
http://dx.doi.org/10.1109/jssc.2013.2274895

Institute of Electrical and Electronics Engineers (IEEE)

Author's final manuscript

Thu Feb 14 04:46:07 EST 2019

http://hdl.handle.net/1721.1/93893

Creative Commons Attribution-Noncommercial-Share Alike

http://creativecommons.org/licenses/by-nc-sa/4.0/
Ultrasonic Imaging Transceiver Design for CMUT: A Three-Level 30-Vpp Pulse-Shaping Pulser With Improved Efficiency and a Noise-Optimized Receiver

Kailiang Chen, Member, IEEE, Hae-Seung Lee, Fellow, IEEE, Anantha P. Chandrakasan, Fellow, IEEE, and Charles G. Sodini, Fellow, IEEE

Abstract—This paper demonstrates a four-channel transceiver chip for medical ultrasonic imaging, interfacing to the capacitive micromachined ultrasonic transducers (CMUTs). The high-voltage transmitter (Tx) uses a three-level pulse-shaping technique with charge recycling to improve the power efficiency. The design requires minimum off-chip components and is scalable for more channels. The receiver is implemented with a transimpedance amplifier (TIA) topology and is optimized for tradeoffs between noise, bandwidth, and power dissipation. The test chip is characterized with both acoustic and electrical measurements. Comparing the three-level pulser against traditional two-level pulsers, the measured Tx efficiency shows 56%, 50%, and 43% more acoustic power delivery with the same total power dissipation at 2.5, 3.3, and 5.0 MHz, respectively. The CMUT receiver achieves the lowest noise efficiency factor compared with that of the literature (2.1 compared to a previously reported lowest of 3.6, in units of mPAM/√mW/Hz). In addition, the transceiver chip is tested as a complete system for medical ultrasound imaging applications, in experiments including Tx beamformation, pulse-echo channel response characterization, and ultrasonic Doppler flow rate detection.

Index Terms—Capacitive micromachined ultrasonic transducer (CMUT), charge recycling, multilevel pulse shaping, noise efficiency factor (NEF), pulse-echo response, transimpedance amplifier (TIA), transmitter (Tx) efficiency, Tx beamformation, ultrasonic Doppler flow rate detection, ultrasonic transceiver.

I. INTRODUCTION

ULTRASONIC imaging is an important modality for medical diagnosis. Compared with other imaging modalities, ultrasound is relatively low cost and harmless to human health and has decent resolution. Modern ultrasonic imaging systems are becoming increasingly complex and powerful, yet compact, benefiting from Moore’s law [1]. Laptop-size ultrasound systems have gained comparable performance to the traditional cart-size machines; handheld devices, such as the GE Vscan [2], indicates the trend toward highly integrated ultrasonic imaging solutions to enable portable or even wearable ultrasound applications in hospital and at home.

Piezoelectric transducers (PZTs) have dominated the ultrasonic transducer technology. However, capacitive micromachined ultrasonic transducers (CMUTs) recently emerged as an alternative technology for better system integration [3]. The CMUT technology offers advantages such as improved bandwidth, ease of fabricating large arrays, and potential for integration with electronics with the through-silicon vias (TSVs) [4]–[6] or monolithic CMUT-CMOS integration [7]–[9]. Interfacing to CMUTs is different from interfacing to PZTs, because the effective load is much more capacitive and has higher impedance. The primary reason for the large parasitic capacitance is the physical structure of the CMUT element, which forms a parallel-plate capacitor [10]. As a result, the transmitter and receiver circuitry need to be designed appropriately to prevent excessive performance degradation caused by the large capacitive load.

For the transmitter, high-voltage linear amplifiers are commonly used to drive the PZT loads to achieve good linearity and acceptable efficiency [11], [12]. To drive a CMUT load, however, linear amplifiers are not optimum. In addition to the amplifier power consumption, a considerable power loss becomes associated with charging and discharging the parasitic capacitance of the CMUT element [10], degrading the overall power efficiency of the transmitter stage. Furthermore, the linearity of the amplifier does not translate to good linearity performance of the transmitter stage, because the CMUT element distorts the amplifier’s output waveform through the nonlinear relationship between the electrical input signal and electrostatic force acting on the element’s membrane [13]. Resonant transmitters with inductors to cancel out the loading capacitances could boost the power efficiency [14]. However, bulky off-chip inductors of several micro-Henries are needed for every transmitting channel, to work with typical loads of 10–200 pF per channel at the ultrasound operating frequency range of 1–20 MHz [15]–[17], which is undesirable for compact integration. Alternatively, the multilevel pulsing technique, which was initially introduced for chip-to-chip interconnects [18], can be applied to reduce the power consumption on the capacitive load. Multilevel techniques have been used in PZT ultrasonic drivers for pulse-shaping and harmonic suppression [15]–[17], [19]. However, the power efficiency was not improved because charge recycling was not implemented between the multiple voltage levels. This work [20] presents the advantage of the multilevel pulsing with charge recycling to improve...
the combined power efficiency of the CMUT transducer and transmitter.

