Session 27 Overview: Biomedical Circuits

IMMD SUBCOMMITTEE

Subcommittee Chair: Makoto Ikeda, University of Tokyo, Tokyo, Japan

Advances in biomedical circuits and systems are essential technology drivers in addressing critical societal needs to increase the effectiveness and reduce the cost of healthcare. This session highlights the latest circuit innovations that contribute to advances in medical devices, sensing and imaging. For implantable and unobtrusive devices, ultrasonic power delivery and telemetry, and improved neural sensing and stimulation are addressed. Advanced medical sensing increasingly combines multiple modalities in a single device, several examples of which are also featured in this session. Finally, circuit innovations enabling improved ultrasonic and magnetic resonance imaging and optical spectroscopy are presented.

1:30 PM
27.1 A 2.8μW 80mVpp-Linear-Input-Range 1.6GΩ-Input Impedance Bio-Signal Chopper Amplifier Tolerant to Common-Mode Interference up to 650mVpp

H. Chandrakumar, University of California, Los Angeles, CA

In Paper 27.1, the University of California, Los Angeles, presents a 2.8μW chopper amplifier optimized for neural recording during stimulation in 40nm CMOS. It has an 80mVpp linear-input-range, –76dB harmonic distortion, and 81dB dynamic range, while handling up to 650mVpp common-mode interference.

2:00 PM
27.2 A 25.2mW EEG-NIRS Multimodal SoC for Accurate Anesthesia Depth Monitoring

U. Ha, KAIST, Daejeon, Korea

In Paper 27.2, KAIST, together with K-Healthwear and Korea University Guro Hospital, demonstrates a combined EEG and Near-Infrared Spectrometry (NIRS) readout, implemented in 65nm CMOS, for continuous quantitative anesthesia depth level monitoring during surgery. The EEG readout achieves an NEF of 3.59 and a LogTIA with a dynamic range up to 60dB is presented for the NIRS readout.
2.30 PM
27.3 All-Wireless 64-Channel 0.013mm²/ch Closed-Loop Neurostimulator with Rail-to-Rail DC Offset Removal
H. Kassiri, York University, Toronto, Canada
In Paper 27.3, York University, together with the University of Toronto, GlaxoSmithKline, and Toronto Western Hospital, present a 64-channel wireless closed-loop neurostimulator with a compact (0.013mm²/ch) and energy-efficient (630nW/ch) channel architecture that merges both amplification and digitization in a single D²S-based neural ADC. The design, implemented in 0.13µm CMOS, yields 1.13mVrms IR noise and an NEF of 2.86.

3.15 PM
27.4 A Sub-1dB NF Dual-Channel On-Coil CMOS Receiver for Magnetic Resonance Imaging
B. Sporrer, ETH Zurich, Zurich, Switzerland
In Paper 27.4, ETH Zurich presents a fully integrated CMOS receiver for medical MRI that can be placed directly on the coil eliminating any RF cabling. The RX features sub-1dB NF and 0dBm IIP3.

3.45 PM
27.5 A Pixel-Pitch-Matched Ultrasound Receiver for 3D Photoacoustic Imaging with Integrated Delta-Sigma Beamformer in 28nm UTBB FDSOI
M-C. Chen, Stanford University, Stanford, CA
In Paper 27.5, Stanford University and STMicroelectronics present a 28nm ultrasound receiver for 3D photoacoustic imaging with the front-end and the modulator integrated within a 250×250µm² pixel. The beamformer achieves 59.9dB SNR with a 7.4× area reduction over comparable prior-art solutions.

4.15 PM
27.6 Single-Chip 3072ch 2D Array IC with RX Analog and All-Digital TX Beamformer for 3D Ultrasound Imaging
S. Kajiyama, Hitachi, Kokubunji, Japan
In Paper 27.6, Hitachi presents a 2D array ASIC with 3072 channel analog RX/TX beamformers for volumetric ultrasound imaging. It features 138Vpp output capability on 0.09mm²/ch. The IC is fabricated in 0.18µm HV-CMOS, and the echo imaging consumes 0.7mW/ch.

4.45 PM
27.7 A 30.5 mm³ Fully Packaged Implantable Device with Duplex Ultrasonic Data and Power LinksAchieving 95kb/s with <10⁻⁴ BER at 8.5cm Depth
T. C. Chang, Stanford University, Stanford, CA
In Paper 27.7, Stanford University demonstrates simultaneous ultrasonic wireless power and duplex data communication for implantable applications. It operates at a depth of 8.5cm tissue phantom at a data rate of 95kb/s. The system consumes 405µW, while receiving power and delivering -9dBm to the transmitter.

5.00 PM
27.8 Fully Integrated Optical Spectrometer with 500-to-830nm Range in 65nm CMOS
L. Hong, Princeton University, Princeton, NJ
In Paper 27.8, Princeton University presents a fully integrated CMOS-based optical spectrometer in a 65nm bulk CMOS process that requires no external optical components. It achieves nearly 10nm resolution and 1.4nm accuracy in peak prediction of continuous-wave excitation in the visible and near-IR range between 500 and 830nm.
27.1 A 2.8μW 80mVpp-Linear-Input-Range 1.66Ω-Input Impedance Bio-Signal Chopper Amplifier Tolerant to Common-Mode Interference up to 650mVpp

Hariprasad Chandrakumar, Dejan Marković
University of California, Los Angeles, CA

Closed-loop neuromodulation with simultaneous stimulation and sensing is designed to administer therapy in patients suffering from drug-resistant neurological ailments. However, stimulation generates large artifacts at the recording sites, which saturate traditional front-ends. The common-mode (CM) artifact can be ~500mV, and the differential-mode (DM) artifact is 50 to 100mV. This work presents a neural recording chopper amplifier that can tolerate 80mVpp CM and 650mVpp CM artifacts in a signal band of 1Hz to 5kHz. To digitize a 2mVpp neural signal to 8b accompanied by an 80mVpp CM artifact requires a linearity of 80dB. Neural recording front-ends also need to function within a power budget of 3 to 3μW/ch, input-referred noise of 4 to 8μVrms, DC input impedance Zin>1GΩ, and high-pass cutoff of 1Hz [1,2]. Prior work has addressed power and noise [2-6], but has low Zin and limited input signal range, making them incapable of performing true closed-loop operation.

The chopper amplifiers in [2,3,5,6] are vulnerable to large CM artifacts (Fig. 27.1.1). The CM signals appearing at the beginning of the pre-charge phase (Fig. 27.1.2) by charge-sharing with Cin, and are disconnected for the remainder of the pre-charge phase. Hence, the input caps Cin are accurately charged to Vin by the end of the pre-charge phase. Thus the settling error in the pre-charge phase is reduced, leading to higher input impedance without increasing power consumption. When f1=0, the aux-buffer bias currents are reduced to 25nA to save power while ensuring that Cin tracks Vin till the next pre-charge phase.

