a 30 cm cable).

D. Design of the Low-Noise Amplifier

As discussed in Section V-B, to minimize noise figure, α should be maximized and the LNA's input referred noise should be minimized. We have just discussed the former task. We now describe our LNA design to perform the latter. As $Z_{\rm LNA} = \infty$ is a prerequisite for maximum α, a common source topology is chosen for the LNA. It has a capacitive input impedance, which may be regarded as infinite at low enough frequencies.

Fig. 9 shows the schematic of our LNA. Since the design frequency, 21.3 MHz, is smaller than the $1/f$ noise corner ($\sim$50 MHz) of nMOS transistors in 0.18 μm technology used, we use pMOS devices with the smaller $1/f$ noise corner ($\sim$1 MHz) [12] as input transistors $\mathbf{M}_1$ and $\mathbf{M}_2$. pMOS transistors bring an additional benefit, the mitigation of the coupling of the substrate noise produced by the digital pulse generation circuits into the LNA input: the substrate noise would otherwise interfere with the weak signal at the LNA input, which is 4 μV even after the maximal passive amplification. On the same token, we must minimize the coupling of the strong local oscillator signal into the LNA input: it occurs first through the mixer, then, through the LNA's input-output coupling capacitors. To minimize the latter coupling, cascode transistors $\mathbf{M}_3$ and $\mathbf{M}_4$ are used. Finally, to suppress the common mode noise, the amplifier is designed in a differential form and a common mode feedback circuit (CMFB) is employed. The capacitor C_3 and the resistor R_3 are for frequency compensation.

With the arrangement above, the total noise of the LNA is dominated by the channel thermal noise of the transistors. The input referred noise of the LNA calculated through the standard small-signal analysis is given by

$$\overline{v_{a,n}^2} \approx 2\overline{v_{n,M_1}^2} + 2\overline{v_{n,M_5}^2}\cdot\frac{g_{m5}^2}{g_{m1}^2} + \frac{2\overline{v_{n,M_3}^2}}{g_{m1}^2 r_{ds1}^2} \tag{8}$$

where $\overline{v_{n,M_i}^2}$ ($i = 1$, 3 and 5) is the transistor $\mathbf{M}_i$'s gate-source referred channel thermal noise. We have ignored the noise of the CMFB, and assumed perfectly matched differential pairs, to obtain the general guideline for how to size the transistors.

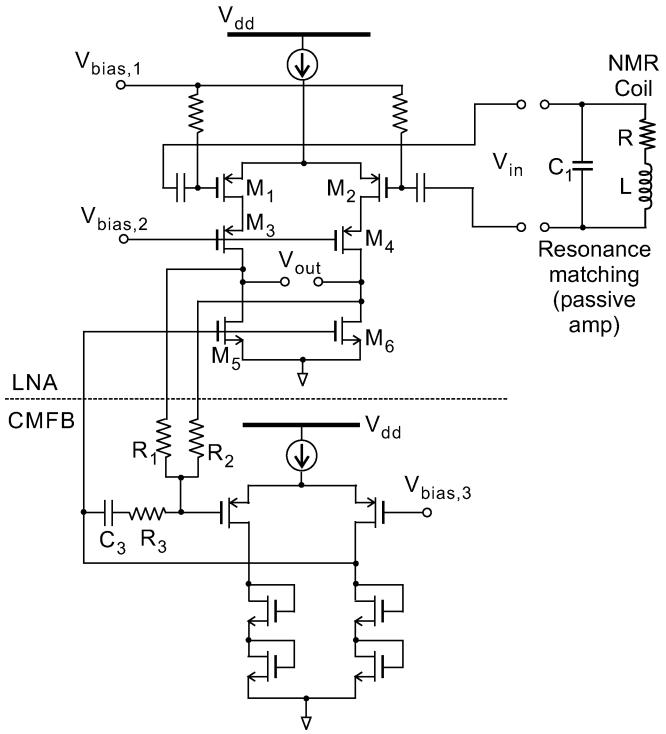


Fig. 9. LNA schematic.

The third term is negligible as $g_{m1}r_{ds1} \gg 1$. To minimize the second term, active loads $\mathbf{M}_5$ and $\mathbf{M}_6$ are made much smaller than input transistors $\mathbf{M}_1$ and $\mathbf{M}_2$: this leads to $g_{m5}/g_{m1} \ll 1$ and the total input-referred noise is dominated by the first term, $\overline{v_{n,M1}^2} \approx 4kT\gamma g_{m1}^{-1}$ where g_{m1} is the transconductance of the input transistors. To reduce this first term, a large tail current (4 mA) and wide input transistors (900 μm) are used to maximize g_{m1}. From SPICE simulation, the input-referred noise of the LNA at 21.3 MHz is 1.3 $\mathrm{nV}/\sqrt{\mathrm{Hz}}$ and its voltage gain is 41 dB. Its input impedance, even with the wide input transistors, is still large enough at 21.3 MHz to guarantee $Z_{\rm LNA} = \infty$.

E. Comparison With the Inductively Source Degenerated LNA

The passive amplification to minimize the noise figure is also employed in the inductively source degenerated LNA [Fig. 10(a)] widely used for wireless RF design [13], [14] although configured differently. As many designers are familiar with this LNA, here we compare it to our LNA.