For the receiver, large input capacitance limits the bandwidth and tends to increase the noise contribution from the input stage transistors, degrading the noise figure. Bulky off-chip inductors are needed to impedance match the source to a traditional PZT pre-amplifier that assumes a low-impedance source [14]. Charge-based amplifiers were attempted for CMUTs. The continuous-time charge amplifier achieved low-noise and low-power performance for CMUT working at kHz range [21], but the large impedance from the dc-setting network limits the bandwidth for a CMUT array working at MHz range for medical imaging applications. The switched-capacitor charge integrating amplifier in [22] could provide enough bandwidth, but issues such as clock feed-through and charge injection are difficult to mitigate for the inherently single-ended CMUT signal path. Moreover, because the sampling clock switches at a higher frequency than input signal bandwidth, the settling requirement for the op-amp demands a higher bandwidth than what is needed in op-amps used as continuous-time buffers, leading to much more power consumption. In this work [20], the transimpedance amplifier (TIA) topology [4], [8], [23] is used to improve the tradeoff between gain, bandwidth and noise, with an inductor-less design at the presence of high input capacitance.

This paper is organized as follows. Section II provides a high-level description of the proposed work. The design methodologies and implementations of the transmitter and the receiver are presented in Sections III and IV, respectively. Section V details the experimental results. The transceiver circuit blocks are characterized and the chip is tested as part of the ultrasonic imaging system for medical applications. Finally, conclusions are drawn in Section VI.

II. HIGH-LEVEL DESCRIPTION OF THE ULTRASONIC IMAGING TRANSCIEVER CIRCUITS

The block diagram of the ultrasonic imaging transceiver is shown in Fig. 1. All CMUT elements in the array are dc-biased with the shared RC network provided off-chip. Resistor $R_k$ and capacitor $C_k$ filter out noise from the high-voltage supply and provide an ac ground for the transducer. The dc bias voltage $V_b$ applied on the CMUT is between 50–85 V. The transceiver chip includes four channels. In each channel, a 30-Vpp high-voltage pulser in the transmitter (Tx) path drives the ultrasound transducer to emit acoustic energy. The emitted ultrasonic wave travels through the medium and is reflected whenever it hits medium boundaries with mechanical impedance mismatch. The reflected echoes are transformed by the CMUT element into a weak electrical signal. A low-noise amplifier (LNA) in the receiver (Rx) path amplifies the weak signal to the output. During transmission, the Rx switch is turned off to prevent the high-voltage transients from breaking the LNA implemented with low-voltage transistors.

After collecting multiple channels’ outputs from one or several transmissions, ultrasonic images can be generated. Medical ultrasound systems use beamformation to improve the image quality. Tx beamformation is realized by controlling and applying different delays across the four Tx channels. Similarly, the received signals are digitized and processed by an off-chip Rx beamformer.

III. TRANSMITTER DESIGN

A. Multi-level Pulsing for Higher Efficiency

The small-signal model for a CMUT element is represented by a capacitor and resistor in parallel, as shown in Fig. 2(a). The capacitor $C$ is the parallel-plate capacitance between the CMUT element’s membrane and the common node. The resistor $R$ is the medium’s mechanical load at the CMUT surface, transformed to the electrical port [10]. The power dissipated by $R$, due to the electrical pulse’s fundamental frequency component, models the useful acoustic power delivered into the medium. The power dissipated while charging and discharging $C$ (dynamic power) does not contribute to the acoustic output and thus is wasted. The CMUT transducer used in this work is a 1-D array. Each CMUT element has a size of 300 $\mu$m x 3000 $\mu$m and is modeled as 40 pF [80 kΩ] [10].

The Tx efficiency is defined as the ratio between the useful acoustic power and the total power dissipated. It models the
combined efficiency of CMUT and the ultrasonic pulser together by capturing both the power loss in the pulser circuitry and the dynamic power dissipated by charging and discharging the CMUT parasitic capacitance. As discussed in Section I, resonant drivers can be used to recycle the dynamic power and increase efficiency, but off-chip inductors for each channel are needed due to the low operating frequencies. Alternatively, the multilevel pulse-shaping technique with charge recycling reduces dynamic power to increase efficiency [18]. It also requires the least off-chip components, as will be seen in Section III-B.

To show how multilevel pulse-shaping increases Tx efficiency, first assume that conventional two-level square-wave pulses are used to drive a 40 pF [80 kΩ load, as shown in Fig. 2(b). The pulse magnitude is 30 Vpp at a frequency of 3.3 MHz. The amplitude of the fundamental frequency component is calculated to be 19.1 V, or 13.5 Vrms. Therefore, the power dissipated on the 80-kΩ resistor, i.e., the transmitted ultrasonic power at fundamental frequency, is 2.28 mW. Meanwhile, the dynamic power wasted on charging and discharging the capacitor C is calculated to be: $CV^2f = 120$ mW.

An N-level pulser, using $(N - 1)$ regulated voltage sources to charge and discharge the capacitor in a stepwise fashion, reduces the wasted dynamic power to $CV^2f/(N - 1)$ [18]. The power saving comes from the charge recycling mechanism during the discharge operation, which is enabled by the regulated voltage supplies. Instead of discarding all the capacitor charge $CV$ to ground as in the square-wave case, a charge packet of $CV/(N - 1)$ is recycled back to the power supply when the capacitor is switched from one voltage level to the next lower one. As many as $(N - 2)$ charge packets of $CV/(N - 1)$ are recycled until the last packet is dumped to ground. As a result, the dynamic power is reduced by a factor of $(N - 1)$. At the same time, the magnitude of the fundamental component is only decreased slightly, leading to overall efficiency improvement. For example, Fig. 2(c) shows three-level pulses with 20-ns middle voltage level steps, out of a 300-ns period. Its fundamental frequency component amplitude is 18.7 V, or 13.2 Vrms. The useful power delivered is 2.18 mW and the dynamic power is $CV^2f/2 = 60$ mW. A comparison with the square-wave example reveals theoretically a 49% total power saving with only a 4.4% acoustic power reduction or, equivalently, 88% more acoustic output power given the same total power dissipation. Measurement results in Section V-A show that this pulse shape achieves maximum efficiency at 3.3 MHz, although with less improvement than the theoretical calculation.