The chopper amplifier for spike and local field potential (LFP) recording is shown in Fig. 27.1.3. The chopping frequency f1 is 23.44kHz. The mid-band gain is set by Cin/Cel=20, and DC-blocking caps Cel are used to avoid chopper ripple at Voff [6]. The servo-loop uses multi-rate duty-cycled resistors with a 10Hz anti-alias filter, Tserv=5ms, f=23.44kHz and f=732.5Hz, which boosts a 350kHz poly-resistor to Rserv=909Ω. Since Cin=12pF, the servo-loop integrator BW is 0.15Hz. The chopper amplifier is fabricated in a 40nm CMOS technology. Figure 27.1.7 shows the chip micrograph. The area is 0.069mm2/ch, and the total power drawn from a 1.2V core supply is 2.8μW. The mid-band gain is 25.7dB, the LP corner is 5kHz, and the HP corner is programmable from 0.12 to 0.3Hz (Fig. 27.1.4) by varying Temax in the MDRCRs. The input-referred noise is 1.8μVrms (1 to 200Hz) and 5.3μVpp (200Hz to 5kHz). Zin at DC is 1.66Ω, which is 76x higher than the input impedance when the aux-path is disabled. Off-chip coupling caps would be needed if Zin<1GΩ. When the assistance from storage caps Cin is disabled, Zin reduced to 600Ω.

When aux-chopping is disabled, a 45mV offset is observed at Vin, and the input-referred noise increases from 1.8μVrms to 4.5μVrms (1 to 200Hz) due to increased flicker noise contribution from the aux-buffers. When aux-chopping is enabled, no discernible offset is present at Vin, and a 5.4μV ripple is observed at 5.86kHz. Rserv=250MΩ and Cin=1nF are used for all measurements.

The total harmonic distortion (THD) for an 80mVpp input at 1kHz is ~76dB (Fig. 27.1.5). A two-tone test was performed (Fig. 27.1.5) in the presence of a 650mVpp CM interferer. When CM is disabled, the signal-to-interferer ratio (SIR) is 7dB, which improves to 38dB when CM is enabled. For an 80mVpp input, the dynamic range of the front-end is 74dB in the spike band and 81dB in the LFP band. Figure 27.1.6 compares the performance of our work with the current state of the art. Our work significantly improves Zin (5.3x), linear input range (2x), introduces tolerance to large CM interferers, increases the maximum resistance of DCRs (32x), requires no off-chip caps and solves the positive-feedback problem in the auxiliary path, while achieving comparable power and noise performance.

Acknowledgements: The authors thank Yuta Toriyama and Dr. Vaibhav Karkare for reviewing the manuscript, Vahagn Hokhinyan for testing support, Dr. Itzak Fried and Prof. Richard Staba for human LFP and spike data, and Lawrence Livermore National Lab for electrodes. This research was developed with funding from the Defense Advanced Research Projects Agency (DARPA). The views, opinions, and/or findings expressed are those of the authors and should not be interpreted as representing the official views or policies of the Department of Defense or the U.S. Government.

References:
Figure 27.1.1: CM cancellation in chopper amps (left) preserves linearity. Multi-rate resistor (right) increases maximum realizable resistance by 32x.

Figure 27.1.2: Conventional auxiliary path amplifies aux-buffer offset (top). Proposed auxiliary path technique (bottom) mitigates the offset and achieves higher Zin.

Figure 27.1.3: Complete implementation of the chopper amplifier.

Figure 27.1.4: Chip measurements: gain with programmable HP corner (top), input-referred noise (center), input impedance (bottom).

Figure 27.1.5: Measured linearity (top). Measured outputs for 2-tone tests with a CM interferer (bottom). The DM and CM tones at 900Hz emulate the stim artifact, and the 1kHz tone is the signal of interest.

Figure 27.1.6: Comparison with current state of the art.
Figure 27.1.7: Chip micrograph (top) of 8-channel IC, and single-channel amplifier (bottom).
27.2. A 25.2mW EEG-NIRS Multimodal SoC for Accurate Anesthesia Depth Monitoring

Unsoo Ha¹, Jaehyuk Lee¹, Jihee Lee¹, Kwanta Kim¹, Minseo Kim¹, Taehwan Roh⁵, Sangsik Choi³, Hoi-Jun Yoo⁴
¹KAIST, Daejeon, Korea
²K-Healthwear, Daejeon, Korea
³Korea University Guro Hospital, Seoul, Korea

There has been recent research into continuous monitoring of the quantitative anesthesia (ANES) depth level for safe surgery [1]. However, the current ANES depth monitoring approach, bispectral index (BIS) [3], uses only EEG from the frontal lobe, and it shows critical limitations in the monitoring of ANES depth such as signal distortion due to electrocautery, EMG and dried gel, and false response to the special types of anesthetic drugs [3]. Near-infrared spectroscopy (NIRS) is complementary to EEG [2], and can not only compensate for the distorted depth level, but also assess the effects of various anesthetic drugs. In spite of its importance, a unified ANES monitoring system using EEG/NIRS together has not been reported because NIRS signals have widely different dynamic ranges (10pA to 10nA), and also signal level variations from person to person and environment are not manageable without closed-loop control (CLC).

In this paper, a multimodal head-patch system that simultaneously measures EEG and NIRS on the frontal lobe is proposed for accurate ANES depth monitoring. A 60dB dynamic range logarithmic TIA (logTIA) is adopted to amplify the photodiode (PD) signal and a CLC driver is used to compensate for the human-to-human variations. Also, an LNA with high Zin (1GΩ) and wide electrode DC-offset (EDO) cancelation range (350mV) is integrated to obtain reliable EEG signals.

Figure 27.2.1 illustrates the ANES depth monitoring system. On the bottom side of the head-patch, 2-ch EEG/EMG electrodes (each 3.14cm²) and a 1-ch NIRS module composed of a red(λ=670nm)/infrared(λ=850nm) vertical-cavity surface-emitting laser (VCSEL) and a PD on silicon holder are integrated on the polyethylene terphathelate (PET) film. The SoC, BLE module and battery are assembled on the flexible-PCB. Compact (26×3.5cm²) and lightweight system (<10g) enables the practitioner to measure signals with high convenience. The system is applied on the patient’s forehead/temple and all the acquired signals are pre-processed in the SoC and then sent to an external device through Bluetooth. The device can display the ANES depth level with the help of a deep neural network to help the anesthesiologist adjust the drug dosages for safe ANES.

Figure 27.2.2 shows the overall block diagram of the SoC. It consists of: (1) a 2-ch EEG/EMG readout for high input impedance (Zin) and large EDO tolerance, (2) a 1-ch NIRS readout for wide dynamic range, (3) a Red/IR VCSEL driver for CLC, (4) a 12b SAR ADC, and (5) a digital module for pre-processing and communication through BLE module.