In the inductively source degenerated LNA of Fig. 10(a) designed for a wireless receiver, $V_{\rm in}$ and $\overline{v_n^2}$ are the input signal and noise from an antenna, while $\overline{v_{gs}^2}$ is the gate-source referred channel thermal noise of the transistor. $R_a = 50\ \Omega$ is the antenna impedance. Inductors L_s and L_g and capacitor C_{gs} arranged around the transistor are chosen such that they resonate at the design frequency, and also present 50 Ω to the antenna: in Fig. 10(a), $Z_{\rm in} = L_s g_m/C_{gs} = 50\,\Omega$. This 50 Ω matching is required in wireless RF design, often reinforced by the band selection filter positioned between the antenna and the LNA. With such inductor and capacitor values, the LC network passively amplifies the input signal and noise, returning their amplified versions across the gate and source of the transistor [Fig. 10(b)] [13], [14]. The gain is $Q_e \equiv \omega(L_g + L_s)/(2R_a)$, the effective

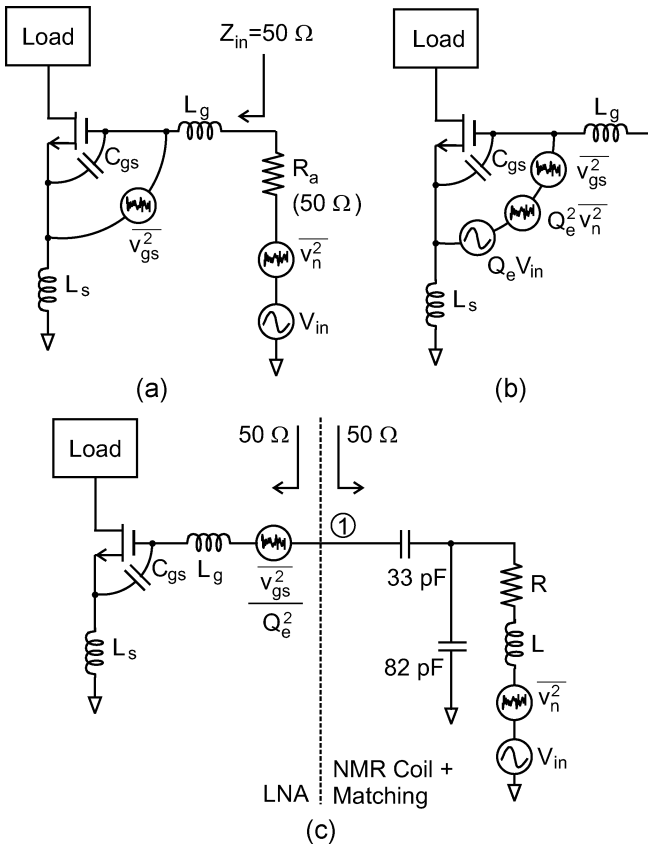


Fig. 10. (a) LNA with inductive source degeneration. (b) Passive amplification in this LNA. (c) Use of this LNA in Fig. 7(b).

quality factor of the LC tank including the antenna resistance (the factor 2 is because $Z_{\rm in} = R_a$). If Q_e is large enough to put $Q_e^2 \cdot \overline{v_n^2}$ far beyond the transistor noise $\overline{v_{gs}^2}$, the noise figure of the LNA will be small, as the SNR is minimally affected by the transistor noise. This is clearly the same passive amplification idea.

Passive amplification is realized differently between our LNA [Fig. 8(c) or 9] and the inductively source degenerated LNA (Fig. 10). In the wireless RF design where 50 Ω matching is required, the source degenerated LNA is well suited, as it simultaneously offers passive amplification and 50 Ω matching. In contrast, NMR system design does not demand 50 Ω matching. This, in conjunction with that the NMR coil has an inherent inductance, makes the implementation of passive amplification for NMR systems much simpler: the "free" inductor of the coil is used as an integral part of the LC passive amplifier, without adding extra inductors.

Note that the inductively source degenerated LNA with a 50 Ω impedance can be used for the widely used 50 Ω matched front-end of the NMR receiver [Fig. 7(b)]: such a design is shown in Fig. 10(c). Here although α is only 1.7, the noise figure is small because the LNA's input referred noise at node ① is significantly reduced, as it is the transistor's gate-source referred channel thermal noise $\overline{v_{gs}^2}$ divided by Q_e^2: this is another way of looking at passive amplification in the source degenerated LNA. This design, however, does not exploit the free inductance of the coil. It instead transforms the coil impedance to 50 Ω, then performs passive amplification, introducing extra inductors. Our design [Fig. 8(c) or 9] is thus a simpler choice as far as NMR system design is concerned.

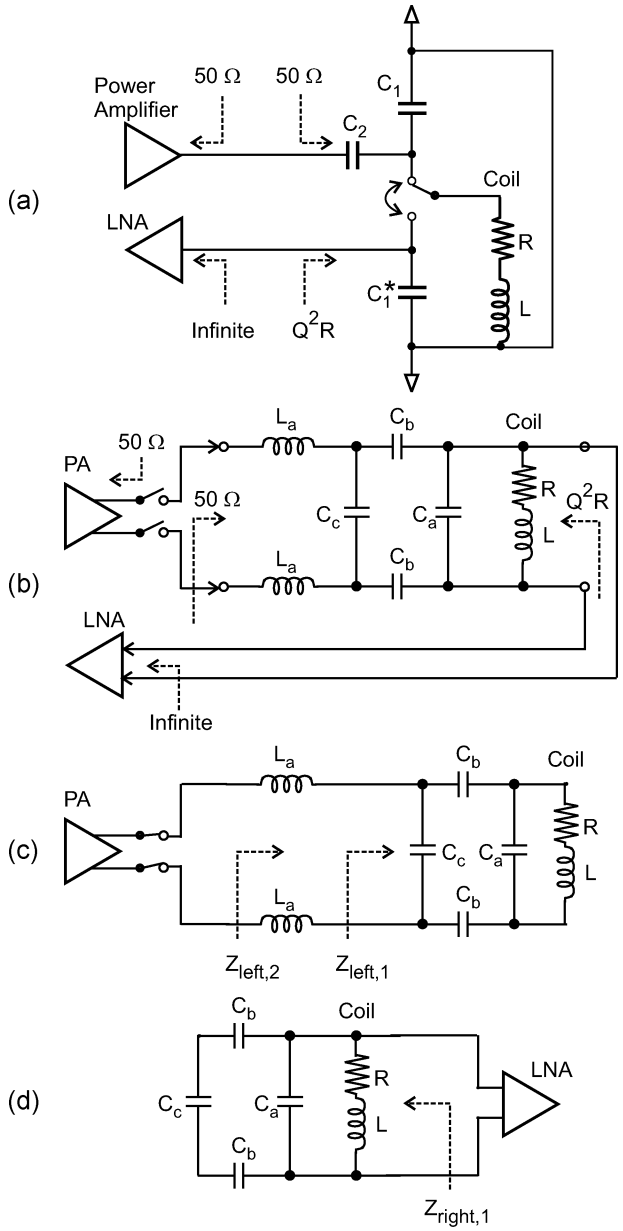


Fig. 11. Noise/power matching (a) with and (b) without a switch in the receiver path. (c) Transmitting mode. (d) Receiving mode.

F. Simultaneous Noise and Power Matching

We have seen that the receiver noise matching [Fig. 8(c)] and the transmitter power matching [Fig. 7(a)] require different passive networks. We may arrange the two networks separately, and to switch the coil to the LNA through the noise matching network during the receiving mode, and to the PA through the power matching network during the transmission mode [Fig. 11(a)]. This strategy, however, suffers from the loss associated with the switch. In the receiving mode, the turn-on loss of the switch will directly translates to noise figure.

To avoid the use of a switch in the receiving mode, we devised a new, differential matching network of Fig. 11(b). In the transmission mode, the switches next to the PA are turned on, and the PA is connected to the coil through the inductors (L_a) and capacitors (C_a, C_b, C_c) [Fig. 11(c)]. In the receiving mode, the switches are off, effectively making the inductors disappear, and the LNA is connected to the coil, whose impedance is modified by the capacitors (C_a, C_b, C_c) [Fig. 11(d)]. The first point of this circuit is that the LNA in the receiving mode and the PA in the transmitting mode see two different impedances, that is, $Z_{\mathrm{left},2} \neq Z_{\mathrm{right},1}$ in Fig. 11. Thus, one can choose capacitors and inductors appropriately to provide power matching for the PA and noise matching for the LNA. Note that the difference between $Z_{\mathrm{left},2}$ and $Z_{\mathrm{right},1}$ arises not merely from the fact that the transmitting path has additional inductors, but more importantly from the fact that the impedance of the network of capacitors and the coil seen from the left is completely different from that seen from the right, that is, $Z_{\mathrm{left},1} \neq Z_{\mathrm{right},1}$ in Fig. 11. The second point is that neither lossy switches nor lossy inductors are present in the receiving mode [Fig. 11(d)], and hence, the matching network hardly degrades the receiver noise figure.

We now consider how to select capacitor and inductor values quantitatively. For simplicity, we introduce the following notation: $X_L \equiv w_L L$, $X_a \equiv -1/(w_L C_a)$, $X_b \equiv -1/(w_L C_b)$, and $X_c \equiv -1/(w_L C_c)$. $Z_{\mathrm{right},1}$ in Fig. 11(d) is given by

$$\begin{aligned} Z_{\mathrm{right},1} &= (R + jX_L) \| jX_a \| (2jX_b + jX_c) \\ &= (R + jX_L) \| j \frac{X_a(2X_b + X_c)}{X_a + 2X_b + X_c} \end{aligned} \tag{9}$$

that is, when seen from the coil, the capacitors in Fig. 11(d) behave like one effective capacitor, whose impedance is to the right of the $\|$ symbol. The noise matching condition [see (6)] for $Q \gg 1$ is that the effective capacitor resonates with L at ω_L. This condition may be written as

$$X_L + \frac{X_a(2X_b + X_c)}{X_a + 2X_b + X_c} = 0. \tag{10}$$

In the transmitting mode, $Z_{\mathrm{left},1}$ of Fig. 11(c) is expressed as

$$\begin{aligned} Z_{\mathrm{left},1} &= jX_c \| [2jX_b + jX_a \| (R + jX_L)] \\ &= \frac{(X_a X_c)^2}{R(X_a + 2X_b + X_c)^2} + j \frac{(X_a + 2X_b)X_c}{X_a + 2X_b + X_c}. \end{aligned} \tag{11}$$

For power matching, the real part of $Z_{\mathrm{left},1}$ should equal the output resistance R_{PA} of the PA, while its capacitive reactance is to be resonated out by the inductors L_a:

$$2\omega_L L_a + \frac{(X_a + 2X_b)X_c}{X_a + 2X_b + X_c} = 0 \tag{12}$$

$$\frac{(X_a X_c)^2}{R(X_a + 2X_b + X_c)^2} = R_{\mathrm{PA}}. \tag{13}$$

If three conditions (10), (12) and (13) are simultaneously met, noise and power matching are simultaneously obtained. Since there are four unknown design parameters (C_a, C_b, C_c, and L_a), we have a large degree of freedom in choosing them.

VI. CMOS RF Transceiver IC Design: Other Circuit Considerations

A. Receiver Linearity

Although the planar microcoil is the ultimate part of our system, it is desirable for the system to be able to accommodate a variety coils for varying experimental needs. The coil's Q varies widely depending on its geometry and construction: e.g., while $Q = 16$ for the planar microcoil, $Q = 200$ for a solenoidal microcoil we built. The signal grows with the coil Q, especially with the resonance noise matching. To provide more amplification for a smaller signal, but to ensure that a larger signal does not enter the nonlinear regime of the mixer,[3] we designed a variable gain amplifier (VGA) following the LNA (Fig. 5). The receiver linearity is important, as we seek to accurately extract T_2.

The VGA is depicted in the middle of Fig. 12. It is a differential common source amplifier with two tunable loads shown within the two dashed boxes. The VGA's gain is altered by tuning the load impedance. Each load has three branches in parallel. Consider the branch built around transistor $\mathbf{M}_1$. As the control voltage V_c is increased from zero, the operation regime of $\mathbf{M}_1$ undergoes transitions from turn-off to pinch-off to triode, decreasing its drain-source resistance, $r_{ds} = \partial V_{ds}/\partial I_{ds}$. Therefore, the first branch serves as a voltage-controlled load. The second branch behaves essentially the same way, with decreasing r_{ds} of $\mathbf{M}_2$ with increasing V_c, but this time, the gate of $\mathbf{M}_2$ sees a voltage lower than V_c due to $\mathbf{M}_3$, thus, the impedance-versus-V_c curve for the second branch is that for the first branch, but translated along the V_c axis. The third branch involving three transistors, $\mathbf{M}_4$, $\mathbf{M}_5$, and $\mathbf{M}_6$, also exhibits the same impedance-versus-V_c curve, but even more shifted along the V_c axis. By combining the three branches in parallel, the load impedance is tuned more smoothly over a wider range of V_c, as compared to the case any single branch is used alone. The voltage gain tuning range of the VGA is from 0.8 to 22. The VGA's noise hardly contributes to the receiver noise figure, as the LNA's gain is large (41 dB). Therefore, little noise minimization effort is required for the VGA.

B. RX-TX Switching Scheme

As discussed in Section V-F, to avoid the use of lossy switches in front of the LNA, in our design the LNA stays connected to the coil during the T_{90° and $2T_{90^\circ}$ pulse transmissions, e.g., Fig. 11(b). In this approach, the large RF signal from the PA will saturate the receiver during the transmission mode. If the receiver is saturated during the transmission mode, restoration to its normal operation at the transition to the receiving phase (e.g., at $t = T_{90^\circ}$) will exhibit a settling behavior over the characteristic time of the receiver, compromising the initial reception of the signal.

To address this issue, during the transmission mode, we reduce the gain of the LNA by closing switch $\mathbf{S}_1$ in Fig. 12, preventing its saturation, and disconnect the rest of the receiver from the LNA by opening switches $\mathbf{S}_2$ and $\mathbf{S}_3$ of Fig. 12. Although $\mathbf{S}_2$ and $\mathbf{S}_3$ lie in the RF signal path, they come *after* the

[3]The input to our mixer can be as large as 10 mV within 5% gain reduction.

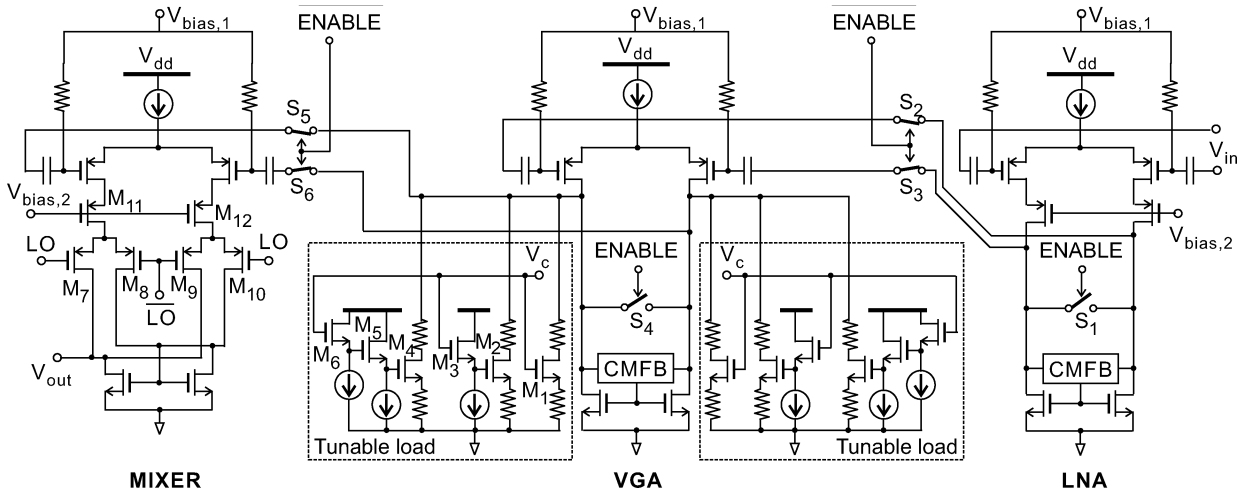


Fig. 12. Schematics of the LNA, VGA, and mixer, and their connections through a set of switches.

LNA substantially amplifies the signal, thus, their turn-on resistances hardly contribute to the overall noise figure. Since $\mathbf{S}_2$ and $\mathbf{S}_3$ are not perfect switches and some of the large RF signal could leak to the VGA via their parasitic capacitors during the transmission mode, we also place switch $\mathbf{S}_4$ to reduce the VGA gain during the transmission phase, and switches $\mathbf{S}_5$ and $\mathbf{S}_6$ to cut-off the mixer from the VGA.

C. Mixer

As shown at the left of Fig. 12, we use a double-balanced active mixer. Although the RF signal is down-converted to a low frequency of $\delta/(2\pi) = 3$ kHz, $1/f$ noise in the mixer does not interfere with the down-converted signal to any significant degree because the signal is amply amplified before the mixer and the LO generates a square wave. This is quantitatively confirmed, resorting to the study in [15]. The voltage gain of the mixer is 26 dB. The cascode transistors $\mathbf{M}_{11}$ and $\mathbf{M}_{12}$ are used to mitigate the undesired feed through from the LO to the LNA's input path, which could otherwise interfere with the weak signal from the coil.

VII. Experiments

We implemented the NMR transceiver in 0.18 μm CMOS technology. The IC occupies an area of 2.0 mm by 1.9 mm (Fig. 1). It contains, among components of Fig. 5, digital circuits for the pulse transmission, and most components associated with the heterodyning, that is, the LNA, VGA, mixers, and local oscillators including the frequency dividers. The matching network, PA, and ADCs were implemented with discrete components.

As we deal with weak echo signals, the most important aspect of our design is the receiver noise figure, which we obtained by directly measuring the receiver input referred noise. After feeding an RF signal to the receiver input (node ②, Fig. 5), we measured the output at node ③ using a spectrum analyzer. From this, we extracted the gain and output noise of the receiver. By dividing the output noise by the gain while factoring out image noise, we obtained the input referred noise of 1.8 $\mathrm{nV}/\sqrt{\mathrm{Hz}}$. Since the LNA has a voltage gain of 41 dB, the input referred noise is primarily contributed by the LNA. This measured input referred noise is larger than the SPICE-simulated value of 1.3 $\mathrm{nV}/\sqrt{\mathrm{Hz}}$. The extra amount originates mostly from the power supply and substrate noise, which were not taken into account in the simulation.