B. Three-Level Pulser Circuit Design

The three-level pulser is implemented in this work, as shown in Fig. 3. The three pulse voltage levels are 30 (HVDD), 15, and 0 V (GND). The 15-V middle voltage is generated from a 2:1 parallel-series switched-capacitor dc–dc converter (M1–M4), which is shared between channels. The only off-chip components are two 0.1-μF capacitors. Because of the charge recycling nature of the proposed three-level pulser and the fact that the CMUT load (roughly 40 pF per channel) is much smaller than 0.1 μF, the converter can operate at a very low frequency (10–100 Hz) to save power, consuming less than 1% of the total four-channel pulsing power.

Three-level pulse-shaping is implemented with four high-voltage switches (M5–M8) in each channel. NMOS M5 and M6 are used for the transitions of $VDD\rightarrow 0$ V and $0\rightarrow 15$ V, respectively, while PMOS M7 and M8 are used for the transitions of $30\rightarrow 15$ V and $15\rightarrow 30$ V, respectively. The on-resistance of each transistor and the CMUT capacitance form an $RC$ time constant that determines pulse voltage level settling. The transistors are sized to be sufficiently wide to keep the settling time within 15 ns, which is less than 1/20 of the pulse cycle. The relative timing differences between each channel’s control signals is digitally adjustable and effectively implements the Tx beamforming. Low-voltage control signals are supplied off-chip from an FPGA running at 100 MHz. Because the signal swing is small, the digital control power is negligible compared to the pulser power. The cross-coupled level shifters in Fig. 4 translates the low-swing signals into high-swing signals that drive gates of the high-voltage transistors in the pulser and the dc–dc converter. The threshold voltage of M1 and M2 is sufficiently low such that they can be completely turned on by the 3.3-V inverters. The level-shifted gate drive signals have a 30-V voltage swing, which is under the rated operation conditions of high-voltage transistors in this process.
This design of multichannel pulser with a shared voltage converter can be extended easily to more Tx channels without additional off-chip components. It could also be revised to implement more voltage levels to achieve more dynamic power reduction. However, this requires the addition of more switches connected between the CMUT and the voltage levels. Due to the large drain capacitance of high-voltage switches, the self-loading effect takes away much of the power savings from introducing additional voltage levels. According to simulation results of the CMOS process used in this work, a three-level pulser dissipates 16% of total power to drive the gate and drain capacitance of M5–M8 in Fig. 3. For a four-level pulser, the dynamic power reduction is counteracted by the power increase to drive more and bigger transistors, leaving the overall efficiency roughly the same as a three-level pulser. A five-level pulser incurs even more power penalty on driving the high-voltage transistors and the efficiency is lower than a three-level pulser.

IV. RECEIVER DESIGN

A. LNA Optimization Methodology

The transimpedance amplifier topology is chosen to buffer the weak signal out of the high-impedance, highly capacitive CMUT element. Fig. 5 shows the small-signal model of the CMUT and LNA. Figs. 6 and 7 plot various circuit transfer functions to help analyze the optimization process for the LNA.

The closed-loop TIA gain is expressed as

$$Z_{CL} = \frac{1}{\frac{1}{1 + sR_fC_f} \cdot \frac{F \cdot A_{OL}}{1 + F \cdot A_{OL}}}$$

(1)

where $F = Z_i/(Z_i + Z_f)$ is the feedback factor, and $A_{OL}$ is the op-amp open-loop gain. From (1), the LNA dc transimpedance gain is $R_f$ and its bandwidth is determined by the smaller of the following two poles:

$$f_p = \frac{1}{2\pi R_fC_f}$$

(2)

$$f_z \approx \sqrt{f_c \cdot f_p} - \sqrt{\frac{1}{2\pi R_f R_u (C_i + C_f)}}$$

(3)

$f_p$ in (2) is due to the $RC$ time constant of the second multiplying term in (1). $f_z$ in (3) is the intersection between $1/F$ and $A_{OL}$ curves, which is approximately the geometric mean of $1/F$’s zero ($f_z$) and the op-amp’s unity-gain frequency ($f_u$), assuming a 20-dB/dec slope in both $1/F$ and $A_{OL}$ curves.

When $f_z < f_p$, as shown in Fig. 7(a), an increase in $R_f$ always improves LNA’s gain–bandwidth product (GBP). This is because gain $= R_f$, while bandwidth $= f_z$, which is approximately proportional to $1/\sqrt{R_f}$ as indicated by (3). GBP improves roughly proportionally with $\sqrt{R_f}$. However, because $f_p$ is proportional to $1/R_f$, as indicated by (2), the increase in $R_f$ leads to faster decrease in value of $f_p$ than $f_z$. When $f_z = f_p$, the LNA achieves maximum GBP available from the op-amp. The phase margin is roughly 45°. Further increase in $R_f$ no longer improves GBP, because the bandwidth becomes limited by $f_p$ and is proportional to $1/R_f$ [Fig. 7(b)], holding the GBP constant. However, as $R_f$ increases, the phase margin continues to improve at the expense of a reduced bandwidth [23].

The optimality condition $f_z = f_p$ also minimizes noise contribution from the op-amp input-referred voltage noise. Fig. 5 shows all noise sources in the circuit. The noise figure is expressed as

$$NF = 1 + \frac{R_i}{R_f} + \frac{\tilde{V}_{op}^2}{\tilde{I}_{op}^2 |Z_i| Z_f^2} + \frac{\tilde{I}_{op}^2}{\tilde{I}_{in}^2} + \frac{2 |V_{op}| \cdot \tilde{I}_{op}}{\tilde{I}_{in}^2 |Z_i| Z_f^2}$$

(4)

From (4), a large $R_f$ is desired to reduce its thermal noise contribution. Moreover, the op-amp’s input-referred voltage noise ($\tilde{V}_{op}$) has a peaking effect due to the impedance drop in $Z_i$ at higher frequencies. It can be mathematically seen from the...
following noise gain expression \( G_n \), defined as the transfer function from LNA input \( (V_{\text{op}}) \) to the output \( (V_{\text{out}}) \):

\[
G_n = \frac{A_{\text{OL}}}{1 + f_r \cdot A_{\text{OL}}} = \left| \frac{1}{f_r} \right| A_{\text{OL}} \approx \min \left( \left| \frac{1}{f_r} \right| A_{\text{OL}} \right). \tag{5}
\]

The dashed line in Fig. 6 shows the graphical interpretation of (5): \( G_n \) is the lower part of \( 1/f_r \) and \( A_{\text{OL}} \) curves, which has a considerable peaking effect within the LNA bandwidth. By comparing the optimal and nonoptimal conditions in Figs. 6 and 7, one can see that the condition \( f_r = f_p \) minimizes the noise peaking effect while exploiting the maximum possible GBP from the op-amp design.

B. LNA Transistor-Level Implementation

Following the guidelines discussed in Section IV-A, the LNA optimization starts with a 5-MHz bandwidth target and the optimal condition: \( f_r \approx f_p \approx BW \). \( R_f \) is maximized while keeping the corresponding \( C_f \), estimated from (2), larger than parasitic capacitances to maintain control over circuit stability. The unity-gain frequency \( f_r \) is estimated from (3) to set the op-amp design target. Further design adjustments keep phase margin above 60°.

Fig. 8 shows the LNA schematic. The input stage devices (M1, M2) are biased at the boundary of strong and weak inversion to achieve high transconductance per unit current and low noise while minimizing size and parasitic capacitance. The differential pair suppresses noise from the power supply and the ground, which is not possible with single-ended topologies [4], [8]. The Miller compensation leg (M9, Cc) keeps the op-amp second pole well beyond the closed-loop bandwidth for good phase margin. The source follower (M7, M8) lowers the op-amp output impedance to enforce accurate feedback.

During high-voltage transmissions, the high-voltage Rx switch (M10) is opened and the low-voltage switches (S1–S6) are closed for protection. The on-resistance of M10 directly impacts LNA noise performance. Its size is chosen such that its noise contribution is only a small portion of the input stage, and its parasitic drain and source capacitance do not degrade phase margin and bandwidth. S1–S6 also implement the op-amp’s sleep mode, during which only the reference current remains conducting for fast wake-up within 1 μs. The sleep mode enables system-level power saving opportunities.

V. EXPERIMENTAL RESULTS

A. Acoustic Power and Tx Efficiency Measurement

1) Measuring Acoustic Output Power: To obtain Tx efficiency, the total pulsing power can be measured electrically. However, the ultrasonic power transmitted into the medium requires acoustic measurements. From

\[
P_{\text{acoustic}} = I \cdot A - \frac{p_{\text{rms}}^2}{Z_m} \cdot A \tag{6}
\]

acoustic power is the product of the acoustic intensity \( I \) at transducer surface and the transducer surface area \( A \). \( I \) is calculated from the rms fundamental frequency component of the acoustic pressure at the transducer surface \( p_{\text{rms}} \) and the acoustic impedance of the medium \( Z_m \).

In practice, the acoustic pressure at the transducer surface cannot be directly measured. Instead, it can be reliably back-calculated from a pressure measurement at another location. According to [24]–[26], when the transducer aperture is close to a square or a circle, the pressure magnitude profile along the axial direction reaches its maximum at the boundary between the near and far fields. The maximum magnitude is roughly twice the pressure magnitude at the transducer surface.

For the back-calculation, an acoustic pressure measurement system is established in the laboratory, as shown in Fig. 9. The CMUT array is submerged in vegetable oil at the bottom of the oil tank. The test chip circuitry is connected to CMUT from under the oil tank. A hydrophone (ONDA HNC0400)
Fig. 9. Photograph of the laboratory setup for measuring the acoustic output power and the Tx efficiency.

Fig. 10. Acoustic output power and Tx efficiency measurement setup.

Fig. 11. Normalized rms pressure along the transducer axial axis: measurement versus simulation. The measurement deviates from the simulation in the near field because the hydrophone tip is too close to the transducer surface, distorting the pressure field.

is mounted on the 3-D translation stage to probe the acoustic pressure magnitude generated by CMUT, in the oil medium.

Fig. 10 shows the detailed configuration to measure the acoustic output power. The four-channel pulser circuitry is parallelized and connected to eight CMUT elements in parallel, in order to form an aperture of 2.4 mm × 3 mm (roughly a square). The solid curve in Fig. 11 is the corresponding acoustic simulation of the pressure field using the Field II software [27], verifying that surface pressure (z = 0 mm) is about half maximum pressure (z = 5.9 mm). Furthermore, the hydrophone is used to probe the acoustic pressure magnitude along the axial direction (z-axis). The measured result in dots in Fig. 11 shows good agreement to both theory and simulation. In the near field, the measured data do not exhibit amplitude fluctuations as predicted by simulation. This is likely caused by the hydrophone tip distorting the pressure field as it approaches the transducer. However, it does not affect the accuracy of the maximum pressure measurement and the surface pressure back-calculation.

2) Measuring Tx Efficiency: Fixing the hydrophone at the near- and far-field boundary (5.9 mm away), acoustic output power is obtained with the aforementioned method. Tx efficiency is thus acquired after dividing the acoustic output power by the total power consumption. Different pulse shapes are generated to evaluate the efficiency improvement. The pulse shape is defined by the Δ/T ratio, as shown in Fig. 12(a), where Δ is the step duration of the middle voltage level and T is the pulse period. When Δ/T = 0, two-level pulses are generated. As Δ/T increases, the pulses turn into three-level, reducing the dynamic power from CV^2f to CV^2f/2 and increasing the efficiency. However, as Δ/T increases further, the acoustic power starts to decrease because less energy is contained within the pulse shape. Since the dynamic power is kept at CV^2f/2, efficiency decreases. Therefore, there is an optimal pulse shape to maximize the Tx efficiency. For example, Fig. 12(b) is a time-domain waveform of optimal three-level pulses at 3.3 MHz.

Fig. 13 shows the measurement results. As an example, Table I(A) compares the optimal three-level pulser against the two-level pulser operating at 3.3 MHz: the optimal three-level pulser dissipates 38% less total power at the cost of delivering 7% less acoustic power. In other words, the three-level pulser outputs 50% more acoustic power at the same power dissipation. The measured improvement is not as big as the theoretical calculation in Section III-A (50% rather than 88%), mainly for two reasons. First, the RC settling transition distorts pulse shapes, with three-level pulses being distorted more severely than two-level pulses, which leads to more acoustic power reduction in a real-world three-level pulser (7% rather than
Fig. 13. Tx efficiency measurement results using different three-level pulse shapes by varying the $\Delta/T$ ratio and at different frequencies.

<table>
<thead>
<tr>
<th>TABLE I</th>
</tr>
</thead>
<tbody>
<tr>
<td>MEASURED PULSER PERFORMANCE SUMMARY AND COMPARISONS. (A) POWER AND EFFICIENCY COMPARISON AT 3.3 MHZ. (B) OPTIMAL THREE-LEVEL PULSER PERFORMANCE SUMMARY</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th></th>
<th>2-level</th>
<th>Optimal 3-level</th>
<th>Change</th>
</tr>
</thead>
<tbody>
<tr>
<td>Acoustic Power</td>
<td>0.56mW</td>
<td>0.52mW</td>
<td>-7%</td>
</tr>
<tr>
<td>Total Power</td>
<td>84.5mW</td>
<td>52.4mW</td>
<td>-38%</td>
</tr>
<tr>
<td>Efficiency</td>
<td>0.66%</td>
<td>1.0%</td>
<td>50%</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Total Power (mW)</th>
<th>2.5MHz</th>
<th>3.3MHz</th>
<th>5.0MHz</th>
</tr>
</thead>
<tbody>
<tr>
<td>Relative Efficiency Improvement Against a 2-level Pulser</td>
<td>56%</td>
<td>50%</td>
<td>43%</td>
</tr>
</tbody>
</table>

4.4%). Second, a three-level pulser uses more high-voltage transistors than a two-level pulser, dissipating more power for driving the transistors’ gate and drain capacitance, which leads to less total power reduction (38% rather than 49%).

The relative efficiency improvements of a three-level pulser over a traditional two-level pulser at 2.5-, 3.3-, and 5.0-MHz pulses are measured to be 56%, 50%, and 43%, respectively. Table I(B) lists the optimal 30-Vpp three-level pulser power dissipation and efficiency at all three measured frequencies. Efficiency improvement is less for pulses with a shorter period, because the same $RC$ settling transition distorts shorter pulse shape more severely, reducing useful acoustic output power. Moreover, higher frequency pulses dissipate proportionally more dynamic power while acoustic output power is kept roughly the same, thus the overall efficiency curve shifts down. Lastly, the optimal $\Delta$ value for the three frequencies is approximately the same (20 ns), which is slightly more than the $RC$ settling time. This is because the optimal pulses use just enough time to settle to the middle level to achieve $CV^2 f/2$ dynamic power, while keeping the middle level as narrow as possible to maintain large fundamental frequency pulse energy delivery. When normalized over pulse period $T$ in Fig. 13, the optimal $\Delta/T$ ratios become different for different pulse frequencies.

By comparing the optimal three-level pulser against the two-level pulser, this work is effectively compared against a range of traditional pulisers. The reason is that, not only for two-level pulisers [4], [10], [29], but also for multilevel pulisers without charge recycling [15]–[17] or pulisers implemented as linear amplifiers [11], [12], the dynamic power dissipation is always $CV^2 f$. Therefore, these traditional pulisers have similar (if not worse, considering the quiescent power dissipation in linear amplifiers) Tx efficiency performance as the two-level pulser used in this work.

B. Tx Beam-Steering Experiment

The ultrasonic lateral beam pattern can be measured based on slight modification of the experimental setup in Fig. 9. In this experiment, each of the four-channel pulisers is connected to one of four consecutive CMUT elements; each pulser drives its CMUT with the 3.3-MHz optimal three-level pulses. The hydrophone is placed at a fixed depth in the transducer’s far field ($z = 7.4$ mm). By moving the hydrophone along the lateral direction ($x$-axis) and collecting the acoustic pressure readings, the lateral beam profile can be plotted. Furthermore, ultrasonic Tx beam-steering is demonstrated on the four-channel transmitter system when varying the relative pulsing delays across four channels.