Figure 27.2.3 shows the circuit schematics and the measurement results of the LNA. The EDO usually increases up to ±300mV almost like the dry electrode [4] due to the long electrode size, sometimes >24 hours [1]. To eliminate the EDO with low noise, a mixed-mode dc-cv servo loop (MM-DSL) is proposed. The MM-DSL is composed of digital (CDSL) DSL for wide range and low noise, and analog (GDSL) DSL for fine resolution. Initially, with the help of SAR, digital DSL cancels EDO coarsely with a 5b DAC (10.5mV LSB). Then, analog DSL (up to 21mV) is activated to decrease the remaining EDO. CDSL is updated according to VOUT,i,n to remove the temporal EDO drift. Right after the start and whenever the value of CDSL changes, fC is increased to >50Hz to reduce the settling time from >60s to <50ms. In this work, a 130ms window is assigned to ensure enough margin. With the help of MM-DSL, the LNA shows 3.59 NEF with ±350mV EDO cancellation range. The impedance-boosting loop [2] enhances Zin about 50× (>1GΩ) at 60Hz.

References:
Figure 27.2.1: Proposed anesthesia depth monitoring system.

Figure 27.2.2: Overall block diagram of the monitoring IC.

Figure 27.2.3: EEG low-noise amplifier with Mixed-Mode DSL.

Figure 27.2.4: NIRS log TIA with ambient light rejection.

Figure 27.2.5: NIRS closed control loop for adaptive duty-cycling.

Figure 27.2.6: System measurement results in the clinical trials.
Figure 27.2.7: Chip micrograph and performance summary.

### Low-Noise Amplifier for EEG (On Chip DSL)

<table>
<thead>
<tr>
<th>Parameter</th>
<th>JBCIC’86</th>
<th>JBCIC’87</th>
<th>JBCIC’88</th>
<th>JBCIC’89</th>
<th>Tech. Notes</th>
</tr>
</thead>
<tbody>
<tr>
<td>VDD (V)</td>
<td>3.3</td>
<td>3.3</td>
<td>3.3</td>
<td>3.3</td>
<td></td>
</tr>
<tr>
<td>Current (μA)</td>
<td>2.8</td>
<td>4.2</td>
<td>6.2</td>
<td>9.8</td>
<td></td>
</tr>
<tr>
<td>CMAN (dB)</td>
<td>100</td>
<td>110</td>
<td>120</td>
<td>130</td>
<td></td>
</tr>
<tr>
<td>f1 (kHz)</td>
<td>1000</td>
<td>1000</td>
<td>1000</td>
<td>1000</td>
<td></td>
</tr>
<tr>
<td>Noise (μV)</td>
<td>50</td>
<td>50</td>
<td>50</td>
<td>50</td>
<td></td>
</tr>
<tr>
<td>NF (dB)</td>
<td>4.1</td>
<td>4.1</td>
<td>4.1</td>
<td>4.1</td>
<td></td>
</tr>
<tr>
<td>Temp. (°C)</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td></td>
</tr>
</tbody>
</table>

### EEG Readout

- **Gain**: 50 dB
- **Bandwidth**: 20 kHz
- **Current**: 3 mA

### LNA

- **Gain**: 50 dB
- **Bandwidth**: 20 kHz
- **Current**: 3 mA

### NIR

- **Gain**: 100 dB
- **Bandwidth**: 100 kHz
- **Current**: 300 μA

### Digital Block

- **Operating Freq**: 50 MHz
- **On-chip SRAM**: 2 Gb
27.3 All-Wireless 64-Channel 0.013mm²/ch Closed-Loop Neurostimulator with Rail-to-Rail DC Offset Removal

Hossein Kassiri1, M. Reza Pazhouhandeh2, Nima Soltani2, M. Tariq Salami1, Peter Carlen4, Jose Luiz Perez Velazquez2, Roman Genov2

1York University, Toronto, Canada
2University of Toronto, Toronto, Canada
3GSK (GlaxoSmithKline), Stevenage, United Kingdom
4Toronto Western Hospital, Toronto, Canada

Accurate capture and efficient control of neurological disorders such as epileptic seizures that often originate in multiple regions of the brain, requires neural interface microsystems with an ever-increasing need for higher channel counts. Addressing this demand within the limited energy and area of brain-implantable medical devices necessitates a search for new circuit architectures. In the conventional designs [1-5], the channel area is dominated by the bulky coupling capacitors and/or capacitor banks of the in-channel ADC, both unavoidable due to the channel architecture, and unscalable with CMOS technology. Additionally, channel power consumption, typically dominated by the LNA, cannot be reduced lower than a certain limit without sacrificing gain and/or noise performance. In this paper, we present a 64-channel wireless closed-loop neurostimulator with a compact and energy-efficient channel architecture that performs both amplification and digitization in a single ΔΣ-based neural ADC, while removing rail-to-rail input DC offset using a digital feedback loop. The channel area and power consumption depend only on the active components and switching frequency, respectively, making the design both technology- and frequency-scalable.

Figure 27.3.1 (top) shows the block diagram of the ΔΣ (Δ+ΔΣ) neural ADC. Because of the added Δ stage, the output bit-stream (Y) of the feedback integrator is comprised of two segments of 4b binary-weighted programmable push/pull current sources. The segments are biased by two currents, different by a factor of 16 for a total of 8b of resolution. The output bit-stream, which is a derivative of the input signal, is fed to a compact and energy-efficient channel architecture that performs both amplification and digitization in a single ΔΣ-based neural ADC, while removing rail-to-rail input DC offset using a digital feedback loop. The channel area and power consumption depend only on the active components and switching frequency, respectively, making the design both technology- and frequency-scalable.

The 0.13μm CMOS SoC was validated in both early detection (experiment 1) and control (experiment 2) of seizures in temporal lobe epilepsy (rat model). Figure 27.3.6 (top, left) shows an example of in vivo online on-chip real-time seizure detection without stimulation. In the second experiment, the SoC was configured to automatically trigger the closed-loop electrical stimulation for the purpose of suppressing upcoming seizures. Figure 27.3.6 (top, right) illustrates the SoC-triggered stimulation upon a seizure onset detection. The SoC is compared with the state of the art both in terms of the channel performance (Fig. 27.3.6 (bottom)) and the system performance (Fig. 27.3.7 (bottom)). The chip micrograph and the channel floorplan are shown in Fig. 27.3.7 (top).

References:

Figure 27.3.1: Block diagram of the ΔΣ-based neural recording channel in both single-ended (top) and differential (bottom) configurations.

Figure 27.3.2: Circuit schematic of the recording channel (top); the input integrator during the two phases of a single clock cycle (bottom).

Figure 27.3.3: Block diagram of the fabricated neurostimulator SoC.

Figure 27.3.4: Experimentally measured performance characteristics of the neurostimulator channel.

Figure 27.3.5: Experimentally measured performance: the on-chip phase calculation, power scalability with bandwidth, area scalability with technology, and wireless transmitters output spectra.

