17.5 50nW 5kHz-BW Opamp-Less ΔΣ Impedance Analyzer for Brain Neurochemistry Monitoring

Maged El Ansary[1], Nima Soltani[1], Hossein Kassiri[1], Ruben Machado[1], Suzie Dufou[1,2], Peter L. Carlen[1,2], Michael Thompson[1], Roman Genov[1]

[1]University of Toronto, Toronto, Canada
[2]Toronto Western Hospital, Toronto, Canada

Potassium (K+) and sodium (Na+) ions are the main signal carriers in the nervous system. The difference in the concentration of both K+ and Na+ across the neuron cell membrane, as regulated by respective ion channels, plays a critical role in the propagation of action potentials, the spike-like signals neurons communicate with, as shown in Fig. 17.5.1 (top, left and middle). Due to their significant role in neuronal signaling, K+ channel malfunctions are linked to over 100 neurological disorders, such as schizophrenia, Alzheimer's disease, spreading depression, and epilepsy. Selective real-time sensing of K+ concentration (denoted as [K+]) is therefore critical for the advancement of many neurological therapies.

Conventional intracranial electroencephalography (iEEG) does not measure [K+] as it cannot distinguish between K+ and Na+ ions. Experimental non-invasive techniques such as MRI are low-accuracy, low-resolution and expensive. Unlike iEEG, K+ amperometry selectively monitors K+ ion channel activity. As shown in Fig. 15.5.1 (top, middle), K+ -sensitive probe molecules are interrogated by applying a small (~10mV) sinusoidal voltage waveform onto the working electrode (WE) and by measuring the corresponding current I. The resulting impedance is thus a selective measure of [K+] in the extracellular space ([Na+] is measured the same way, but yields a distinctly different impedance spectrum). Double-barrel glass electrodes are the gold standard for *in vivo* K+ amperometry, but are bulky – thus have low temporal and spatial resolution, and are not implantable.

We present a CMOS-based miniature technology for low-cost high-spatial-resolution monitoring of [K+] in an animal brain *in vivo*. As shown in Fig. 17.5.1 (bottom, right), CMOS top-metal aluminum pads are plated with Ni/Pd/Au to make the surface inert and biocompatible. As shown in Fig. 17.5.1 (bottom, left and middle), we have developed a method to deposit a mixed ultrathin surface monolayer onto a gold surface in order to both provide sensitivity to K+ in the cerebrospinal fluid and to minimize interference caused by protein and other interferers adsorption. The two modified K+ -binding crown ethers (the cyclic chemical compounds in Fig. 17.5.1 (bottom) capture K+ ions and as a result modify the surface impedance. The water molecules, present in the cerebrospinal fluid, bond with the MEG-OH branches, which then shield the electrode from the interfering hydrophobic molecules. As compared to conventional probe-based techniques, the electrode surface does not undergo significant changes after implantation. The impedance changes only when [K+] changes, and not as a result of interferers attaching themselves to the electrode. The lumped model of an electrode-tissue interface can be viewed, at a given frequency, as a complex impedance Z as shown in Fig. 17.5.1 (right). [K+] in the brain ranges from ~1mM to ~100mM leading to 10pA-1nA current range for the developed electrodes.

Figure 17.5.2 (top, left) depicts how the Re/Im components of Z are efficiently computed by Frequency Response Analysis (FRA) for two cases: (1) conventionally, the current is first digitized and then multiplied by cos/sin to get Re/Im terms, respectively; (2) here we fully eliminate the TIA and the multi-bit multiplication (Fig. 17.5.2 (top, middle)) and introduce a compact and low-power current-input ΔΣ ADC with a single-bit multiplying counter (i.e., by a square wave) (Fig. 17.5.2 (top, right)). In the ΔΣ topology shown in Fig. 17.5.2 (bottom), instead of an active integrator, a simple charge-sharing DAC with no opamp integrates onto a grounded capacitor. Due to noise shaping, the harmonics resulting from the square wave are strong only at high frequencies. As the summation weight of these harmonics decreases hyperbolically with frequency, the stronger higher-frequency harmonics do not significantly corrupt the output ($\leq$ 2% error).

Figure 17.5.3 shows the superimposed z-domain models of both the conventional active-integrator (black) and the presented ΔΣ ADC (black and red). Removing the op-amp causes two additional branches with the weights α and β to appear. The absence of a TIA causes the 1-bit DAC to have an additional unwanted branch connecting it to the comparator input. The two designs become approximately equivalent when the coefficients α, β, and γ are << 1. This is true when: (1) the sampling period is negligible compared to the time constant of the tissue-electrode interface, (2) C_{INT} is much greater than the equivalent shunt capacitance of the microelectrode, and (3) the input current is much smaller than the full-scale input current of the ADC. In the context of in vivo potassium amperometry, condition (1) is satisfied when the ADC operates at sampling frequencies in the ~MHz range (above 800kHz) for the input bandwidth of a few kHz (5kHz). Condition (2) is met by keeping size of the WE small (~50μm x ~100μm). C_{INT} is the MIM capacitor which shields the WE from underneath, so the WE capacitance is effectively equal to the fringe capacitance between the WE and the adjacent reference electrode (RE). Condition (3) is met by choosing a high oversampling ratio (OSR) of 500 or higher.

Figure 17.5.4 details the circuit implementation of the opamp-less channel. The DAC pushes/pulls charge by the small capacitor C_{DAC} to/from the much larger capacitor C_{INT}, depending on the comparator output. A high OSR (at least 500) is selected so that the digital feedback loop maintains V_{INT} approximately equal to V_{REF}. The input current is integrated directly by the input capacitor C_{INT} leading to a voltage range smaller than that on the conventional opamp feedback capacitor, but this does not increase susceptibility to comparator noise. The opamp-less ADC rejects in-band noise added by the comparator, such as DC offset or 1/f noise, because, as shown above, its transfer function is approximately equivalent to that of an active-integrator ΔΣ ADC. Hysteresis and path dependence in the comparator are mitigated by resistive degeneration in the comparator. Comparator low duty cycle helps to further save power.

Figure 17.5.5 (top, left) shows the SNDR plot of the ADC at two sampling frequencies: 846kHz and 8.46MHz. The latter yields a higher dynamic range, due to an increased range of currents that can be pushed/pulled by the DAC. Figure 17.5.5 (top, right) shows that the OSR improves the SNDR by ~28dB/decade. Figure 17.5.5 (bottom, left) depicts the ADC output versus the input current at the two sampling frequencies. The experimental results differ from the simulated results for small currents as this is near the 1pA sensitivity. Figure 17.5.5 (bottom, right) shows the error in the digital impedance output using both a 10-bit sine wave and its 1-bit (square wave) approximation. For OSR of 500 or more, the computation error is less than 2%.

The IC was validated in vivo in an anesthetized immobilized mouse. To avoid bonding wire damage, the IC was mounted on a flexible polyimide electrode array shown in Fig. 17.5.6. (bottom, left), with its four sensory gold strips chemically functionalized as described in Fig. 17.5.1 (bottom, left). This sensor array was then placed on a section of the cerebral cortex of a mouse exposed by craniotomy. Figure 17.5.6 also depicts experimentally measured: calibration curve for both [K+] and [Na+] (top, left), [K+] during pulsed optogenetic stimulation (top, right), and [K+] on four electrodes during a seizure (bottom, right). The recordings are consistent with those by a gold-standard glass electrode, also shown. In fact, the Au electrode yields a significantly higher temporal resolution than what is currently attainable by the glass electrode.

Figure 17.5.7 shows the micrograph and the comparison table. This design achieves SNDR of 50.3dB over 5kHz bandwidth, which is 21.8dB less than the best relevant design but with the advantage of over two orders of magnitude in power, and over an order of magnitude in area.

References:

[1] M. Stanacevic, et al. "VLSI potentiostat array with oversampling gain modulation for wide-range neurotransmitter sensing." *IEEE TBioCAS*, vol. 1, no. 1, pp. 63-72, 2007.
[2] F. Heer, et al. "CMOS Electro-Chemical DNA-Detection Array with On-Chip ADC." *ISSCC Dig. Tech. Papers*, pp. 168-169, Feb. 2008.
[3] H. Jafari, et. al., "Nanostructured CMOS wireless ultra-wideband label-free PCR-free DNA analysis SoC", *IEEE JSSC*, vol. 49, no. 5, pp. 1223-1241, 2014.
[4] Y. Liao, et. al., "A 3-μW CMOS glucose sensor for wireless contact-lens tear glucose monitoring," *IEEE JSCC*, vol. 47, no. 1, pp. 335-344, 2014.
[5] B. Bozorgzadeh, et. al., "Neurochemostat: a neural interface soc with integrated chemometrics for closed-loop regulation of brain dopamine," *IEEE TBioCAS*, vol. 10, no. 3, pp. 654-667, 2016.

 978-1-5090-4940-0/18/$31.00

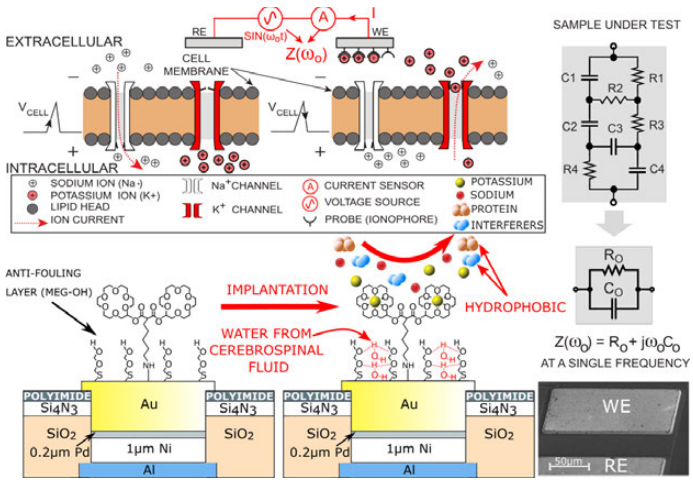


Figure 17.5.1: Sensing extracellular action potential ions using Potassium sensitive amperometry.

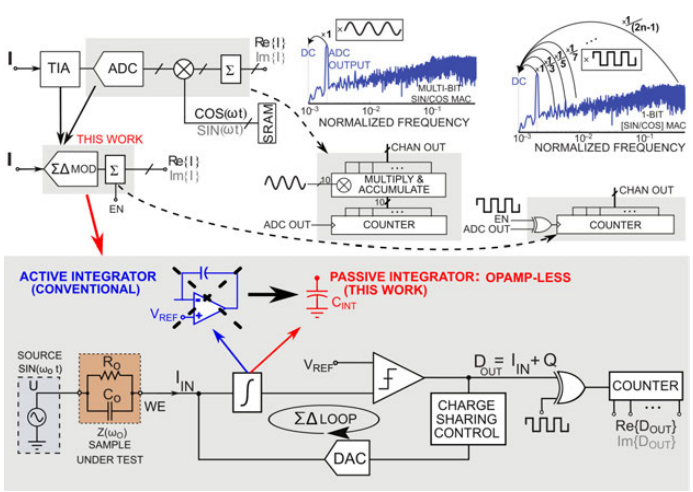


Figure 17.5.2: Opamp-less ADC-based potentiostat block diagram.

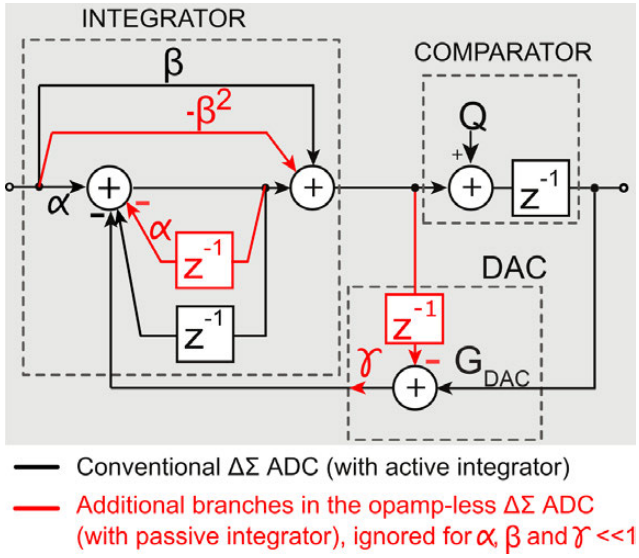


Figure 17.5.3: Opamp-less ADC-based and conventional DS ADC equivalency.

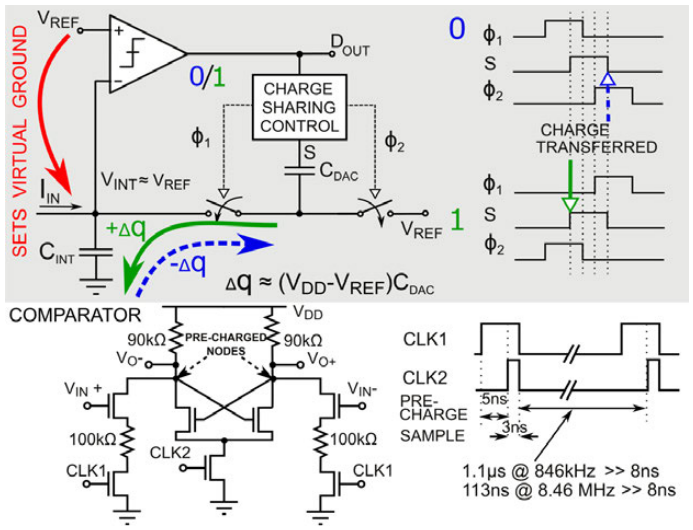


Figure 17.5.4: Charge-mode DAC and comparator operation.

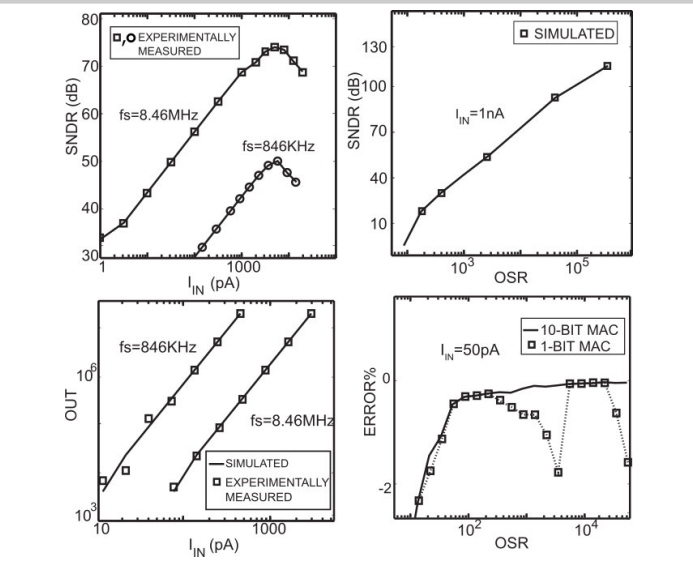


Figure 17.5.5: Experimentally measured opamp-less SD ADC results.

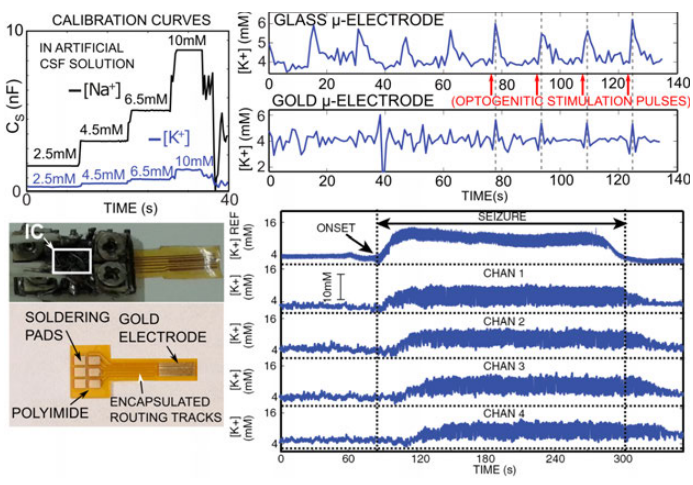
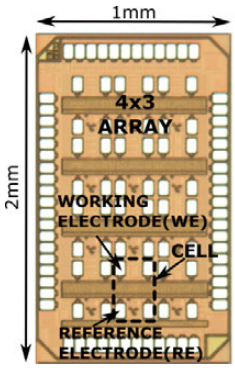


Figure 17.5.6: In vivo [K+] recordings experimentally measured on flexible gold electrodes.



	This Work	[1]	[2]	[3]	[4]	[5]
Technology [µm]	0.13	0.5	0.18	0.13	0.13	0.35
No. of Channels	12	16	24	54	1	1
Bandwidth [Hz]	5000	4000*	-	3400	0.2	4880
Sensitivity [pA]	1	0.1	97	8.6	20	55
SNDR [dB]	50.3	50	70**	56.5	-	72.1
Area [mm^2]	0.0039	-	0.03*	0.06	0.22	0.111
Power [W]	50n	3.4µ	-	8µ	500n	9.5µ

*Estimated **External 5nF ceramic capacitor used at the input

Figure 17.5.7: Die micrograph and table of comparison.

978-1-5090-4940-0/18/$31.00

17.6 A 200Mb/s Inductively Coupled Wireless Transcranial Transceiver Achieving 5e-11 BER and 1.5pJ/b Transmit Energy Efficiency

Wen Li[1], Yida Duan[2], Jan M. Rabaey[1]

[1]University of California, Berkeley, CA
[2]Inphi, Santa Clara, CA

Recent advancements in medical neural science and brain research have enabled the potential of uninterrupted simultaneous recording of thousands of neurons. To minimize the risk of infection to the patients, wireless data transmission is the preferred option. Therefore, the large amount of data generated by the neural recorder must be transmitted through the scalp and skull bone to an outside data processor to translate into action or to be analyzed. A next-generation 1024-channel implanted neural recorder that uses 20KS/s 8b ADC to capture action and field potential can generate up to 164Mb/s data, which imposes a stringent requirement on the communication throughput of the implanted transmitter (TX). In addition, TX power consumption must be kept as low as possible for longer battery life or safe wireless power transfer, while the power consumption of the receiver (RX) chip outside the scalp can be relaxed. Current state-of-the-art designs that use backscattering [1], pulse harmonic modulation (PHM) [2] or ultra-wide-band (UWB) communication [3,4] suffer either from limited throughput or low TX energy efficiency. Other techniques such as ultrasound [5] usually have data-rates limited to tens of Kb/s and do not penetrate through the skull. In this paper, we utilize inductive coupling for transcranial wireless data transfer to achieve 200Mb/s data-rate and ultra-low TX power. Series de-Q resistors are employed to alleviate inter-symbol-interference (ISI) caused by the inductor self-resonance. The entire TX chip uses a single 0.5V V_{dd} to save power. To generate 200MHz TX clock from 10MHz reference at such low supply, an injection-locked PLL with fully digital frequency tracking and spur suppression is used.

The TX chip (Fig. 17.6.1) consists of an injection-locked PLL, a current driver, and a PRBS7 generator for testing. Similar to [6], the output stage (M_{1-4}) directly controls the polarity of the current in the TX coil (I_{TX}) by turning on either $M_{1,4}$ or $M_{2,3}$. Since the PMOS's ($M_{3,4}$) conduct current only when *ck* is high (*X* or $\overline{X}$ is low), I_{TX} pulse lasts ~1/2 Unit Interval (UI) and returns to zero before the next bit arrives (Fig. 17.6.1). The 2 transitions (dI_{TX}/dt) of every I_{TX} pulse couple to the RX coil to produce 2 voltage pulses V_{RX}, with a positive pulse followed by a negative pulse representing bit 1 and the opposite representing bit 0. For simplicity, only the first V_{RX} pulse of each bit is used in the RX; the subsequent pulse is discarded. Since large coupling between the inductor coils is desired for high RX signal swing, a 10×10mm 2-turn TX coil is employed to increase inductance while keeping the form factor small (Fig. 17.6.2). This significantly lowers the self-resonance to ~300MHz, potentially causing excessive ringing in pulse response and severe ISI. Equalization techniques such as decision-feedback-equalization can be used to remove post-cursors, but are ineffective in this wireless context due to the large ISI magnitude and channel variation. To reduce ringing, we simply distribute 7 series de-Q resistors on the off-chip TX coil so that it is critically damped. The added resistors however increase noise and lower the SNR, so de-Qing is used only in the TX coil where the signal is large. The RX inductor is a single-turn coil with high resonance frequency (2.4GHz). As shown in the Fig. 17.6.2, de-Qing removes most post-cursors at the cost of reduced RX peak swing.

The RX consists of 4 stages of common-source amplifier followed by a StrongArm comparator (Fig. 17.6.3). Capacitors (C_{BW}) are added to Stage-1 to filter out ringing caused by the RX coil. The Stage-4 input is AC-coupled to eliminate offset by Stage-1 to 3. The output of Stage-4 can be brought off-chip to plot the eye-diagram. A 1.5V supply is used for the 4-stage amplifiers to increase the headroom and gain. The rest of the RX chip uses a 1.1V supply. The amplifier chain achieves 59dB gain and 22.4uV input referred noise. A CML divider followed by a phase interpolator (PI) adjusts the sampling instance to the peak of the eye.

Using 0.5V V_{dd} in the TX chip imposes severe challenges for clock generation. Low power ring oscillator (RO) can be extremely noisy. Large inductor size at 200MHz makes LC oscillator unattractive. An injection-locked RO (ILRO) emerges as a viable solution because of its large noise suppression bandwidth. However, ILRO has large spur if its frequency is not calibrated well. In-between injection pulses, ILRO oscillates at its natural frequency. The injection pulses align the phase of ILRO to the reference clock by changing its period at instance of injection (T_{20} in Fig. 17.6.4). Therefore, monitoring the ILRO period right at and before the injection (T_{20} and T_{19}) provides a way to continuously tune ILRO [7]. Unlike [7] in which a time-to-digital converter (TDC) is used, the sign of the difference between T_{20} and T_{19} is measured by a 1b time-comparator (2 integrators and a comparator) to save power (Fig. 17.6.4). Its results are averaged before feeding back to ILRO through a 10b DAC to reduce noise. The time-comparator offset is calibrated in foreground. Frequency tracking is accomplished by counting ILRO pulses in 1 reference period and by adjusting its frequency until the count reaches 20, thus guaranteeing a division ratio of 20. Spur-suppression is enabled only after frequency is locked. The ILRO consists of 5-stage current-starved inverters and consumes only 20uW. To achieve fine tuning step with moderate DAC resolution, K_{VCO} is kept low by using back-gate voltage of the bottom NMOS's in deep n-well (DNW) as control voltage. Thanks to 0.5V V_{dd}, the forward bias body-source junction current in these NMOS's is negligible. The PLL achieves -43dBc residue injection-spur and 59ps total rms jitter while consuming 130uW.

To demonstrate the transceiver architecture, TX/RX prototypes are fabricated using TSMC 65nm GP CMOS. The coupled-inductors are implemented with 200um wide trace on 2-layer PCB, on which the chips are directly wire bonded. The wireless link is characterized for 2 different channels: The 11mm air gap and the 11mm thick scalp and skull bone of an 8-week primordial piglet carcass closely resembles that of humans (Fig. 17.6.2). The 10MHz TX reference is synchronized to the 400MHz RX clock during testing. Although small frequency offset between TX and RX may exist in practice due to crystal oscillator ppm offset, various clock recovery techniques can be used in RX to address this issue. A small aluminum cover is placed on top of the RX board to reduce radio interference. The eye diagram and bathtub curves for both channels are shown in Fig. 17.6.5. Note two eyes exist because of 2 I_{TX} transitions per UI; the asymmetry between them is caused by difference in dI/dt between turning on and off PMOS's ($M_{3,4}$) in TX (Fig. 17.6.1). Only the 1st eye (left) is used in RX bit-detection. The prototype achieves 5e-11 BER over the scalp and skull bone and <1e-12 BER over the air. The entire TX chip consumes only 300uW, with 130uW for PLL, 165uW for TX current driver, and 5uW for PRBS generator and buffers (Fig. 17.6.6). The RX consumes 37.2mW.

In this paper, we have demonstrated combining inductively coupled data transmission with injection-locked clock generation improves the energy efficiency of bio-implantable TX by nearly 7× (Fig. 17.6.6). In addition, the prototype achieves over 2× higher date-rate (200Mb/s) and 3 orders-of-magnitude lower BER (5e-11) than state-of-the-art wireless TX for biomedical implants.

Acknowledgements:
The authors would like to thank DARPA Subnets, Starnet Human Intranet, BWRC for funding and support; K. Abdelhalim, B. Helal, C. Chu, R. Mohanavelu, J. Pernillo, L. Tse of Inphi for valuable input and help during testing; Y. Lu of Qualcomm, M. Mark of Infineon, P. Nuzzo of USC for useful discussion; TSMC for chip fabrication.

References:
[1] R. Muller, et al., "A Miniaturized 64-Channel 225μW Wireless Electrocorticographic Neural Sensor", *ISSCC Digest Tech. Papers*, pp. 412-413, Feb. 2014.
[2] F. Inanlou, et al., "A 10.2 Mbps Pulse Harmonic Modulation Based Transceiver for Implantable Medical Devices", IEEE JSSC, vol. 46, no. 6, pp. 1296-1306, June 2011.
[3] M. Chae, et al., "A 128-Channel 6mW Wireless Neural Recording IC with On-the-Fly Spike Sorting and UWB Transmitter", *ISSCC Digest Tech. Papers*, pp. 146-147, Feb. 2008.
[4] K. Abdelhalim, et al., "64-Channel UWB Wireless Neural Vector Analyzer SOC With a Closed-Loop Phase Synchrony-Triggered Neurostimulator", *IEEE JSSC*, vol. 48, no. 10, pp. 2494-2510, Oct. 2013.
[5] T. Chang, et al., "A 30.5mm3 Fully Packaged Implantable Device with Duplex Ultrasonic Data and Power Links Achieving 95kb/s with <10-4 BER at 8.5cm Depth", *ISSCC Digest Tech. Papers*, pp. 460-461, Feb. 2017.
[6] Y. Yoshida, et al., "Wireless DC Voltage Transmission Using Inductive-Coupling Channel for Highly-Parallel Wafer-Level Testing", *ISSCC Digest Tech. Papers*, pp. 470-471, Feb. 2009.
[7] B. Helal, et al., "A Highly Digital MDLL-Based Clock Multiplier That Leverages a Self-Scrambling Time-to-Digital Converter to Achieve Subpicosecond Jitter Performance", *IEEE JSSC*, vol. 43, no. 4, pp. 885-863, Apr. 2008.

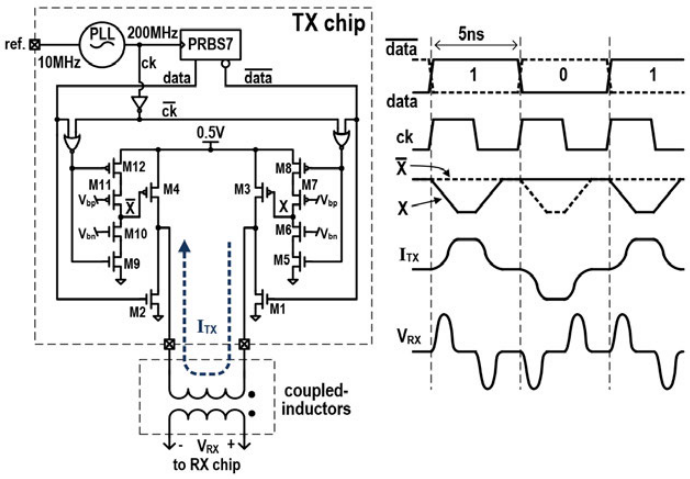


Figure 17.6.1: The TX architecture and its signal waveform.

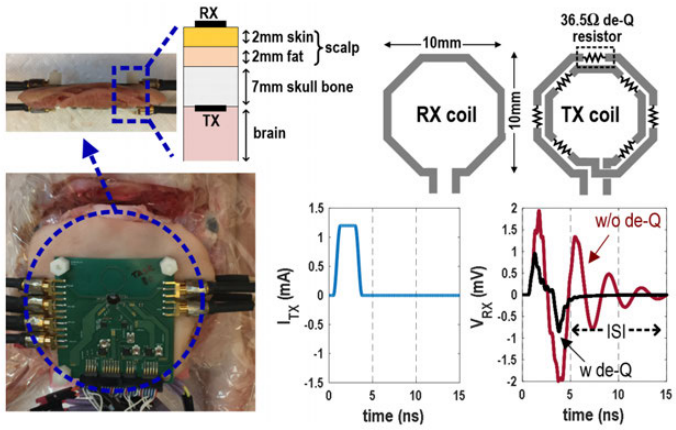


Figure 17.6.2: The 11mm piglet scalp and skull bone channel, the TX/RX coils, and the simulated pulse response.

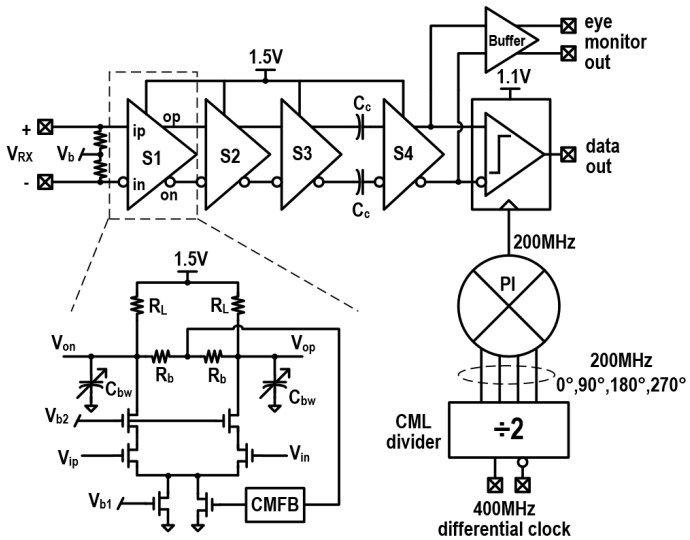


Figure 17.6.3: RX architecture.

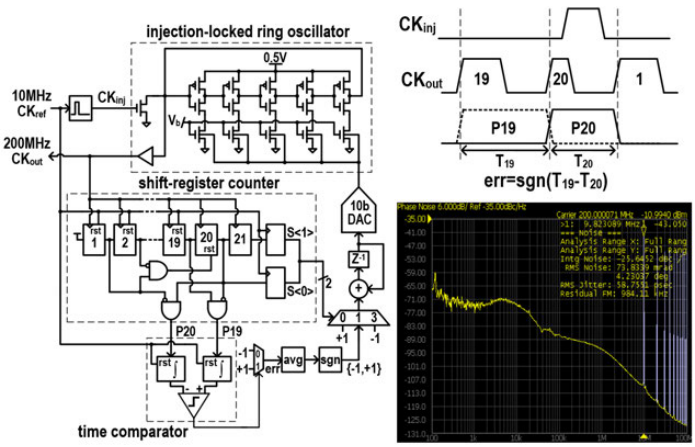


Figure 17.6.4: Architecture, signal waveform, and measured phase noise of injection-locked PLL.

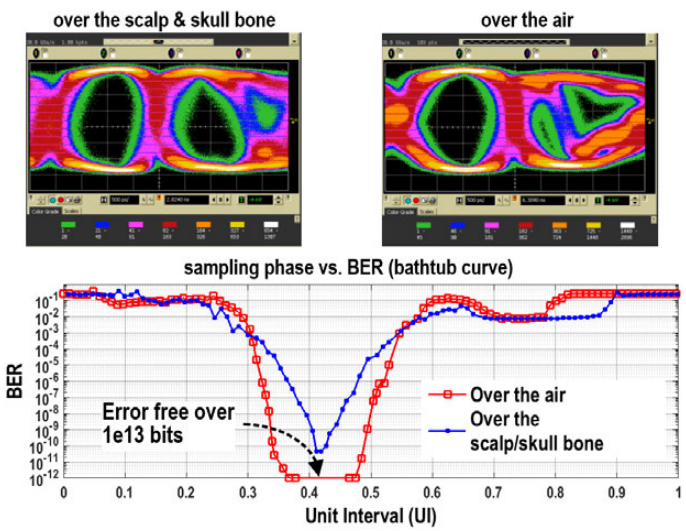
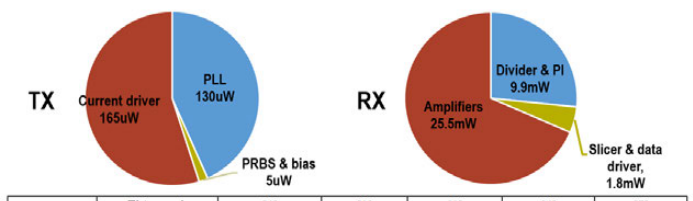


Figure 17.6.5: Measured RX eye diagram and bathtub curve.



	This work		[1]	[2]	[3]	[4]	[5]
Technology	65nm CMOS		65nm CMOS	0.5um CMOS	0.35um CMOS	130nm CMOS	65nm CMOS
Data-rate	200Mb/s		1Mb/s	10.2Mb/s	90Mb/s	10Mb/s	95Kb/s
Modulation	Inductive-coupling		Backscattering	PHM	UWB	UWB	Ultrasound
TX antenna size	10X10mm		6.5X6.5mm	10X10mm	10mmX5mm*	N/A	0.55X0.55mm
Channel distance	11mm		10mm	10mm	N/A	5cm	8.5cm
Channel media	scalp & skull	air	in-vivo	N/A	N/A	N/A	animal tissue
BER	5e-11	<1e-12	<1e-7	6.3e-8	N/A	5e-3	<1e-4
TX power	300uW		13uW	3.52mW**	1.6mW**	100uW**	157uW***
TX FOM	1.5pJ/b		13pJ/b	345pJ/b**	17.8pJ/b**	10pJ/b**	1.65nJ/b***

* Antenna size estimated from PCB photo
** Does not include clock generation power
*** Power estimated by adding peak PA power and clock generator power

Figure 17.6.6: Power breakdown and comparison with state-of-the-art bio-implantable TX.

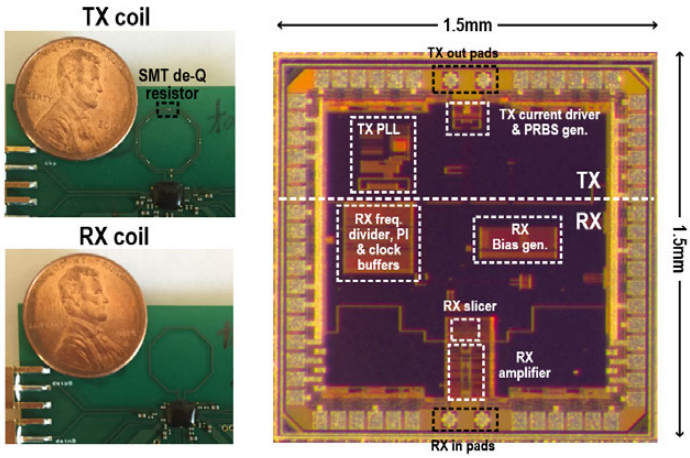


Figure 17.6.7: Coupled inductor layout and chip micrograph.

978-1-5090-4940-0/18/$31.00

17.7 A 330μm×90μm Opto-Electronically Integrated Wireless System-on-Chip for Recording of Neural Activities

Sunwoo Lee, Alejandro J. Cortese, Paige Trexel, Elizabeth R. Agger, Paul L. McEuen, Alyosha C. Molnar

Cornell University, Ithaca, NY

Recording neural activity in live animals *in vivo* poses several challenges. Electrical techniques typically require electrodes to be tethered to the outside world directly via a wire, or indirectly via an RF Coil [1], which is much larger than the electrodes themselves. Tethered implants result in residual motion between neurons and electrodes as the brain moves, and limits our ability to measure from peripheral nerves in moving animals, especially in smaller organisms such as zebra fish or fruit flies. On the other hand, optical techniques, which are becoming increasingly powerful, are nonetheless often limited to subsets of neurons in any given organism, impeded by scattering of the excitation light and emitted fluorescence, and limited to low temporal resolution [2]. Here we present the electronics for an untethered electrode unit, powered by, and communicating through a microscale optical interface, combining many benefits of optical techniques with high temporal-resolution recording of electrical signals.

Figure 17.7.1 illustrates the concept of such a microscale, optically-transduced electrode site with I-V curves of a custom, dual-functioning AlGaAs Photo-Voltaic/Light Emitting Diode (PVLED) unit in its PV and LED modes, which is mounted on top of a CMOS die. About 98% of the time, the PVLED acts as a power source, transducing incoming light into electrical power, providing at least 1μA at ~0.9V. During the remaining ~2% of the time, the PVLED acts as an optical transmitter, emitting optical pulses to transmit encoded measurement data to an external receiver at a longer wavelength. This allows the system to be more compact than the previously reported RF [1] and ultrasonic [3,4] approaches. The bottom-right of the Fig. 17.7.1 depicts the integration of the PVLED on 180nm CMOS where the underlying CMOS circuitry incorporates recording electrodes, amplification, pulse-position encoding, and a PVLED interface to arbitrate power and communications [5]. Since neural tissue is primarily scattering (as opposed to absorbing), optical power and information can propagate well beyond the range of traditional optical imaging, allowing such a system to function at depths greatly exceeding that of imaging, but without the tethers required by most electrodes.

Figure 17.7.2 shows a block diagram of the proposed system and a schematic of the amplifier that boosts the differential signal between the two sensing electrodes that are spaced ~150μm apart to sample the electric fields generated by nearby neurons. Approximately one half of the total current from the PVLED (500nA of 1μA) is used to provide the low noise amplification through the input differential pair (M1 & M2). A pair of NFETs (M3 & M4) act as high-pass active loads: the amplifier output is fed back to the gates from through transistors acting as pseudo resistors and shunted by MOS capacitors. Thus, M3 and M4 provide a low impedance at low frequencies (<<1Hz) but a high resistance in the neural band of interest (>10Hz). Finally, a pair of diode-connected PFETs (M5 & M6) provide a matched load for a controlled mid-band gain, with parallel MOS capacitors setting a low-pass corner at about 10KHz to suppress higher-order aliasing terms. It should be noted that, because the high-pass load would lead to a prohibitively long start-up time while illumination may be transitory, the high-pass resistors are briefly set to a low resistance state during VDD startup, to rapidly calibrate out DC offsets and bias state before switching to their normal high-resistance state. The amplifier and all other circuits are biased from a supply-invariant PTAT-like current source to provide immunity to variations in VDD during fluctuations in illumination or output optical pulse generation.

We have implemented Pulse Position Modulation (PPM) for signal encoding for its high information-per-photon efficiency [6]. Figure 17.7.3 shows that the amplifier drives the pulse-position encoder where a 10KHz relaxation oscillator generates a periodic pulse, which charges capacitor C1 to VDD. After this reset, the capacitor is discharged by one of a pair of differential currents generated from the output of the amplifier. The result is a square-wave whose duty cycle reflects the inverse of the measured voltage. Fixed currents bound the duty cycle to a range between 20% and 80%. A T-flip-flop selects which of the two complementary currents discharges the capacitor at any given time, alternating between clock cycles – like chopping, this allows the separation of signals from fluctuations due to slow-changing light level. The resulting square-wave is passed through a delay-line of current-starved inverters, and edges are combined to generate pulses on both the rising and falling edges of the square wave. The timing of these signals is illustrated in Fig. 17.7.3. A wider pulse disconnects VDD from the PVLED for 1ms, and two other pulses switch a 3-capacitor (1.2pF each) charge pump, switching from a parallel configuration to a series configuration, and connecting to the PVLED to deliver a sharp (<100ns) current pulse. Each cycle of the relaxation oscillator generates two light pulses through the PVLED, one at the beginning of the cycle, and the other between 20μs and 80μs later, where this time difference denotes the input voltage. One extra benefit of using such low duty-cycle optical pulses (<0.1%) is that multiple units could be active simultaneously with minimal overlap between their pulse trains, where PVLEDs with different wavelengths can provide further disambiguation via wavelength multiplexing.

To ensure that the VDD does not drop excessively during the pulsing events (when it is disconnected from the PVLED), 16pF of decoupling capacitance is installed. In addition, because the PVLED can only supply a finite amount of instantaneous current, and to avoid an excessive supply ripple, the charge pump capacitors are recharged slowly over ~10μs. A 20μs minimum pulse spacing ensures that the charge pump is fully charged before each pulse. Finally, to ease the assembly of PVLED and CMOS, a cross-coupled rectifier (polarity corrector) is implemented to ensure the system functionality regardless of the polarity of the PVLED on its pads.

The CMOS circuit is fabricated in 0.18μm CMOS, with an active area of 210μm × 90μm. For testing, we have bonded the CMOS to a PVLED. When we illuminate the hybrid system with ~50nW/μm^2 of band-passed white light (380nm–720nm), which is about 1/6th of the safe limit for brain tissue [7], light pulses are measured as expected as shown in the top-left of Fig. 17.7.4. Figure 17.7.4 also shows that the pulse-position modulated optical pulses can be demodulated to reconstruct a 1KHz test input signal applied to the input electrodes. The system has a transduction gain of ~70ns/μV across 1Hz to 10KHz and the gain compresses for inputs larger than 6mV_{PP} (2.1mV_{RMS}), whereas the input-referred noise floor is ~42μV_{RMS}. We examine the wake-up characteristics of the system for its potential use in pulse-powered environment (as opposed to continuous exposure). Figure 17.7.5 (left) demonstrates that our system wakes up in under 1ms.

Finally, to demonstrate the system's capability to encode real neural signals, we have connected the input electrodes to the ventral nerve cord of an earthworm using probes, with a commercial neural amplifier connected in parallel to provide a reference baseline. Figure 17.7.5 (right) clearly shows that the composite spikes have been accurately encoded in the output optical pulses, even when communication and power are purely optical.

Figure 17.7.6 shows a breakdown of the design by power consumption (top-left) and by silicon area (top-right). As emphasized earlier, the power consumption is dominated by the main amplifier and the charge pump. Area is dominated by the amplifier (for lower flicker noise), LED driver, and decoupling. The bottom of the Fig. 17.7.6 shows a table of comparison against prior art.

Acknowledgements:
The authors would like to thank Chris Xu, Jesse H. Goldberg, Haining Wang, and Changhyuk Lee for critical discussions. Research reported herein was supported by the National Institutes of Health under award number 1R21EY027581 and the Cornell Center for Materials Research with funding from the National Science Foundation MRSEC program (DMR-1719875).

References:
[1] R. R. Harrison, et al., "A Low-Power Integrated Circuit for a Wireless 100-Electrode Neural Recording System," in *IEEE J. