A fully integrated IQ-receiver for NMR microscopy

Jens Anders *, Paul SanGiorgio, Giovanni Boero

Ecole Polytechnique Fédérale de Lausanne (EPFL), Station 17, CH-1015 Lausanne, Switzerland

Abstract

We present a fully integrated CMOS receiver for micro-magnetic resonance imaging together with a custom-made micro-gradient system. The receiver is designed for an operating frequency of 300 MHz. The chip consists of an on-chip detection coil and tuning capacitor as well as a low-noise amplifier and a quadrature downconversion mixer with corresponding low-frequency amplification stages. The design is realized in a 0.13 μm CMOS technology, it occupies a chip area of 950/800 μm² and it draws 50 mA from a supply voltage of 1.8 V. The achieved time-domain spin sensitivity is 5 × 10¹⁴ spins/√Hz.

Images of phantoms obtained in our custom-made gradient system with 8 μm isotropic resolution are reported.

© 2010 Elsevier Inc. All rights reserved.

1. Introduction

Nuclear magnetic resonance (NMR) based methods are amongst the most powerful analysis techniques in both modern chemistry and medicine. As scaling down the size of the detection coil in principle increases the spin sensitivity of the NMR receiver [1,2], the unprecedented sensitivity of miniaturized coils opens up a new range of applications. In NMR spectroscopy this improved spin sensitivity is used to study mass-limited samples with volumes in the microliter and nanoliter range [3,4]. Miniaturized detection coils are also used to perform high spatial resolution magnetic resonance imaging (MRI) on small objects [5–7]. Here, the improved spin sensitivity ensures reasonable acquisition times even for spatial resolutions in the micrometer range.

As explained in [8], co-integrating the receiver electronics together with the detection coil on the same CMOS chip offers two major advantages for NMR applications: First, the complexity of the off-chip electronics can be drastically reduced, a fact which is particularly important for forming large-scale arrays of many channels. Furthermore, placing the low-noise amplifier (LNA) in close proximity to the detection coil makes the use of a 50 Ω-matching network unnecessary. Since the tuning capacitor can be co-integrated on-chip, very dense arrays with virtually no spacing between the coils become feasible. Due to the aforementioned advantages, the use of integrated-circuit (IC) technologies for the design of both fully integrated NMR probes [8–11] and NMR receiver electronics for hybrid solutions with external detection coils [12–15] has recently attracted attention in the NMR community.

In this paper the first fully integrated CMOS quadrature receiver for NMR spectroscopy and imaging applications is introduced. The receiver is designed to be used as a stand-alone block or as a unit cell to build large-scale dense arrays of NMR receivers. If used as a building block for large-scale arrays, the on-chip frequency downconversion potentially reduces the crosstalk between channels and also facilitates channel multiplexing to reduce the number of output pins. Thanks to an improved RF frontend and the on-chip quadrature detection the spin sensitivity has been improved by a factor of three compared to the work presented in [8]. This spin sensitivity improvement allows for a reduction in measuring time by a factor of about 10.

The paper is organized as follows: First, a detailed description of the on-chip electronics is given in Section 2. In Section 3, the chip is characterized both electrically and using NMR-based methods. In Section 4, we introduce our custom-made gradient system for NMR microscopy and present images obtained using the CMOS receiver in Section 5. The paper concludes with a brief outlook on future work in Section 6.

2. Receiver electronics

In this section, we describe the on-chip electronics of the receiver chip in detail. After a general explanation of the receiver architecture, the details of the integrated-circuit implementation, realized in a 0.13 μm IBM CMOS technology, will be discussed.