Figure 27.3.6: In vivo seizure detection and closed-loop stimulation results for a rat epilepsy model (top), neurostimulator channel performance comparison (bottom).
Figure 27.3.7: Chip micrograph and channel floorplan (top), and a system-level comparison table (bottom).
27.4 A Sub-1dB NF Dual-Channel On-Coil CMOS Receiver for Magnetic Resonance Imaging

Benjamin Sporrer1, Linbao Wu2, Luca Bettini1, Christian Vogt1, Jonas Reber1, Josip Marjanovic2, Thomas Burger2, David O. Brunner3, Klaas P. Prüssmann4, Gerhard Tröster2, Quting Huang1

1ETH Zurich, Zurich, Switzerland
2University and ETH Zurich, Zurich, Switzerland

Magnetic Resonance Imaging (MRI) is a widely used medical imaging technique. It employs a strong static magnetic field (1.5 to 10.5T for human imaging) to split the spin states of the 1H nuclei in the body, and RF excitation to induce transitions and coherence among them. Gradient fields are superimposed to modulate the 1H resonance frequency, which enables spatially distinguishable signals to be picked up by RF receive coils. A high-field MRI provides better sensitivity and resolution but requires better receivers (RX), as signal DR and H resonance increase (128MHz for 3T, 300MHz for 7T). Overall sensitivity and imaging speed can be enhanced by closely surrounding the target anatomy with tens of RX coils (as in MIMO) [1], at the expense of as many shielded RF cables to carry the information out of the field. Progress in PCB size has allowed multi-channel RX to be placed inside the magnetic field (in-bore), reducing the RF cable length to less than 1m [2,3]. Ultimately, the RX should be placed directly on-coil to avoid bulky coaxial cables and improve patient comfort and safety by acquiring data in-bore and sending them digitally to the MRI scanner via an optical fiber link. The latter is cheap, flexible and insensitive to magnetic fields. The immediate vicinity of the coils and the patient is, however, a hostile as well as sensitive electromagnetic environment, which tolerates only the smallest of PCBs and virtually no magnetic material in its components. Integration of the full RX chain in a CMOS chip, which is small, non-magnetic and low power, holds the key to the next wave of compact MRI coil arrays for advanced medical imaging. This paper presents a fully integrated dual-receiver RFIC for coil arrays intended for (ultra-) high field (1.5 to 10.5T and 64 to 450MHz) scanners for clinical MRI, where requirements are considerably stricter than previously reported transceiver ICs [4,5] on palm-held NMR devices for spectroscopy or lab-on-chip applications. Figure 27.4.1 shows the overview of an MRI system with on-coil RX, and associated timing diagrams for a receiving slot. The reference to the (2nd-stage) RF PLL is typically sensitive to magnetic field). A highly stable oven-controlled crystal (OCXO) on the chip implementation. This research is funded by the Swiss Nano-Tera project WearableMRI.

Acknowledgments:
The authors would like to thank ACP Advanced Circuit Pursuit AG for supporting the chip implementation. This research is funded by the Swiss Nano-Tera project WearableMRI.

References:
**Figure 27.4.1:** Block diagram of the on-coil PCB and receiver states during MRI operation.

**Figure 27.4.2:** Fully integrated CMOS direct-conversion receiver with reflective LNA and 2-stage PLL for system synchronization.

**Figure 27.4.3:** Measured phase noise of divided LO (64MHz) in a 3T MRI scanner. Influence of 1st stage digital PLL is illustrated.

**Figure 27.4.4:** Noise-cancelling LNA with configurable low input impedance (Re(Z_in) $\ll$ RS). Re(Z_in) can be set between 25 and 100 $\Omega$.

**Figure 27.4.5:** MRI receiver benchmark and measured RX IC performance for a 7T setup (300MHz). Jitter is measured in 3T.

**Figure 27.4.6:** Acquired images from a human wrist with an on-coil receiver in a one channel setup shown in the bottom right corner.
Figure 27.4.7: Integrated MRI receiver chip micrograph.
27.5 A Pixel-Pitch-Matched Ultrasound Receiver for 3D Photoacoustic Imaging with Integrated Delta-Sigma Beamformer in 28nm UTBB FDSOI

Man-Chia Chen1, Aldo Peña Perez2, Sri-Rajasekhar Kothapalli2, Philippe Cathelin2, Andrea Cathelin2, Sanjiv Sam Gambhir1, Boris Murmann3

1Stanford University, Stanford, CA 2STMicroelectronics, Crolles, France

A variety of emerging applications in medical ultrasound rely on 3D volumetric imaging, calling for dense 2D transducer arrays with thousands of elements. Due to this high channel count, the traditional per-element cable interface used for 1D arrays is no longer viable. To address this issue, recent work has proven the viability of flip-chip bonding [1] or direct transducer integration [2]. This shifts the burden to a CMOS substrate, which must provide dense signal conditioning and processing before the massively parallel image data can be pushed off chip. A common approach for data reduction is to employ subarray beamforming (BF), which applies delay and sum operations within a group of pixels. To implement such functionality within the tight pixel pitch, prior works have implemented the delays using simple S/H circuits [2] or analog filters [3], and typically suffer from a combination of issues related to limited delay, coarse delay resolution and limited SNR.

This work leverages the integration density of modern CMOS to demonstrate a pitch-matched digital subarray beamforming receiver (RX) with signal conditioning and $\Delta \Sigma$ modulator ($\Delta \Sigma M$) integrated within a pixel area of $250\times250\mu m^2$ (see Fig. 27.5.1). Our proof-of-concept IC supports a subarray of 4x4 pixels and is flip-chip bonded to a Capacitive Micromachined Ultrasound Transducer (CMUT) chip that is similar to the one used in [1]. Since our application is photoacoustic imaging (receive-only using external laser pulses), we did not integrate a transmitter interface. However, in a large-scale array implementation of our concept, it is conceivable to add this functionality using a subset of the pixels for transmit [2].

Figure 27.5.2 compares our approach with prior art: analog BF [2-3] and digital BF using a per-channel Nyquist ADC. The latter approach is popular for 1D arrays, but difficult to integrate within a pitch-constrained 2D array. In addition, the Nyquist ADC must typically oversample to provide sufficient timing resolution, which further exacerbates the integration issue. The work of [4] combines analog and digital Nyquist-rate BF, but the area per element is ~5x larger than our pixel size. To enable area-efficient digital BF, this work uses a $\Delta \Sigma$ approach similar to [5]. The oversampling of the $\Delta \Sigma M$ naturally provides sufficient timing resolution for BF, enables low-complexity analog design with small passives, and simplifies the signal routing (1b outputs). In our chip, the 16 bitstreams are routed to a global digital block for decimation filtering (DF1) and beamforming (BF = FIF0 + summation), followed by a final decimation filter (DF2) off chip. Within the on-chip block, the BF is placed after DF1, which was identified as the preferred option due to the lower FIF0 clock speed and the commensurate reduction in power (see Fig. 27.5.3). Placing the BF before DF1 (as in [5]) would lead to a slightly lower gate count (since DF1 is shared), but the savings are insignificant due to the relatively low complexity of the employed cascaded integrator comb (CIC) filter. We expect the advantages of the DF-first option to become more pronounced for larger arrays, where early clock rate reduction is critical. Despite the decimation by DF1, the delay resolution is still 83.3ns, which is sufficient for a 5MHz CMUT center frequency. The implemented FIFOs have a depth of 2, providing the required delay range for our 4x4 subarray (1.06µs).

Figure 27.5.4 shows the analog front-end. The transimpedance amplifier (TIA) provides five gain levels using a programmable $R_{\text{network}}$. The TIA output is taken against a replica to facilitate supply noise cancellation as the succeeding lowpass filter (LPF) performs single-ended to differential conversion. Both the TIA and LPF are designed using 1.5V thick oxide devices (for large DR), while all other circuits use core devices (1V supply). The VGA uses a Padé approximation to provide fine linear-in-db gain tuning. The 1b $\Delta \Sigma M$ (see Fig. 27.5.5) uses a 3rd-order architecture with an OSR of 48 to provide 60dB peak SNR in a 10MHz BW. The employed inverter-based SC integrator is similar to [6]. It uses three gain stages to achieve the required gain with minimum L, and it is designed to slew for the most of the clock period. The large swing at the 3rd stage input during slewing leads to small devices and a compact layout. The input bias of the 3rd stage is established using diode replicas and stored on $C_S$. In comparison to [6], this obviates the need for special high $V_T$ devices and resistors. As illustrated in Fig. 27.5.5, the designed ADC is the smallest published among designs with similar BW and SNDR.

Our chip is fabricated in a 28nm UTBB FD-SOI CMOS process. The 16 RX pixels occupy 1mm$^2$ and consume 356mW, while the synthesized digital block occupies 0.4mm$^2$ and consumes 173mW. The $\Delta \Sigma M$ occupies 1/4th of the pixel area and consumes 6.65mW. The $\Delta \Sigma M$ was measured in isolation (test pixel), showing $\text{SNR}_{\text{peak}} = 59.9$dB and $\text{SNDR}_{\text{peak}} = 58.9$dB for a 2MHz input. To evaluate the entire RX, a tiled 4x4 2D CMUT array is flip-chip bonded onto the 28nm chip. The receiver is tested within a photoacoustic imaging setup, where the acoustic signals are induced by light absorbing wire targets (see Fig. 27.5.6). The cross-sectional view from the y-z plane shows three parallel wires at different depths, while the view from the x-z plane captures their diagonal placement.

Figure 27.5.7 shows the top view of the chip stack and the RX chip, along with a comparison to the state of the art (focusing on BF performance). Relative to the hybrid analog/digital BF approach of [4], our work has comparable delay resolution and power dissipation, while achieving 7-4x smaller area and 7dB improvement in single-channel SNR. Our maximum delay range is lower due to the different requirements imposed by our 4x4 array, but it is straightforward to extend it through a longer FIFO. A direct comparison to analog BF ICs [2-3] is more difficult to make, due to the significantly different performance parameters. If we relax the SNR to 40dB and reduce the delay range to 200ns, we estimate an 8x and 5x power reduction for our $\Delta \Sigma M$ and BF, respectively. This would yield a BF power of 2.99mW/channel, which lies between [2] and [3]. In summary, we view the demonstration of in-pixel A/D conversion and efficient $\Delta \Sigma M$ BF as the most important aspects of this work. We believe that the presented approach offers a viable path toward larger arrays with pitch-matched electronics, high-fidelity readout and digital subarray BF.

Acknowledgement:
Silicon fabrication was provided by STMicroelectronics through CMP. We thank Romain Feuilleul, Christophe Bernicot (ST) and Jean-Francois Paillotin (CMP) for design support, Astrid Tomada (SLAC), PacTech, and Hai Nguyen (Silicon) for chip assembly. Prof. Khuri-Yakub, Anshuman Bhuyan, Byung-Chul Lee, and Ji-Hoon Jang for discussion and preparation of CMUT. This work was funded in part by Stanford’s Initiative on Rethinking Analog Design (RAD) and the C2S2 Focus Center, one of six research centers funded under the Focus Center Research Program (FCRP), an SRC subsidiary.

References:
Figure 27.5.1: Block diagram of the implemented RX and pixel layout.

Figure 27.5.2: Comparison of beamformer architectures.

Figure 27.5.3: Comparison of two ΔΣ beamforming options.

Figure 27.5.4: RX front-end (contained in each pixel).

Figure 27.5.5: Block diagram of ΔΣM (contained in each pixel) and SC integrator half circuit.

Figure 27.5.6: Photoacoustic imaging setup and results.
Figure 27.5.7: Photo of chip assembly and RX die, along with a comparison to the state of the art.
A diagnostic ultrasound (US) system transmits acoustic waves at several to tens of MHz into the human body for clinical purposes and detects the reflected waves to observe the internal organs without having a medical operation or radiation exposure. The system is composed of a main unit and probe connected via coaxial cables. The probe is very small because medical technicians laboriously grab and manipulate it for a long time. To avoid image obscurity depending on medical technicians, high-speed and high-resolution 3D/4D imaging is necessary. For this reason, several thousands of lead bulk piezoelectric material transducers (TD) need to be squeezed into the small probe. Since the number of cables is limited to several hundreds, the probe needs to include beamforming functionality and a 2D array IC [1-6], which includes thousands of US transceivers.

Figure 27.6.1 shows a block diagram of the diagnostic US system and the proposed 2D array IC, which includes 3,072 US transceivers. Since a TD pillar is 300×300μm, each US transceiver size needs to be the same pitch. The TD is connected to an IC via a low-temperature co-fired ceramic (LTCC) interposer to alleviate fabrication difficulties. On the other side of the pillars, there is an acoustic impedance matching layer and a lens. The 2D array IC bilaterally interfaces from the 5,072-ch TD to the 128-ch coaxial cables. The 24 TD and US transceivers are grouped to coherently execute transmitter and receiver US signal processing. The main unit digitizes the signal to perform beamforming and image rendering.

The subarray block diagram is shown in Fig. 27.6.2. The transmitter is an all-digital architecture, and is composed of a digital beamformer, a switched-capacitor delay (SCD), and a tunable amplitude 3-level pulser (TA3LP). Therefore, it is free from waveform distortion and timing error caused by cable propagation between the main unit and the probe. The pulser power consumption is generally far less than that of linear amplifiers, but the pulser’s constant amplitude leads to limited image quality. TA3LP in this work has a built-in amplitude control function, and the TX and RX chain share the SCD to squeeze all the necessary blocks in a die. The TD TRX signal IOs include zero-power-TRX-isolation switches (ZTRSW) to protect subsequent blocks with ordinal rating elements. Although conventional ZTRSW uses a zener diode as a floating switch gate bias and inherently needs a lot of power, ZTRSW in this work only consumes 10μW during RX mode using MOSFET active floating-gate bias topology. The next stage is a programmable gain-and-input impedance low-noise amplifier (PGZLNA). PGZLNA boosts the weak signal to a reasonable level for the next stage. The programmable input impedance is used to achieve >85dB dynamic range. The charge-domain adders (CDADDs) perform, in the charge domain, correlated-signal summation of 24 SCD outputs, and the CDADDs are placed in a subarray without signal headroom concerns. If each SCD setting is ideal, and their correlation is ideal, RX SNR is >24 times better. It means the noise level is >24 times lower. Therefore, the next stage block, which is the low-noise cable buffer (LNCBUF), needs to be low noise and have heavy coaxial cable drivability. Finally, the main unit digitizes the signal to perform beamforming and image rendering.