Fig. 14(a) shows the measured beam profile in dots with zero delay between channels. The beam is steered to the center, i.e., broadside. Fig. 14(b) shows the profile when 30-ns delay is applied between channels. The figures also show the Field II simulation results of the same experimental configurations for each case. The simulation and measured data match well. Hand calculations based on classical wave propagation provide another verification for Fig. 14(b). The beam lateral displacement $\Delta x$ and the channel delay, $\tau_d = 30$ ns, are related to each other by

$$\frac{\Delta x}{z} \approx \frac{\tau_d \cdot c}{d}$$

where depth $z = 7.4$ mm, sound speed $c$ in vegetable oil is measured to be 1460 m/s, and CMUT element pitch $d = 300 \mu$m. The calculated beam lateral displacement $\Delta x = 1.08$ mm, which is consistent with the measured result.

C. LNA Characterization

The LNA is first tested as a single amplifier block, and Table II summarizes its performance. The LNA sleep power is measured when all LNA biasing currents are kept on to achieve fast wake-up time. In Table III, the LNA is compared against other CMUT LNAs in the literature.

Being used for different medical ultrasound applications, the CMUT arrays are very different in size, impedance and operating frequency. Thus, the corresponding LNA specs are also vastly different and difficult to compare. To establish a figure of merit for fair comparison and to be able to apply the data available in CMUT LNA literature, the noise efficiency factor (NEF) commonly used for instrumentation amplifiers [30] is revised for use in this paper. The revised NEF’ is expressed in

$$\text{NEF'} = \frac{V_{\text{rms}}}{\sqrt{B \cdot W}} \cdot \sqrt{\frac{P_{\text{tot}}}{P_{\text{in}}}}$$

where $V_{\text{rms}}$ is the input-referred rms noise amplitude in-band and $V_{\text{rms}}$, $V_{\text{rms}}$ is the input-referred noise spectral density averaged inside the passband. Note that both $V_{\text{rms}}$ and $V_{\text{rms}}$ are acoustic pressure noise, input-referred all of the way to the charge recycling [15]–[17] or pulisers implemented as linear amplifiers [11], [12], the dynamic power dissipation is always $CV^2 f$. Therefore, these traditional pulisers have similar (if not worse, considering the quiescent power dissipation in linear amplifiers) Tx efficiency performance as the two-level pulser used in this work.

2 For example, the CMUT used in this work is designed as an alternative to 1-D PZT linear arrays operating at 2–5 MHz. The LNA bandwidth target of DC-5 MHz is set accordingly.
TABLE II
MEASURED LNA PERFORMANCE SUMMARY

<table>
<thead>
<tr>
<th>LNA Specs</th>
<th>Measured Result</th>
</tr>
</thead>
<tbody>
<tr>
<td>Process</td>
<td>0.18 μm CMOS</td>
</tr>
<tr>
<td>Target CMUT Area</td>
<td>300 μm x 3000 μm</td>
</tr>
<tr>
<td>Active Power</td>
<td>143 μW</td>
</tr>
<tr>
<td>Consumption</td>
<td>1.5 mW</td>
</tr>
<tr>
<td>Sleep Power</td>
<td>5.2 MHz</td>
</tr>
<tr>
<td>Consumption</td>
<td>96.6 dBΩ</td>
</tr>
<tr>
<td>Gain</td>
<td>162 mV/μPa</td>
</tr>
<tr>
<td>Input-referred Noise</td>
<td>0.56 mPa/√Hz at 3 MHz</td>
</tr>
<tr>
<td>Output-referred Noise</td>
<td>91 nV/√Hz at 3 MHz</td>
</tr>
<tr>
<td>Noise Figure</td>
<td>10.3 dB at 3 MHz</td>
</tr>
<tr>
<td>Output P1dB</td>
<td>618 mVpp at 3 MHz</td>
</tr>
<tr>
<td>4-Ch Gain Mismatch</td>
<td>&lt;0.11 dBΩ</td>
</tr>
<tr>
<td>Crosstalk</td>
<td>≤ 50 dBc at 3 MHz</td>
</tr>
<tr>
<td>Wake-up/Sleep Time</td>
<td>≤ 1 μs</td>
</tr>
</tbody>
</table>

mechanical side at the CMUT element surface, in the unit of \( \text{Pa} \) and \( \text{Pa}/\sqrt{\text{Hz}} \), respectively. This input-referred method normalizes the effect of CMUT receive sensitivity and LNA gain. Moreover, the input-referred noise spectral density at the center frequency of the passband is used to approximate \( V_n \text{rms} \) (the input-referred noise spectral density averaged inside the passband) for \( \text{NEF}' \) calculation, because it is the more accessible measurement result in the literature.

The \( \text{NEF}' \) in (8) handles CMUT element size scaling correctly. For example, a CMUT element with \( 2 \times \) bigger surface area presents approximately \( 2 \times \) bigger input capacitance to the LNA. If two of the same LNAs are parallelized to buffer the \( 2 \times \) CMUT element, the same bandwidth and noise figure targets are achieved. Although the parallelization reduces the input-referred noise amplitude by \( \sqrt{2} \times \) and increases the power consumption by \( 2 \times \), the \( \text{NEF}' \) is held unchanged. This is expected since the same LNA design is used in both cases. Another example to show the usefulness of \( \text{NEF}' \) is [8] in Table III. It achieves a very low noise performance, as indicated by the noise figure. On the other hand, excessive power is dissipated on a very small CMUT element, which leads to a relatively high \( \text{NEF}' \).

Table III suggests that the LNA design in this work achieves the lowest \( \text{NEF}' \), indicating the best power efficiency for noise and bandwidth performance. In addition, this design achieves a good linearity performance as shown by the P1dB numbers.

D. Pulse-Echo Experiment

The LNA measurement is followed by the pulse-echo experiment, which characterizes the complete ultrasound signal chain. The test setup in Fig. 9 is revised to perform the experiment. As shown in Fig. 15, the pulser drives a single CMUT element with a wideband pulse as an approximation to the ideal impulse excitation. The narrowest pulse that can be generated from the pulser is a two-level 30-Vpp pulse with 20-ns pulse width [Fig. 16(a)]. The excited ultrasonic wave then propagates through the oil medium and is reflected back at the oil-air boundary 26 mm away from the transducer (the hydrophone is not needed for this experiment). The reflected echo is received by the same CMUT element and amplified by the LNA [Fig. 16(b)]. Because the CMUT blocks the dc component and acts as a differentiator, the received echo looks similar to the derivative of the transmitted pulse, with a positive peak and a negative peak corresponding to the rising edge and the falling edge of the transmitted pulse. Moreover, the pulse’s high frequency components get attenuated more, leading to the dispersed received echo. The echo duration is about 0.3 μs, corresponding to the dominant frequencies (3–5 MHz) that go through the ultrasound signal chain. The echo’s FFT in Fig. 16(c) confirms the intuition. It shows the total channel impulse response, including CMUT, the oil medium, and LNA. It mainly reflects the bandpass characteristic and the wide bandwidth of the CMUT device, with a center frequency of 4.5 MHz and a 6-dB fractional bandwidth of 116%.

E. Ultrasonic Doppler Experiment for Volumetric Flow-Rate Detection

1) Test Setup: The ultrasonic Doppler experiment utilizes the transceiver system to demonstrate the volumetric flow rate detection. In order to generate the flow for the experiment, a water circulation system is assembled in the laboratory, as shown in Fig. 17. The system includes a centrifugal pump, a pressure gauge, a check valve, a needle valve, a flow meter, and a water reservoir. One segment of a rubber tube with the circulating water is placed above the CMUT array inside the oil tank, tilted at 45°, so that the ultrasonic transmitter can emit ultrasonic power into the tube and the receiver detects the scattered Doppler signal from the flowing water for further Doppler processing to estimate the flow rate.
TABLE III
CMUT LNA PERFORMANCE COMPARISON

<table>
<thead>
<tr>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td>Target CMUT Area</td>
<td>300 x 3000</td>
<td>63 x 1037</td>
<td>250 x 250</td>
<td>70 x 70</td>
<td>200 x 200</td>
</tr>
<tr>
<td>Power [P_{in}] (mW)</td>
<td>14.3</td>
<td>3.81</td>
<td>4.0</td>
<td>6.6</td>
<td>2.0</td>
</tr>
<tr>
<td>Bandwidth (MHz)</td>
<td>5.2</td>
<td>20</td>
<td>10</td>
<td>10~20</td>
<td>1.2~3.5</td>
</tr>
<tr>
<td>TIA Gain (dB)</td>
<td>96.6</td>
<td>94.0</td>
<td>112.7</td>
<td>129.5</td>
<td>N/A</td>
</tr>
<tr>
<td>Input-referred Noise [V_{in},in] (mPa/\sqrt{Hz})</td>
<td>0.56@3MHz</td>
<td>2.18@10MHz</td>
<td>1.8@5MHz</td>
<td>3.0@15MHz</td>
<td>12@2MHz</td>
</tr>
<tr>
<td>Noise Figure (dB)</td>
<td>10.3@3MHz</td>
<td>10.5@10MHz</td>
<td>N/A</td>
<td>1.8@10~20MHz</td>
<td>N/A</td>
</tr>
<tr>
<td>Output PdB (mVPp)</td>
<td>61@3MHz</td>
<td>84.2@8MHz</td>
<td>N/A</td>
<td>N/A</td>
<td>N/A</td>
</tr>
<tr>
<td>NEF [V_{in,ref}, -\sqrt{P_{in}}] (mPa/\sqrt{mW/Hz})</td>
<td>2.1</td>
<td>4.2</td>
<td>3.6</td>
<td>7.7</td>
<td>17</td>
</tr>
</tbody>
</table>

The test setup is used to mimic the scenario of detecting human blood flow rate inside a main artery, such as the common carotid artery. The vegetable oil in the oil tank is similar to human body fat in acoustic property, and the inner diameter of the rubber tube is chosen to be a quarter inch (6.35 mm) and is close to the average common carotid artery diameter of an adult person [31]. The water flow mimics the blood flow, and the air bubbles in the water mimics the blood cell, which is responsible for scattering the incoming ultrasonic wave generated from the transmitter. The water flow rate is tunable with a needle valve between 0.32~2.52 L/min and the actual volumetric flow rate is measured by the flow meter as a reference. The volumetric flow rate range of 0.32~2.52 L/min corresponds to a flow speed range of 0.17~1.33 m/s inside the quarter-inch rubber tube, covering enough range between the end-diastolic minimum velocity (0.2 m/s) and the peak-systolic maximum velocity (1.0 m/s) in human common carotid artery [32].

2) Doppler Processing Principle: The classical continuous-wave (CW) Doppler system emits a continuous sinusoidal ultrasonic wave; the receiver detects the frequency shift in the reflected wave due to moving particles, which is proportional to the velocity. Because the continuous wave does not tell the traveled distance before reflected back to the transducer surface, CW systems lack range differentiation [33], [34].

The pulsed-wave (PW) Doppler mode is more widely used in medical ultrasound for its better range resolution. PW systems apply short pulses and can resolve flow information at different depths at the same time. However, the classical Doppler frequency shift cannot be detected in PW mode because the received echoes are of short duration and wideband. Instead, it is through detecting slight timing differences in the received echoes across multiple pulse transmissions, with either subsampling or autocorrelation methods, that the flow information can be extracted [33], [35]. Equivalently, the processing can be intuitively explained from the frequency domain. The spectrum of a train of echoes spaced out at a fixed pulse repetition period (PRP) is a sampled version of a single echo’s spectrum. The spacing of the frequency sampling is the pulse repetition frequency (PRF = 1/PRP). If the scene is not moving, the resultant echo train is steady, and all peaks in its spectrum are located at multiples of the PRF. In contrast, the echo train coming off a moving scene would have slight timing variations between
consecutive echoes, leading to small shift of peaks in the spectrum. When the echo train is subsampled or autocorrelated at the exact same PRF rate, all spectrum peaks are aliased to baseband (0-PRF) in the frequency domain. For a still scene, all peaks add up at dc. However, for a moving scene, the peaks accumulate at a frequency that is proportional to the velocity in the scene, hence the “Doppler shift” $\Delta f$. The velocity $v$ can be calculated from

$$ v = \frac{c \cdot \Delta f}{2 \cdot f_0 \cdot \cos(\theta)} $$

(9)

where $c$ is the sound speed, $f_0$ is the pulse frequency, and $\theta$ is the angle between the flow direction and the ultrasonic beam direction. Rigorous derivations of the PW Doppler principle are found in [33] and [35].

3) Experimental Results: A subsampling-based, PW Doppler flow detection system is implemented in this work, as shown in Fig. 18. The four-channel transceiver chip is driving four consecutive CMUT elements in parallel; five cycles of 3.3-MHz three-level pulses are used. The subsampler down-samples a 25-MHz (ADC sampling rate) data stream of the echo train into a 12.5-kHz (PRF) data stream. The DSP block performs FFT and velocity estimation. The time difference between the subsampling instance and beginning of each repetition cycle, i.e., the pulse transmission, represents the depth where the Doppler information is estimated. The subsampling instance can be adjusted through the PRF synchronization signal, such that all depths can be swept at the same time for Doppler information.

The Doppler measurement result is shown in Fig. 19. At the top of the figure, a typical received ultrasound echo waveform is plotted as a reference. As the waveform progresses in time, it carries information about the acoustic properties of the material at the corresponding distance. Therefore, the $x$-axis of the figure can be interchangeably labeled in time or distance, which are related to each other by

$$ \text{Distance} = c \times \text{Time}/2 $$

(10)

where the factor of 2 represents the fact that the pulse travels twice the distance to reach the medium and come back. As indicated by the reference echo waveform, high-voltage pulses are transmitted at transducer surface ($\text{Distance} = 0$ cm). At roughly $\text{Distance} = 3$ cm, the pulse reaches the front rubber tube wall, exciting visible stationary reflections. After getting through the tube wall, the ultrasound pulse wave is scattered by the moving air bubbles. The scattered echoes are amplified by the receiver and processed, leading to the flow velocity estimations between $\text{Distance} = 4$–5 cm. The different estimated velocity profiles from different flow rates can be clearly differentiated in Fig. 19.

By integrating each estimated flow velocity profile in Fig. 19 across the tube diameter, the volumetric flow rate can be calculated. Fig. 20 shows the calculated volumetric flow rates compared against the actual readings from the flow meter. The estimations match the readings quite well.

There are several error sources in this experiment. First, the window placement for the velocity profile integration could be slightly different than the actual rubber tube position in the space. Second, the ultrasonic Doppler estimation detects the instantaneous flow velocity within the data collection duration, while the flow meter reading is the time averaged flow rate.
Since the flow generated by the centrifugal pump is not perfectly steady and constant, the estimated instantaneous flow rates could deviate from the average flow rate readings.

Fig. 21 shows a die photograph of the test chip fabricated in TSMC 0.18-μm high-voltage CMOS process. Each channel occupies an area of 300 μm x 1100 μm. The shared middle voltage generation circuit occupies an area of 300 μm x 600 μm.

VI. CONCLUSION

This paper describes a transceiver for compact low-power medical ultrasonic imaging applications. The multilevel pulse-shaping technique is successfully applied to the CMUT transmitter for increased power efficiency than traditional designs. The transmitter requires minimum off-chip components and can be easily extended to more channels. The receiver is implemented with the transimpedance amplifier topology in the presence of a highly capacitative source. The LNA achieves the best power–noise tradeoff compared with other works. Furthermore, the combined electrical and acoustic measurements demonstrate complete system functionality. Not only the performance of individual blocks are measured, the test chip is also used as part of an ultrasonic imaging system to perform experiments including ultrasonic beamformation, pulse-echo channel characterization, and PW Doppler flow rate detection.

ACKNOWLEDGMENT

The authors would like to thank the TSMC University Shuttle Program for chip fabrication and Prof. Khuri-Yakub’s group at Stanford University for providing CMUT devices.

REFERENCES


[17] “TX724 Data Sheet: Quad Channel, 3-Level RTZ, ±75 V, 2-Channel Integrated Ultrasonic Pulse,” Texas Intrums., Dallas, TX, USA.