2.1. Receiver architecture

The chip architecture is displayed in Fig. 1. It is, for the most part, a standard homodyne architecture as used in many communication systems with the main difference being the lack of explicit...
on-chip filtering stages. This somewhat simplified architecture is possible due to the absence of strong interferers in the relevant NMR spectrum. Furthermore, the tuned LC circuit in the receiver frontend provides some attenuation for very high- and very low-frequency out-of-band signals. As illustrated in the figure, the on-chip electronics consist of an RF frontend followed by a quadrature downconversion mixer and the corresponding baseband electronics. The RF frontend is composed of the on-chip detection coil, which together with the on-chip tuning capacitor forms the tuned LC resonant circuit, and the low-noise amplifier (LNA). The output of the LNA is AC-coupled to two mixers which are driven by two LC resonant circuits, and the low-noise amplifier (LNA). The output which together with the on-chip tuning capacitor forms the tuned electronics. The RF frontend is composed of the on-chip detection coil, which together with the on-chip tuning capacitor forms the tuned LC resonant circuit, and the low-noise amplifier (LNA). The output of the LNA is AC-coupled to two mixers which are driven by two local oscillator signals which are phase-shifted by 90° with respect to each other to form the in-phase (I-) and quadrature channel (Q-channel) of a quadrature downconversion stage. Due to the tight phase noise requirements in NMR spectroscopy necessary to achieve acceptable linewidths, we use an off-chip reference for the local oscillator (LO). The in-phase and quadrature local oscillator signals are generated from this off-chip reference at around four times the Larmor frequency using an on-chip divide-by-four frequency division stage. The external LO signal is applied at four times the Larmor frequency to avoid inductive coupling from the printed circuit board (PCB) lines carrying the LO signal to the detection coil. In previous designs, such a coupling lead to a fairly large DC offset voltage at the mixer output due to self-mixing of the coupled signal. The offset voltage generated by this effect was found to be as large as 50 mV referred to the input of the low-frequency amplification stage, severely degrading its performance. In another test chip this offset was canceled by means of an on-chip digital offset calibration scheme. This method also worked but we experienced problems with the digital electronics losing their calibration after excitation pulses.

The low-frequency amplification stage serves two purposes: First, it further boosts the signal level to make it compatible with standard analog-to-digital converters (ADCs) and, second, it provides the capability to drive the capacitance of the long cable connecting the chip to the anti-aliasing filter (AAF) and the ADC outside the magnet.

### 2.2. Integrated electronics

The detection coil occupies an area of 345 μm × 345 μm. This value was found as a compromise between scaling down the coil size to increase its unitary magnetic field \( B_n \) – improving its intrinsic spin sensitivity – and being large enough such that the entire quadrature receiver can fit underneath the coil to allow the use of the detector as a building block for large-scale dense arrays. In contrast to the technology used in [8], the IBM process used here offers two thick top metal layers for the design of integrated inductors, one 4-μm thick Al layer and a 3-μm thick Cu layer. To fully benefit from the additional design freedom associated with the second metal layer, an optimization tool was written in MATLAB (Mathworks, Natick, USA) to maximize the SNR of the receiver frontend (i.e. detection coil, tuning capacitor and LNA). To this end, we first estimated the minimally achievable input-referred noise of the LNA for a given bias current. A simplified version of the input stage is shown in Fig. 2 together with the corresponding equivalent noise model. From this model we calculated the excess noise factor of the LNA, \( F = 1 \), i.e. the ratio of the noise power added by the amplifier and the noise power of the source conductance, \( G_0 \), to be:

\[
F - 1 = \frac{\gamma_{nd} G_m}{\beta_{nd}} + \frac{2 \delta_{ng}}{G_m} \left( 1 + \frac{1 + \alpha^2}{2} \left( \frac{\gamma_{nd}}{\beta_{nd}} - 2 \frac{\gamma_{nd}}{\beta_{nd}} \right) \right),
\]

where \( \gamma_{nd} \) is the drain thermal noise excess factor, \( \beta_{ng} = \delta_{ng}/5\), and \( \delta_{ng} \) is the gate thermal noise excess factor and \( \alpha \) the slope factor, \( \alpha = |c_{gip}|/c_{gon} \) is the ratio of the PMOS and NMOS gate-to-source capacitances, \( c_p \) is the imaginary part of the correlation coefficient between drain noise and induced gate noise in a MOSFET, \( G_m \) is the gate-to-bulk transconductance of a single MOSFET (it is assumed that the PMOS and NMOS transistors are sized to obtain equal transconductances, i.e. \( \alpha \approx \mu_n \delta / \mu_p \) with \( \mu_n \) and \( \mu_p \) being the mobility of electrons and holes, respectively), and \( G_0 \) is the conductance connected between the gates of the two input pairs. A plot of the noise figure vs. gate transconductance using this model applied to the detection coil and the LNA employed in the chip is shown in Fig. 3. We clearly see that for a gate transconductance above 105 mS, the noise figure would be dominated by gate induced noise, which emphasizes the need to include gate induced noise in our analysis, even at the rather moderate frequency of 300 MHz. More details for the device noise modeling of MOSFET transistors can be found in [16].

The equivalent coil circuit parameters (self inductance, AC resistance, quality factor, etc.) required by the MATLAB optimizer were obtained using the freeware EM simulation tool ATITIC (Analysis and Simulation of Spiral Inductors and Transformers for ICs) which can be easily called in batch mode from a MATLAB script. The actual operating point chosen by the MATLAB optimizer lies at 87 mS. This value was then used as a starting point for further simulations using SpectreRF (Cadence Design Systems Inc., San Jose, USA). Similarly, to obtain more accurate parameters of the detection coil, the optimum coil design obtained by the MATLAB optimizer was resimulated using Agilent Momentum (Agilent Technologies, Santa Clara, USA). The optimum coil resulting from the optimization process described above has the following parameters: five turns in each of the two metal layers, the two metal lay-
ers being connected in series, a conductor width of \(w = 15 \mu m\) and a pitch of \(p = 20 \mu m\).

The tuning capacitor is realized using a dual metal-inter-metal (MIM) capacitor. This type of capacitor not only displays a high intrinsic quality factor \(Q > 150\), but also shows small parasitic capacitances to the substrate since it resides in the very back end of line (BEOL) of the process between the last and the second to the last metal layers. Furthermore, the intrinsic parameter variations of this type of capacitor are relatively small and are well modeled by the foundry, so that the value of the resonant frequency of the input LC circuit can be well centered even in the presence of manufacturing tolerances.

The schematic of the LNA is depicted in Fig. 4. The current-reuse topology of the input stage has already been discussed in detail in [8]. However, in contrast to the design introduced in [8], the LNA presented here is fully differential, i.e. differential input and differential output, significantly reducing the complexity of the output stage compared to [8]. In that chip, the differential-to-single-ended conversion was necessary to reduce the power consumption of the 50 \(\Omega\) output stage by a factor of four per channel compared to a fully differential 50 \(\Omega\)-output stage. In this design which incorporates a co-integrated downconversion stage, the fully differential topology aids in suppressing any common-mode interferers and noise which might be present on the supply lines.

The second major change compared to the previous design is the use of minimum length transistors \((L_{\text{min}} = 0.13 \mu m)\) in the input differential pairs. This change reduces the parasitic capacitances associated with the input devices, which significantly lowers the loading of the input LC circuit. Furthermore, the reduction in parasitic capacitance associated with the input devices enhances the bandwidth to more than 800 MHz. The bandwidth enhancement reflects the fact that due to the large size of the input devices the dominant pole of the LNA is located at the drains of the input differential pairs. The final significant change compared to the design in [8] is the common-mode feedback (CMFB) loop which has been added to ensure proper functionality of the design in the presence of process variations in the chip manufacturing process. Devices which are part of the CMFB loop are displayed in gray. This measure became necessary because statistical and corner simulations revealed that the variations in the threshold voltage due to tolerances in the manufacturing parameters could cause the input differential pairs to leave the saturation region and thus display a drastically degraded transconductance over bias current ratio, \(g_m/I_d\), which directly effects the noise performance of the LNA. This effect is much more pronounced in the chip presented here than in the previous design presented in [8] due to the reduced nominal supply voltage of 1.5 V compared to 3.3 V for the previous chip.

The mixer is a Johnson-type mixer employing a current-reuse technique in the input differential pair to improve the noise performance for a given bias current. In contrast to the LNA, no CMFB loop is necessary here because the bias currents are significantly smaller which renders the design more immune against process variations. The size of the switching transistors is optimized for low flicker noise to allow for a homodyne downconversion of the NMR signal. An RC-filter at the output suppresses both the sum frequency and limits the noise bandwidths.

The low-frequency amplification stage consists of a fully differential, buffered Miller operational amplifier (OP-Amp) with corresponding CMFB and feedback resistors. The output driving stage is
realized by a differential source follower. The driving capability of the output stage was chosen to allow for driving of some hundred picofarads corresponding to several meters of cable connecting the chip inside the magnet to the ADC in the control room.

Although nominally designed for a supply voltage of 1.5 V, it was found that the supply had to be increased to 1.8 V to achieve an optimal performance in terms of gain and input-referred noise. This behavior was also seen in simulations and can be attributed to the produced chip being close to a process corner.

3. Electrical characterization and μ NMR spectroscopy

In this section we characterize the chip both electrically and using NMR-based methods.

3.1. Electrical characterization

To verify that the input LC circuit is tuned to the right operating frequency of 300 MHz, a normalized gain characteristic of both the I- and Q-channels was measured. The setup used for this measurement is illustrated in Fig. 5. A small coil which is broadband 50-O-terminated is connected to a frequency generator (Anritsu MG3633A) to induce a voltage in the on-chip coil. The frequency of this oscillator is set to a value of \( f_{RF} + \Delta f \). The local oscillator for the on-chip mixer at a frequency of 4\( f_{RF} \) is provided by a second signal generator (Rohde & Schwarz SMR20). The gain vs. frequency characteristic is measured by sweeping \( f_{RF} \) in the frequency range from 50 MHz to 400 MHz and recording the chip’s output signal at the difference frequency \( \Delta f \). The resulting normalized gain curves for both I- and Q-channels is shown in Fig. 6. From the figure we see that the input LC circuit is properly tuned around 300 MHz. Furthermore, the peak height of Fig. 6 suggests a quality factor of the detection coil around 3.5. This value seems surprisingly low given the fact that modern CMOS technologies with their thick top metal layers can produce coils with quality factors around 10 even at the moderate operating frequency of 300 MHz. However, the coil is not optimized for the highest possible quality factor but for the highest possible receiver SNR including the added noise from the LNA. Our simulations (performed using the MATLAB optimizer described in Section 2) indicate that it is beneficial to trade off coil quality factor for increased induced voltage and reduced LNA noise figure by increasing the number of turns. Simulations using Agilent Momentum show that the resulting optimum design is supposed to display a quality factor of 6. The discrepancy between the measured value and simulations can be attributed to the receiver circuitry underneath the detection coil probably degrading its quality factor. It is also interesting to note that the 3 dB bandwidth of the peak in Fig. 6 corresponds to the simulated value of the quality factor of six rather than a value of 3.5 as suggested by the peak value. The measured maximum phase imbalance between the I- and Q-channels is below 5° corresponding to an image frequency suppression of about 25 dB and a noise figure degradation of less than 0.1 dB.

3.2. NMR spectroscopy

The setup used to perform the NMR spectroscopy experiments is illustrated schematically in Fig. 7. The probe consists of three PCBs made of FR4 and a silicon heatsink. As illustrated in Fig. 7, the chip is mounted on a silicon heatsink using a thermally con-
ductive glue. The heatsink itself is mounted on the backside of a PCB containing a cutout for the chip. The chip is then wire bonded to the PCB by means of aluminum wedge bonding. The differential LO signal is supplied to the chip by two 50-Ω striplines. The two ground planes of the stripline are placed on the backside of PCB 1 and PCB 3, respectively. PCB 2 merely serves the function of a spacer PCB ensuring the same distance between the stripline conductor and the two ground planes. A 50 Ω resistor, which is AC-coupled to both sides, is placed as a termination at the end of the stripline close to the chip to avoid reflections. Apart from being used as a ground plane for the stripline, PCB 3 also contains the excitation coil with the corresponding tuning and matching network. The excitation coil is mounted orthogonally with respect to the on-chip detection coil to avoid direct inductive coupling during the excitation pulse between the two. The size and shape of the coil is chosen to obtain a $B_1$-homogeneity over the sensitive volume of the coil better than 1%.

To estimate the time-domain and frequency-domain spin sensitivities according to the definitions introduced in [8], we used a polydimethylsilane (PDMS) sample with a size of approximately $(300 \mu\text{m})^3$. The measured spin sensitivity in the time-domain is $5 \times 10^{14}$ spins/√Hz. The spin sensitivity in the frequency domain, evaluated using an optimal time-domain filter with a time constant of 1.5 ms corresponding to the $T_2^*$ value of the sample, is $1.5 \times 10^{15}$ spins/√Hz. The real part of the spectra of the PDMS sample is shown in Fig. 8.

To evaluate if the achievable spectral resolution is sufficient for micro-imaging applications, a capillary filled with pure water (capillary dimensions: inner diameter ID = 320 μm, outer diameter OD = 450 μm) was placed above the detection coil. The corresponding real part of the FFT is displayed in Fig. 9. The achieved full-width half maximum (FWHM) linewidth is 50 Hz. A close-up of the FFT indicating the linewidth more clearly is shown as an inset of Fig. 9.

4. Micro-gradient system

To use the receivers for imaging, we have built a custom micro-gradient system based on the design of Weiger et al. [17]. Copper wires (outer diameter 0.4 mm) are pressed into two polyoxymethylene (POM) and aluminum supports on either side of the support PCBs in a bi-planar configuration. A schematic of the coils on both planes is shown in Fig. 10. The inter-plane spacing is 4.9 mm and the strength of the gradients produced in each of the three axes is $(G_x, G_y, G_z) = (96, 62, 67) \text{ mT/(m A)}$. The gradients are within 5% linearity in a region of approximately $1.2 \times 1.9 \times 1.8 \text{ mm}^3$. The three coils are powered by three amplifiers (AE Techron 7792), capable of driving the coils with trapezoidal current pulses with an amplitude of 100 A and a rise time of approximately 50 μs. The coils are thermally coupled to a water–cooled copper support which provides adequate cooling to run a current of 100 A through all three coils at a 1% duty cycle. At lower currents, they can be cycled much more quickly, but this is not necessary for the pulse sequence employed in this work.

5. NMR microscopy using the integrated receiver

Images were taken using a simple constant-time-imaging (CTI) sequence similar to [18], and as suggested by Choi et al. [19] as being particularly well-suited to micro-imaging – the short and i-
tense pulses are easily produced by single-turn, low-impedance
coops and the lack of a read gradient keeps the duty cycle of the gra-
dients light. A schematic of the pulse sequence used for imaging is
presented in Fig. 11. A non-selective RF pulse (\(T_p = 10 \times 10^{-6}\) s) ex-
cites the entire sample, after which a phase-encoding gradient on
all three axes is applied (\(T_{grad} = 400 \times 10^{-6}\) s). After the pulse,
data is acquired (\(T_{acq} = 40\) ms) and then the sequence is repeated
(\(T_R = 75 \times 200\) ms). Using the largest possible gradients (i.e. about
10 T/m), the theoretically achievable resolution is roughly
1.2 \times 1.9 \times 1.8 \mu m^3. As our current system does not include RF
pulse shaping that would allow for slice selection, taking a full 3-
D scan over the sensitive region of the microcoil at the maximum
resolution would take more than 2 weeks and is thus implausible.

Preliminary scans were thus taken at more modest resolutions to
keep total imaging time to less than a couple of days.