The 2D array IC is assembled and implemented in the probe and main unit with a real-time rendering engine that is optimized and caters for the probe. The system can capture 3D tissue images with several arbitrary 2D cutting planes within the angle of view (AoV) of this system. Arbitrary 2D cutting plane images are easily extracted from the full-volume 3D dataset. The System AoV is more than 90°×90°, and it can look over the whole human heart during echo-cardiography. Figure 27.6.5 shows the 2D and 3D image of a phantom (white ball: agar / black ball: agar and graphite mixture) as an example. Figure 27.6.6 shows the performance comparison. Our work is only a 2D array IC that includes the RX analog, all-digital TX beamformer, and related peripheral blocks. The silicon area occupies 0.09mm 2 /ch and consumes 0.7mW/ch during B-mode capture. Regardless of the size, US transducers use CRM, an operational amplifier (OPA), and a CDADD. Since CRM occupies most of the area, TX and RX chain share the CRM. While the RX chain uses CRM as an analog memory, the TX chain uses them as a 2D D-latch. During TX mode, the OPA operates as a comparator. The switches, being carefully placed and deployed, achieve the optimal metastability to avoid signal distortions. The CRM, which is composed of 32 capacitors, controls the time differences of write (charge) and read timing to make the necessary delay. The write and read signals are non-overlapped 40MHz pulse trains and cover the 25- to 750ns delay range. During RX mode, dynamically changing the focal points leads to fast image rendering. For this reason, duplicated write pulse or doubling the read pulse width can increase the delay time. Adversely doubling the write pulse width or skipping read pulse can shorten it. Figure 27.6.4 shows the TA3LP circuits, which can launch plus, zero, and minus level signals. The output stage takes on the source follower push-pull topology (PP). PP consumes a lot of power only during the pulse transition durations and alleviates the previous inverter (INV) stage bias current and transistor sizes. The INV output has grounding switches to make zero level (RZ: return to zero circuit). PP and INV will operate within a safe operating area to avoid device destruction. Since TA3LP does not use feedback topology, rise and fall time adjustment is necessary. As previously noted, TA3LP has a 256 level amplitude adjustment module to compete with a linear amplifier system and cover many diagnostic modes necessary for the US system. The amplitude transition can settle to <4.4μs.

A die micrograph of the 2D array IC implemented in a 0.18μm HV SOI CMOS process is shown in Fig. 27.6.7. The total area is 417mm2 including the digital beamformer and the other control and the interface circuits.

References:
Figure 27.6.1: Block diagram of diagnostic ultrasound system with 3,072-ch 2D array IC.

Figure 27.6.2: Block diagram of 2D array IC.

Figure 27.6.3: Switched-capacitor delayer circuit (top). Timing chart of fixed focus (bottom left) and dynamic focus (bottom right).

Figure 27.6.4: Circuit diagram of tunable amplitude 3-level pulser and its control block.

Figure 27.6.5: Evaluated 3D image (top right) and 2D cutting plain images from the full-volume 3D dataset (bottom right).

Figure 27.6.6: Comparison table of ultrasound array IC.
Figure 27.6.7: Die photograph.
The next generation of implantable medical devices focuses on minimally invasive miniaturized solutions that operate reliably at large depths, providing duplex communication for closed-loop therapies, and enabling multi-access for a network of implants to gather information or provide systemic interventions. Using ultrasound (US), power and data can be efficiently transferred through the body as its wavelength at MHz is comparable to a mm-sized receiver, resulting in improved focusing, coupling, and acoustic-to-electrical conversion efficiency. Furthermore, thanks to the low propagation loss (~1dB/cm/MHz) and 7.2mW/mm² safety limit, several mW of power is obtainable at the receiver, enabling high-power, complicated functionalities.

While initial experiments using US for power transfer to implants have been successful [1,2], there is still a significant need to have reliable uplink data communication at depth. Other studies have worked on RF uplink [3,4] or backscatter [2], which have limited depth and/or multi-access capability restricting coordinated therapies throughout the body. We demonstrate a miniaturized fully packaged implant that receives both US power and data and also transmits US data for uplink. Figure 27.7.1 depicts the conceptual diagram of the system as well as the block diagram of the implant. It consists of two transducers (US power and data receiver (RX) and data transmitter (TX)), a discrete capacitor, and a CMOS chip in TSMC 65nm GPP technology, which includes power-management, data and clock-recovery circuits, frequency generators, finite state machine, pseudorandom binary sequence (PRBS), and the power amplifier (PA). While both RX and TX are made from piezoelectric materials, the operating frequency of the receiver is chosen to be ~1MHz for low propagation loss and mm-size as the resonance of the receiver is inversely proportional to its thickness. Using clock recovery and frequency generation circuitry, the carrier frequency of the output data (fout) is ~2.6× of the input frequency (fin) to avoid self and external interference from the high-power downlink and its harmonics.

Figure 27.7.2 illustrates the timing diagram during operation. The implant is charged until the LDO voltage is established. The implant then listens for the data and clock-recovery circuits, frequency generators, and resonance of the TX driven differentially with 1V supply should be <5kΩ. For successful transmission at large tissue depth, the desired power consumption is 100μW for successful transmission at large tissue depth. Therefore, the desired impedance of the TX driven differentially with 1V supply should be <5kΩ at resonance. Using PZT4, the TX is sized 0.55×0.55×0.4mm³, giving a resonance near 2.5MHz with impedance ~4kΩ as shown in Fig. 27.7.5. Simulated PA efficiency is 86%, delivering a peak power of 125μW to the TX. A fully wireless end-to-end test with external power/data transmitter and data receiver in castor oil, commonly used as a tissue phantom with loss ~0.6dB/cm/MHz, is performed. The carrier frequency of the received voltage is 2.45MHz for fin of 950kHz. The received voltage waveform for an 8.5cm link after filtering out the harmonics of fout is also plotted; the interference signals are mostly due to multipath reflections and can be mitigated. Bit error rate (BER) is calculated by sending the on-chip PRBS 10 times; no errors are found for both 3.5 and 8.5cm links, achieving a BER of <10⁻⁴ for a data rate of 95kb/s. The average signal-to-interference ratio (SIR) is computed to be 27.4 and 15.8dB respectively; SIR can be further improved with better alignment and larger receiver gain.

The next generation of implantable medical devices focuses on minimally invasive miniaturized solutions that operate reliably at large depths, providing duplex communication for closed-loop therapies, and enabling multi-access for a network of implants to gather information or provide systemic interventions. Using ultrasound (US), power and data can be efficiently transferred through the body as its wavelength at MHz is comparable to a mm-sized receiver, resulting in improved focusing, coupling, and acoustic-to-electrical conversion efficiency. Furthermore, thanks to the low propagation loss (~1dB/cm/MHz) and 7.2mW/mm² safety limit, several mW of power is obtainable at the receiver, enabling high-power, complicated functionalities.

While initial experiments using US for power transfer to implants have been successful [1,2], there is still a significant need to have reliable uplink data communication at depth. Other studies have worked on RF uplink [3,4] or backscatter [2], which have limited depth and/or multi-access capability restricting coordinated therapies throughout the body. We demonstrate a miniaturized fully packaged implant that receives both US power and data and also transmits US data for uplink. Figure 27.7.1 depicts the conceptual diagram of the system as well as the block diagram of the implant. It consists of two transducers (US power and data receiver (RX) and data transmitter (TX)), a discrete capacitor, and a CMOS chip in TSMC 65nm GPP technology, which includes power-management, data and clock-recovery circuits, frequency generators, finite state machine, pseudorandom binary sequence (PRBS), and the power amplifier (PA). While both RX and TX are made from piezoelectric materials, the operating frequency of the receiver is chosen to be ~1MHz for low propagation loss and mm-size as the resonance of the receiver is inversely proportional to its thickness. Using clock recovery and frequency generation circuitry, the carrier frequency of the output data (fout) is ~2.6× of the input frequency (fin) to avoid self and external interference from the high-power downlink and its harmonics.

Figure 27.7.2 illustrates the timing diagram during operation. The implant is charged until the LDO voltage is established. The implant then listens for the data input (falling edge) from the external source to generate a NOTCH and prepare for uplink transmission. The Recovered CLK is obtained from fin. DATA CLK divides Recovered CLK by 10; thus, the data rate is close to 100kb/s for fin near 1MHz with OOK modulation, sufficient for a range of applications including pressure, temperature, or neural recording. An on-chip 2¹–1 PRBS is used for data generation. The next input falling edge stops the transmission. If sufficient energy is retained on the storage capacitors, the process can be immediately restarted, re-enabling uplink transmission.

The power-management circuits, which are shown in Figure 27.7.3, use a two-path architecture. The main path includes an active full-wave rectifier for AC-DC conversion, a charge pump to boost the generated voltage, and a low-dropout regulator (LDO) regulating supply at 1V; the low-power auxiliary path sets up the biasing, similar to what [1] presents. PMN-PT with dimension of 0.9×0.9×0.5mm³ is chosen for RX; it has a measured impedance of 900Ω and resonance of 0.95MHz as desired for impedance matching and low propagation loss. Chip efficiency, defined as load power over input rectifier, along with measured voltage waveforms at four nodes are also plotted.

Figure 27.7.4 shows the frequency generator and the differential class-D PA driving the TX. The frequency generator uses an open-loop architecture to ensure fast settling and robust fout. In the setup phase, the frequency of a free-running current-starved RO (f_ref) is resolved with a 9b asynchronous counter controlled by COUNT. Five MSBs are stored in the registers, effectively dividing fout by 16. During transmission, a 5b asynchronous counter counts the RO cycles, which is compared to the stored value; if they match, a pulse is sent to a 5b Johnson counter, dividing the frequency of the pulse train by six to generate fout. An example calculation for an fout of 250MHz and fin of 1MHz is shown in gray text. This method is insensitive to process and temperature variation. In addition, a Widlar bias is used so that fout is stable for small perturbations of supply. fout is recounted in the setup phase for each transmission; and fin and fout have a stepped relationship because of the open-loop operation, leading to a more reliable system. Measured fout across different LDO voltages shows that it stays constant for LDO voltages from 0.84 to 1.0V. As OOK modulation is used, the generator is disabled when DATA = 0 and the output of the PA is cycled and pulled to GND to save energy; for DATA = 1, the generator can start up and settle in <1μs and consumes 32μW during runtime.

Taking into account the loss, the transmitted output power needs to be at least 100μW for successful transmission at large tissue depth. Therefore, the desired impedance of the TX driven differentially with 1V supply should be <5kΩ at resonance. Using PZT4, the TX is sized 0.55×0.55×0.4mm³, giving a resonance near 2.5MHz with impedance ~4kΩ as shown in Fig. 27.7.5. Simulated PA efficiency is 86%, delivering a peak power of 125μW to the TX. A fully wireless end-to-end test with external power/data transmitter and data receiver in castor oil, commonly used as a tissue phantom with loss ~0.6dB/cm/MHz, is performed. The carrier frequency of the received voltage is 2.45MHz for fin of 950kHz. The received voltage waveform for an 8.5cm link after filtering out the harmonics of fout is also plotted; the interference signals are mostly due to multipath reflections and can be mitigated. Bit error rate (BER) is calculated by sending the on-chip PRBS 10 times; no errors are found for both 3.5 and 8.5cm links, achieving a BER of <10⁻⁴ for a data rate of 95kb/s. The average signal-to-interference ratio (SIR) is computed to be 27.4 and 15.8dB respectively; SIR can be further improved with better alignment and larger receiver gain.

Acknowledgement: The authors thank Prof. Khuri-Yakub for valuable discussions. The chip fabrication was made possible by the TSMC University Shuttle Program. We also acknowledge Mentor Graphics for the use of the Analog FastSPICE (AFS) Platform. This material is based upon work supported by the DARPA YFA and the NSF CAREER award ECCS-1454107.

References:
Figure 27.7.1: Conceptual diagram of the wireless ultrasonic powering and communication system and the block diagram of the implant.

Figure 27.7.2: Timing diagram during operation.

Figure 27.7.3: Diagram and measurement of power-management circuits.

Figure 27.7.4: Frequency generator and measured $f_{out}$ versus LDO voltage and $f_{in}$.

Figure 27.7.5: Measured impedance of the US TX and the measured received voltage waveform in castor oil (ultrasound tissue phantom).

Figure 27.7.6: Demonstration of US wireless power and data communication through 6cm animal tissue with packaged implant.
Figure 27.7.7: Die photo and comparisons of implantable devices with wireless powering and data communication.
27.8 Fully Integrated Optical Spectrometer with 500-to-830nm Range in 65nm CMOS
Lingyu Hong, Kaushik Sengupta
Princeton University, Princeton, NJ

Next-generation IoT systems are expected to be enabled by compact, low-cost, low-power, smart sensing devices that provide a wealth of information to build new applications and capabilities. Among sensing modalities, optical spectrometry is one of the rapidly growing areas of interest due to its wide range of applications from environment monitoring, industrial, and home applications to healthcare [1-3]. As shown in Fig. 27.8.1, current optical spectrometers are large and bulky with non-integrated components that limit their application potential. In this paper, we present a fully integrated CMOS-based optical spectrometer in a 65nm bulk process that requires no external optical components. The spectrometer achieves nearly 10nm resolution and 1.4nm accuracy in peak prediction of continuous-wave (CW) excitations between 500 and 830nm.

The principle of operation of a classical optical spectrometer relies on light incident on a grating structure getting diffracted, focused and sensed by a photodetector array, which enables us to correlate the measured voltages of the photodetectors to the incident optical spectrum (Fig. 27.8.1). The architecture of the presented CMOS optical spectrometer is shown in Fig. 27.8.1. The incident light is delivered through an optical fiber into the CMOS chip through a 3μm-sized opening aperture on the 7th metal layer (M7) and is converted into a propagating mode in metal-insulator-metal (MIM) waveguides between M4-M7 [5, 6]. The propagating wave is incident on a concave grating structure (M4-M7) that is designed to disperse the incident mode and create a wavelength-dependent focal plane 400μm away within the MIM waveguides. The spatial distribution is sensed by deflecting the focused light with a metal shield into an array of photodetectors underneath (Fig. 27.8.1). The signal is then processed on-chip with low-noise circuitry and digitized off-chip for analysis. The entire structure measures 650×540μm².

The layout of the dispersive optical elements, the MIM waveguide, and the constituent circuits is shown in Fig. 27.8.2. It shows focusing of the optical waves of different wavelengths within the chip on the focal plane. The resolution is determined by the numerical aperture of the grating structure and the slit aperture size at the fiber input, and is simulated to be 5nm. Unlike free-space propagation in a classical spectrometer, mode propagation within the copper-metal waveguides at optical frequencies can be very complex and the interaction with the grating structure on-chip can be dependent on the mode profiles in addition to wavelength. To remove this extra dependence, the MIM waveguides are designed to suppress higher-order modes, allowing only the lowest fundamental mode to interact with the grating structure. Figure 27.8.2 shows the attenuation of the higher-order modes by >20dB relative to the fundamental over the wave travel distance across the spectrum.

Figure 27.8.3 shows the array architecture where the photodetector arrangement follows the curvature of the focal plane. The array of 60 detectors is realized with n-well/ p-sub junctions, and the signals are integrated by capacitive TIAs to remove the dependence on the diode capacitance in this non-custom imager process. To partially suppress the effect of dark currents, each signal is sensed differentially with a reference diode and a differential TIA which suppresses common-mode perturbations. It is also important to ensure that the detectors are accommodated within the 4μm spot size of the optical focus points so that the resolution is not degraded by the photodiode layout (Fig. 27.8.3). Therefore, the grating structure and the array architecture need to be co-designed for optimal performance and resolution. As shown in Fig. 27.8.3, the signals are multiplexed and finally processed through a correlated double sampling architecture to remove offsets and low-frequency drifts and ultimately digitized by a 16b ADC off-chip and analyzed.

Stray light scattering can cause unintended errors in spectral estimation, which are typically eliminated by elaborate optical shielding. For a chip-scale spectrometer, such external effects can be minimized by proper packaging, while inside the chip, the propagation of light in undesired directions can be shielded with metal optical walls (Fig. 27.8.2). This work, we show spectral estimation in the unpackaged die (Fig. 27.8.7) not completely shielded from external scattered light. The spectrometer is first characterized against CW excitations with a wide-band source and a tunable narrow-band filter (-10nm) realized with a linear variable filter on a motorized translational stage (Fig. 27.8.4). Over the range of measured wavelengths, the array response creates a responsivity matrix defined as

\[ R_{\text{opt}} \in \mathbb{R}^{M \times N} \]

where \( N \) is the number of diodes and \( M \) is the number of wavelengths of characterization. When a spectrum approximated by \( S_{\text{est}} \in \mathbb{R}^M \) is incident on the chip, the array response is given by \( V_{\text{op}} = R_{\text{opt}} S_{\text{est}} + V_{\text{noise}} \), where \( V_{\text{op}} \in \mathbb{R}^M \) is the random noise voltages of the outputs. Given \( V_{\text{op}} \) and measured responsivity \( R_{\text{opt}} \), we estimate the incident spectrum \( S_{\text{est}} \) by the minimizing the following:

\[ \min_{S_{\text{est}} > 0} \left\| V_{\text{op}} - R_{\text{opt}} S_{\text{est}} \right\|^2 + \lambda \left\| S_{\text{est}} \right\|^2. \]

The regularization parameter \( \lambda \) allows for robust estimation in presence of noise by eliminating solutions with undesirable spikes due to the inverse estimation process [4]. The measured responsivity matrix which is a one-time characterization is shown in Fig. 27.8.4 by removing the average spectral responses of the sensors. The spectral dependence of the spatial distribution of intensity is evident from the figure. The presence of scattering of stray light creates the non-focus profile, but the wavelength-dependent variation of the spatial distribution still allows us to achieve spectral estimation across the range.

The chip is tested with CW excitations of light and Fig. 27.8.5 shows the spectral estimation from the measured responses showing robust estimation with regularity only 40mW estimated to be entering the chip. When the chip is excited at a wavelength between the characterization wavelengths, the reconstruction shows two peaks near the center wavelength. The estimation of the peak can be progressively narrowed reaching down to 1.4nm when the estimation is averaged over five responsivity matrices with 1nm spacing. The chip is then tested with excitation of varying spectra and the estimation reasonably matches the measured spectra without optical packaging and external shielding (Fig. 27.8.6). The normalized RMS error of reconstructed spectra given by:

\[ e = \sqrt{\frac{\sum |S(\lambda) - S_{\text{est}}(\lambda)|^2}{\sum |S(\lambda)|^2}} \]

was measured to be around 15 to 20%. The chip is powered by a 3V supply and dissipates 30mW of total DC power. The chip-scale spectrometer shows the feasibility of integrating complex optical systems-on-chip for various applications in sensing by exploiting metal-optical structures in CMOS through a co-design approach.

References:
Figure 27.8.1: (a) Traditional optical spectrometer (b) and (c) principle of operation, light-path, the grating and detection structure in the CMOS integrated spectrometer.

Figure 27.8.2: (a) Integrated grating structure layout, MIM waveguide, and circuitry. (b) Optical simulation: focusing of different wavelengths along focus curvature. (c,d) EM simulations showing mode profiles of 0th and 1st order, showing higher losses for the higher-order modes that are eliminated. (e) Loss of the first 2 modes showing >20dB rejection of all higher-order modes compared to the fundamental.

Figure 27.8.3: Architecture and circuit components. Spatial distribution of light intensity is detected by array of 60 photodetectors whose outputs are integrated differentially through CTIAs, and multiplexed before processing through CDS. Measured analog output waveforms for sequential addressing of pixels during one integration time.

Figure 27.8.4: (a) Spectrometer responsivity characterization set-up with broadband source and linear variable narrow-band filter between 500 and 830nm. (b) Measured spectral responsivity shown by pixel outputs against wavelength by removing average response. Presence of scattering of stray light creates non-focus profile, but wavelength-dependent variation of spatial distribution still allows spectral estimation across range.

Figure 27.8.5: (a) Measured spectral estimation with regularization with CW-illuminated chip at characterization frequencies. (b) Measured spectral estimation when illuminated at between characterization wavelengths. (c) Progressive narrowing of peak estimation by averaging with 5 responsivity matrices shifted by 1nm (d) Averaging narrows peak estimation accuracy to 1.4nm.

Figure 27.8.6: (a-c). Measured spectral estimation when excited with broadband optical sources showing reasonably good agreement with incident spectra. (d) Normalized errors in spectral estimation of the three test spectra.
Figure 27.8.7: Chip photo and the measurement set-up.