From the measured input referred noise and the measured impedance of the coil, we obtain the noise figure of the receiver. For the resonance noise matching with $\alpha = 16$ [Fig. 8(c)], the noise figure is only 0.7 dB. If we use the configuration of Fig. 8(c) but with C_1 and C_2 arranged for 50 Ω matching, $\alpha = 3.4$ and the noise figure is 7.0 dB. These results agree well with our time domain measurements of the echo signal. For $\alpha = 3.4$, we averaged 64 echo signals to obtain a clean echo signal. In contrast, for the resonance matching with $\alpha = 16$, only 4 time averaging was sufficient to obtain a clean echo signal, from which we were able to reliably extract T_2. This is an unequivocal proof of substantial reduction of the noise figure via the maximum passive amplification.

We assembled the entire NMR system of Fig. 1 by putting together the RF IC with auxiliary electronics on the printed circuit board, along with the planar microcoil and the magnet. The magnet and planar microcoil shown in Fig. 1 are borrowed from another work of ours [4], where the microcoil and its interface with sample were fully developed with extensive biosensing experiments but the electronics were at the discrete level with lower sensitivity and larger size. The magnet is a commercial product. The planar microcoil was in-house fabricated using standard photolithography. A 5 μL water sample is held on the microcoil inside a microfluidic container fabricated on top. In the following, we present NMR and biosensing experiments performed with the system.

A. NMR Experiments

We first performed proton NMR experiments in a phosphate buffered saline (PBS) solution, which is essentially biocompatible water. The top of Fig. 13 shows a down-converted echo signal measured at the output of the ADC of Fig. 5, which is

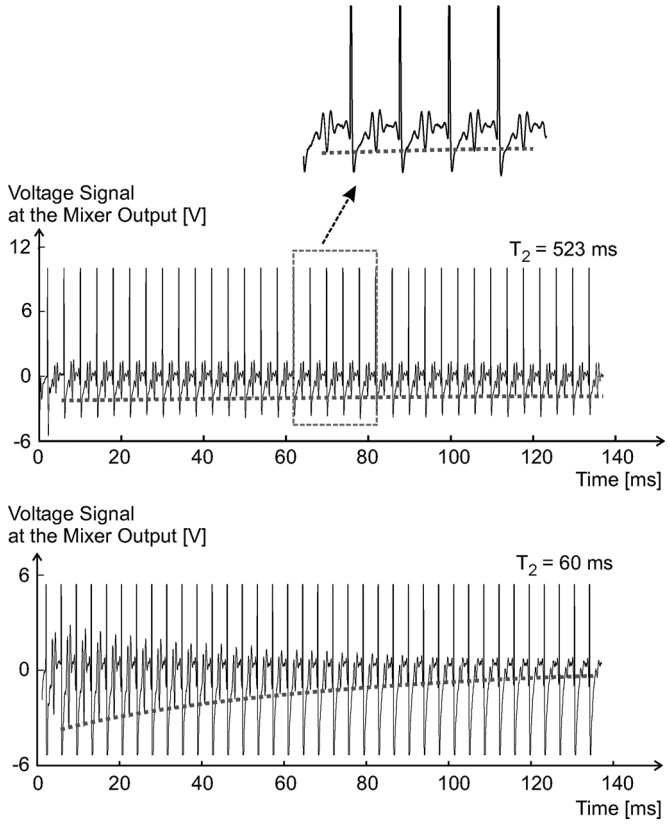


Fig. 13. Measured, down-converted echo signals: (top) pure PBS; (bottom) PBS with magnetic nanoparticles (0.17 mM).

directly interfaced with a computer. From the exponentially decaying envelope of the echoes, shown as a dotted line, we extract $T_2 = 523$ ms. The repeated large spikes are coupling of the transmitted RF sequence through unidentified coupling paths, but they do not compromise the observation of the conspicuous echo signal, as they occur at different times. As magnetic nanoparticles [Fe] (diameter: 30 nm) are suspended in the PBS with a concentration of 0.17 mM, the measured T_2 is decreased to 60 ms (Fig. 13, bottom), as discussed in Section II-C. We remark that the data presented here and in what follows were obtained using an external signal generator instead of the on-chip local oscillator. Although the on-chip oscillator yielded echo signals from which T_2 values could be extracted, they were noisier as the on-chip oscillator was a free running oscillator.

We discussed that T_2 decreases ($1/T_2$ increases) with an increasing concentration of the magnetic nanoparticles [Section II-C]. The NMR theory tells more specifically that $1/T_2$ increases linearly with the concentration [16]. Measured $1/T_2$ with a varying concentration of the magnetic particles in Fig. 14 conform to the theory, affirming our system's functionality.

While the planar microcoil ($Q = 16$) used to obtain Fig. 13 was ultimately part of our system due to its fabrication advantage, we performed NMR experiments using a higher-Q solenoidal microcoil ($Q = 200$) to show the echo signals more clearly, far beyond the spikes, as shown in Fig. 15. As the magnetic particle concentration changes from 0.1 mM to 1 mM, T_2 is reduced from 150 ms to 15 ms.

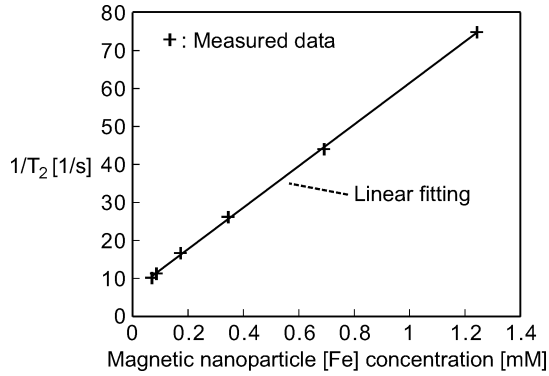


Fig. 14. Measured $1/T_2$ versus magnetic nanoparticle concentration in PBS.

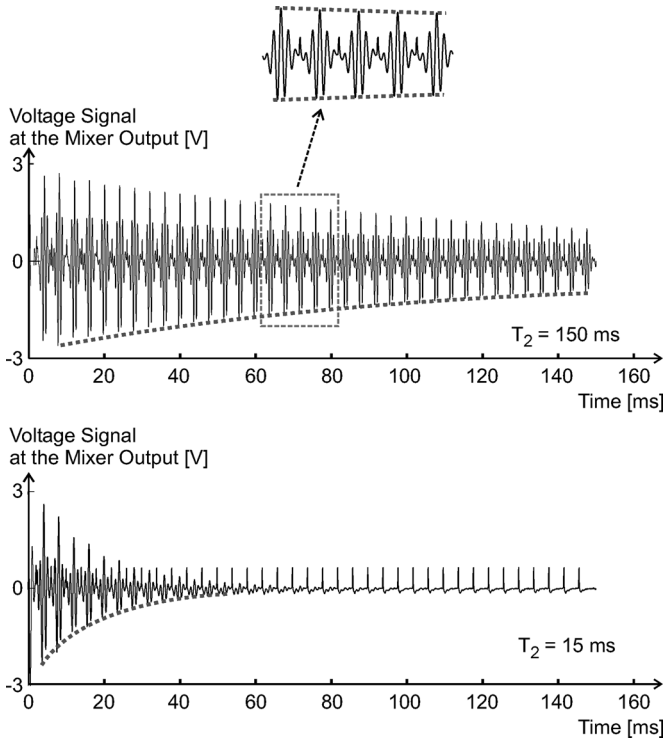


Fig. 15. Measured, down-converted echo signals for two different concentrations of magnetic nanoparticles; with a solenoidal microcoil.

B. Biomolecular Sensing

Our NMR system with the planar microcoil can be used to sense biomolecules, as described in Section II-C. Consider, for example, the detection of avidin (protein) in the PBS. Magnetic nanoparticles with surfaces coated with biotin (vitamin H) that specifically bind to avidin are suspended in the sample (Fig. 16). When avidin was absent in the sample, the magnetic particles were mono-dispersed, yielding T_2 of 73 ms (Fig. 16, top). When avidin existed in the sample with concentration of 320 nM, the magnetic particles self-assembled into clusters [2], reducing T_2 to 31 ms (Fig. 16, bottom), thus indicating the presence of avidin.

We repeated the experiment with a decreasing concentration of the avidin. The minimum detectable quantity of avidin was 20 fmol and 1.4 ng in the 5 μL sample volume (Fig. 17). In comparison to the state-of-the-art commercial benchtop NMR

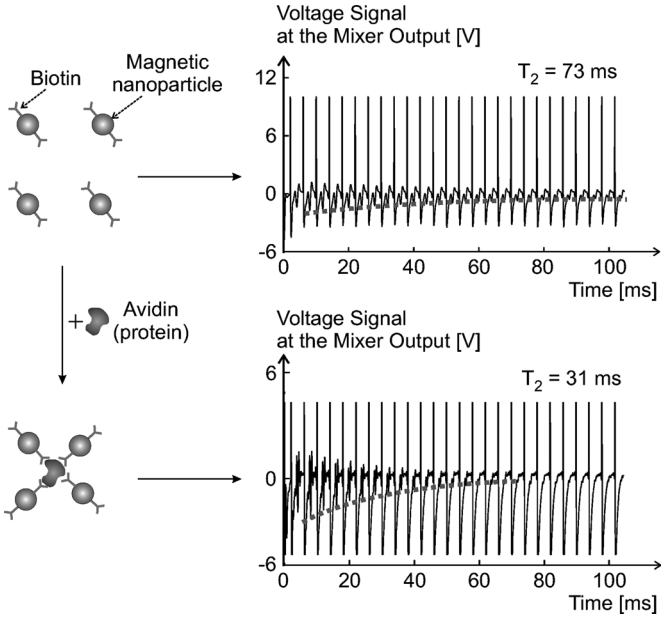


Fig. 16. Experimental avidin detection.

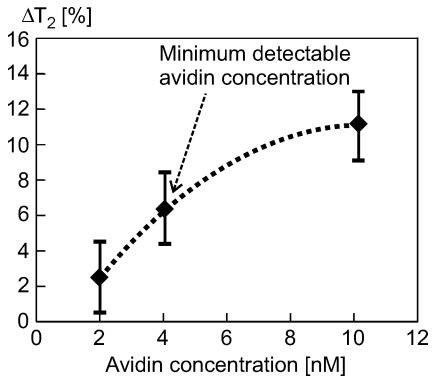


Fig. 17. Measured ΔT_2 for various avidin concentrations.

system [3], our system is 60 times more sensitive in terms of the absolute detectable quantity (20 fmol), thus, our system is far more advantageous in desired mass-limited sensing applications. Even a better detection threshold can be attained via improvement in the bioassay, for our current detection threshold is limited not by our transceiver, but by the specific bioassay used [4], [5]. We also successfully carried out the detection of another biomolecule, folic acid, whose presentation is omitted here due to page limitation.

For more extensive biosensing experiments, we refer readers to another work of ours in [4], where the same magnet and planar microcoil are used while the transceiver electronics are implemented in a discrete form with a lower receiver sensitivity.

VIII. Comparison to Other Miniaturization Works

Before concluding, we compare this work to other NMR system miniaturization efforts [4], [9], [17]–[20]. As shown in Table II, these works miniaturized only parts of the NMR system, but not the entire NMR system as we did. Thus, they are all much larger than our system.

TABLE II
NMR System Miniaturization Efforts

	Transceiver integration	Small (portable) magnet	Micro -coil
This work	◯ Receiver; pulse gen.	◯	◯
[4]	×	◯	◯
[17]	×	◯	◯
[9]	◯ LNA	×	◯
[18]	◯ Receiver	×	◯
[19]	◯ LNA; PA	×	◯
[20]	◯ LNA	×	◯

For the works that (partially) integrated the transceiver [9], [18]–[20], none of them performed the NMR experiment with the small magnet that would substantially reduce the NMR signal as in our case, and thus, the circuit performance optimization was not the major issue. The integration level in [9], [19], [20] is low. The integration level in [18] is the highest among them, and this work represents a significant advance in art. Their target application is the measurement of various magnetic field strengths, and due to the difficulty of broadband noise matching, they directly connected their LNA to the coil with $\alpha = 1$, inevitably degrading the noise figure performance.

Overall, in terms of the entire system size, integration level, and circuit performance, our work is a step forward from the existing NMR miniaturization works.

IX. Conclusion

Combining RF microelectronics and biotechnology in the unifying framework of NMR, this work showcases how silicon RF ICs can be used not only for cell phones, but also for biosensing aimed at improving human health care. The portable, low-cost NMR system, made possible by the development of the highly sensitive CMOS RF transceiver with a noise figure of 0.7 dB, is an addition to the growing library of emerging biosensors aimed at early disease diagnosis and/or at personalized medical care. We are now developing an even smaller, handheld NMR system by using an even smaller magnet: this increases the associated burden on the transceiver IC design, thus more joy for circuit designers.

Acknowledgment

The authors thank M. Cole of Oxford University, W. Hill, W. Andress, and K. Donoghue of Harvard University, and D. Ricketts of Carnegie Mellon University for their valuable suggestions.

References

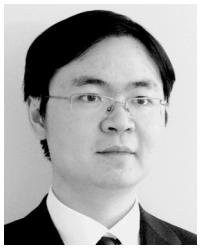
[1] D. Ham and R. M. Westervelt, "The silicon that moves and feels small living things," *IEEE Solid-State Circuits Society Newsletter*, vol. 12, no. 4, pp. 4–9, Oct. 2007, Fall issue.

[2] J. M. Perez, L. Josephson, T. O'Loughlin, D. Hoegeman, and R. Weissleder, "Magnetic relaxation switches capable of sensing molecular interactions," *Nature Biotechnology*, vol. 20, pp. 816–820, Aug. 2002.

[3] [Online]. Available: http://www.brukeroptics.com/minispec.html

[4] H. Lee, E. Sun, D. Ham, and R. Weissleder, "Chip-NMR biosensor for detection & molecular analysis of cells," *Nature Medicine*, vol. 14, no. 8, pp. 869–874, Aug. 2008.

[5] I. Koh, R. Hong, R. Weissleder, and L. Josephson, "Sensitive NMR sensors detect antibodies to infuenza," *Angew. Chem. Int. Ed. Engl.*, vol. 47, no. 22, pp. 4119–4121, 2008.
[6] D. Canet, *Nuclear Magnetic Resonance: Concepts and Methods*. New York: Wiley, 1996.
[7] Y. Liu, N. Sun, H. Lee, R. Weissleder, and D. Ham, "CMOS mini nuclear magnetic resonance system and its application for biomolecular sensing," in *IEEE Int. Solid-State Circuits Conf. Dig. Tech. Papers*, Feb. 2008, pp. 140–141.
[8] H. Lee, Y. Liu, R. M. Westervelt, and D. Ham, "IC/microfluidic hybrid system for magnetic manipulation of biological cells," *IEEE J. Solid-State Circuits*, vol. 41, no. 6, pp. 1471–1480, Jun. 2006.
[9] T. Cherifi *et al.*, "A CMOS microcoil-associated preamplifier for NMR spectroscopy," *IEEE Trans. Circuits Syst. I*, vol. 52, no. 12, pp. 2576–2583, Dec. 2005.
[10] R. Meller and D. Hartill, "Pulsed NMR magnetometers for CESR," in *Proc. Particle Accelerator Conf.*, 2003, pp. 2339–2341.
[11] "Signal/noise ratio optimization tuning system," U.S. patent 5,347,222, Sep. 13, 1994.
[12] D. M. Binkley, J. M. Rochelle, B. K. Swann, L. G. Clonts, and R. N. Goble, "A micropower CMOS direct-conversion, VLF receiver chip for magnetic-field wireless applications," *IEEE J. Solid-State Circuits*, vol. 33, no. 3, pp. 344–358, Mar. 1998.
[13] T. H. Lee, *The Design of CMOS Radio-Frequency Integrated Circuits*. Cambridge, MA: Cambridge Univ. Press, 1998.
[14] D. K. Shaeffer and T. H. Lee, "A 1.5-V, 1.5-GHz CMOS low noise amplifier," *IEEE J. Solid-State Circuits*, vol. 32, no. 5, pp. 745–759, May 1997.
[15] H. Darabi and A. A. Abidi, "Noise in RF-CMOS mixers: A simple physical model," *IEEE J. Solid-State Circuits*, vol. 35, no. 1, pp. 15–25, Jan. 2000.
[16] A. Roch, R. N. Muller, and P. Gillis, "Theory of proton relaxation induced by superparamagnetic particles," *J. Chem. Phys.*, vol. 110, no. 11, pp. 5403–5411, Mar. 1999.
[17] G. Eidmann, R. Savelsberg, P. Blümler, and B. Blümich, "The NMR mouse, a mobile universal surface explorer," *J. Magn. Resonance*, ser. Series A, vol. 122, pp. 104–109, 1996.
[18] G. Boero *et al.*, "Fully integrated probe for proton nuclear magnetic resonance magnetometry," *Rev. Sci. Instruments*, vol. 72, pp. 2764–2768, Jun. 2001.
[19] L. Fan *et al.*, "Miniaturization of magnetic resonance microsystem components for 3-D cell imaging," in *IEEE ISSCC Dig. Tech. Papers*, Feb. 2007, pp. 166–167.
[20] R. Magin *et al.*, "Miniature magnetic resonance machines," *IEEE Spectrum*, p. 51, Oct. 1997.



Nan Sun (S'06) received the B.S. degree in electrical engineering from Tsinghua University, Beijing, China, in 2006, where he ranked the top first out of 160 students in the Electrical Engineering Department every single year, and graduated with the highest honor and outstanding undergraduate thesis award. He is currently working towards the Ph.D. degree in electrical engineering at Harvard University, Cambridge, MA.

His Ph.D. research is twofold. First, he has developed several CMOS RF biomolecular sensors utilizing nuclear magnetic resonance (NMR) to pursue early disease detection and low-cost medicine. These systems can be used not only for low-cost medical diagnostics, but also for oil detection and quantum computing on silicon. Second, he is developing digital background calibration techniques for pipelined analog-to-digital converters. In addition to these two works, he has also made a key contribution to proving the chaotic nature of uncontrolled soliton oscillators by performing a Lyapunov analysis.

Mr. Sun received the first-class Tsinghua University Outstanding Student Award in each year from 2003 to 2006. He won the Top Prize in the Intercollegiate Physics Competition in 2003. He is the recipient of the Samsung Outstanding Student Award in 2003, Hewlett Packard Outstanding Student Award in 2006, and Analog Device Outstanding Student Designer Award in 2007. He won the Harvard Teaching Fellow Award in 2008. He has filed 4 US provisional patents.

Yong Liu (S'03–M'07) received the B.S. and M.S. degrees from Tsinghua University, Beijing, China, in 2000 and 2003, and the Ph.D. degree from Harvard University, Cambridge, MA, in 2007, all in electrical engineering. He is currently with the IBM T. J. Watson Research Center, Yorktown Heights, NY. His Ph.D. work examined applications of CMOS ICs in medicine and biotechnology, by directly interfacing CMOS ICs with biological systems. Specifically, his work consisted of two projects. First, he developed CMOS ICs in conjunction with microfluidic systems to magnetically manipulate individual biological cells for cell sorting applications. Second, he developed a CMOS RF biosensor utilizing NMR for medical diagnostics in a portable platform. Additionally, he worked on RF and mixed-signal ICs for communication and computation systems during his stay at Harvard, including autonomic PLLs for multi-domain synchronous clocking. His current research at IBM is on wireline transceiver design. He has 14 publications, two U.S. patents, and three provisional patents filed.

In 2001, he was with Tsinghua Tongfang Microelectronics Co., working on the second-generation Chinese RF ID card. In 2005 and 2006, he was with the Mixed-signal Communications IC Group at IBM T. J. Watson Research Center as a summer intern. He was the recipient of the Seagate Scholarship, Motorola Scholarship, Second Prize in the China National Graduate EDA Competition, and Analog Devices Outstanding Student Designer Award. He is a co-winner of the Beatrice Editorial Excellence Award in the 2009 IEEE International Solid-State Circuits Conference (ISSCC).

Hakho Lee (M'05) is an Instructor in Biomedical Engineering program at the Center for Systems Biology at the Massachusetts General Hospital (MGH) and Harvard Medical School (HMS). He received the Ph.D. degree in physics from Harvard University in 2005, and joined the Center for Molecular Imaging Research at MGH/HMS as a Research Associate. He has contributed to developing a new type of micro-total-analysis-systems by combining integrated circuits and microfluidics and developed the miniature NMR system. Dr. Lee's research topics include the synthesis of novel magnetic particles and the development of microelectronic systems for medical diagnosis.

Ralph Weissleder is a Professor at Harvard Medical School, Director of the Center for Systems Biology at Massachusetts General Hospital (MGH), Director of the Center for Molecular Imaging Research, and Attending Clinician at MGH. Dr. Weissleder is also a member of the Dana Farber Harvard Cancer Center, an Associate Member of the Broad Institute (Chemical Biology Program), and a member of the Harvard Stem Cell Institute (HSCI). Dr. Weissleder's research interests include the development of novel molecular imaging techniques, tools for detection of early disease detection and development of nanomaterials for sensing and systems analysis. His research has been translational and several of his developments have led to advanced clinical trials with anticipated major impacts when these methods become routinely available. He has published over 500 original publications in peer reviewed journals and has authored several textbooks. His work has been honored with numerous awards including the J. Taylor International Prize in Medicine, the Millennium Pharmaceuticals Innovator Award, the AUR Memorial Award, the ARRS President's Award, the Society for Molecular Imaging Lifetime Achievement Award, the Academy of Molecular Imaging 2006 Distinguished Basic Scientist Award, and the 2008 RSNA Outstanding Researcher Award.

Donhee Ham (S'99–M'02) is John L. Loeb Associate Professor of the Natural Sciences and Associate Professor of Electrical Engineering at Harvard University, Cambridge, MA, where he is with the School of Engineering and Applied Sciences (Programs: Electrical Engineering and Applied Physics).

He received the B.S. degree in physics from Seoul National University, Korea, in 1996, where he graduated *summa cum laude* with the Valedictorian Prize as well as the Presidential Prize, ranked top 1st across the Natural Science College, and also with the Physics Gold Medal (sole winner). Following 1.5 years of mandatory military service in the Republic of Korea Army, he proceeded to California Institute of Technology, Pasadena, CA, where he received the M.S. degree in physics in 1999 working on general relativity and gravitational astrophysics, and Ph.D. degree in electrical engineering in 2002 winning the Charles Wilts Doctoral Thesis Prize, Best thesis award in Electrical Engineering. His doctoral work examined the statistical physics of electrical circuits. He was the recipient of the IBM Doctoral Fellowship, IBM Faculty Partnership Award, IBM Research Design Challenge Award, Li Ming Scholarship, Silver Medal in National Math Olympiad, and the fellow of the Korea Foundation of Advanced Studies. He shared Harvard's Hoopes prize with William Franklin Andress. He was recognized by *MIT Technology Review* as among the world's top 35 young innovators in 2008 (TR35), for his group's work on CMOS RF biomolecular sensor using nuclear spin resonance to pursue early disease detection and low-cost medicine.

Dr. Ham's work experiences include Caltech-MIT Laser Interferometer Gravitational Wave Observatory (LIGO), IBM T. J. Watson Research Center, IEEE conference technical program committees including the IEEE International Solid-State Circuits Conference (ISSCC) and the IEEE Asian Solid-State Circuits Conference (ASSCC), advisory board for the IEEE International Symposium on Circuits and Systems (ISCAS), international advisory board for the Institute for Nanodevices and Biosystems, and various U.S., Korea, and Japan industry, government, and academic technical advisory positions on subjects including ultrafast electronics, science and technology at the nanoscale, and the interface between biotechnology and microelectronics. He was a guest editor for the IEEE JOURNAL OF SOLID-STATE CIRCUITS (JSSC, January 2009) and a co-editor of *CMOS Biotechnology* with Springer (2007).

The research focus of his research group at Harvard University is on: 1) RF, analog and mixed-signal ICs; 2) GHz/THz 1-dimensional electronic and plasmonic transport in quantum wires; 3) soliton and nonlinear wave electronics; 4) applications of CMOS ICs in biotechnology.