The sample chosen for imaging was a micro-capillary (outer
diameter 450 \(\mu m\), inner diameter 320 \(\mu m\)) filled with a 40 mM
solution of CuSO\(_4\) in water and polystyrene beads with a diameter
of 75 \(\mu m\) (inset of Fig. 11). As epoxy protecting the bonding wires
prevented the capillary from being placed in direct contact with
the chip, it had to be placed with its bottom edge approximately
250 \(\mu m\) above the chip surface, as shown in Fig. 12. The average
coil sensitivity at this distance relative to the average sensitivity
if the capillary were in contact with the chip is roughly 6 – i.e.
the induced signal would be 6-times larger (and the imaging time
36-times shorter) if the capillary were in contact. Sample contain-
ers having lateral dimensions smaller than 650 \(\mu m\) \(\times\) 520 \(\mu m\)
can be placed in close contact with the detection coil already in the
present design. Future designs of the chip will have altered bond-
ing layouts with the bonding pads being placed in two columns on
each side of the receiver. This will allow for a close contact also for
sample containers with large dimensions in the \(B_0\) direction, e.g.
capillaries. Scans were taken with an isotropic resolution of
(8 \(\mu m\))^3, a scan range of 400 \(\times\) 400 \(\times\) 600 \(\mu m^3\), and 20 averages
per point which took a total scan time of 83 h. At close contact,
only one average per point would be needed to produce a better
signal-to-noise ratio and the total time could be reduced to
250 min. One slice in the \(X\)-plane and two slices in the \(Z\)-plane
are shown in Fig. 13: (a) shows a slice of the capillary parallel to
the chip surface in which a single bead is present; (b) shows the
cross-section of the capillary also showing the same bead; and
(c) is a different cross-section slice, where no beads are present.
The circles and lines drawn on the figure represent the nominal
capillary and sphere diameters. Although the calculated spin den-
sities in the images have been adjusted to account for the spatial
dependence of the sensitivity of the coil, some variation still exists
which might be accounted for by an improperly shimmed \(B_0\)-field.

6. Outlook

The ultimate goal of our work is the realization of an imaging
system capable of performing parallel imaging with a field of view
of several millimeters and an isotropic spatial resolution down to a
To be used in such large arrays, the power consumption of a single receiver has to be reduced to avoid excessive self-heating of the chip. The two best candidates for power reduction are the RF front-end and the baseband amplification stage. The baseband amplification stage can easily be optimized by splitting the combined amplification buffering stage with differential outputs into a fully differential amplification stage and a dedicated single-ended buffer stage capable of driving the large load capacitance more power efficiently, e.g. using a class AB topology. Preliminary simulations show that in this way, the power consumption can be reduced from about 25 mA in the current design to below 10 mA. To achieve the aim of further reducing the power consumption in the LNA we performed an exhaustive search of the entire design space. The result of these simulations is that we can decrease the current in the LNA to values around 5 mA (the current design consumes about 12 mA) without significantly degrading the receiver SNR. Note that for different values of bias current different optimal coil designs are found by the optimizer, so it would not be enough to simply optimize the power consumption of the current LNA for the given detection coil to reach this goal. Consequently, we expect to be able to reduce the power consumption of a single reception channel from about 75 mW in the current design to about 30 mW with negligible degradations in SNR. Thermal simulations indicate that with such a level of power consumption per channel the surface of theforeseen 6 × 6 single-chip array can be easily kept at room temperature.

Acknowledgment

This work is supported by the European Union under contract no. 028533.

References


Fig. 13. MRI of a micro-capillary filled with CuSO4 doped water and a 75 μm-poly styrene bead. The isotropic resolution is (8 μm)3 and the field of view is 400 × 400 × 600 μm3. The microcoil is located in the y–z plane at x = 0. (a) Slice parallel to the X-plane showing the outline of the capillary and the bead. (b) A slice perpendicular to the capillary showing the cross-section and the same bead. (c) A different cross-sectional slice without the bead.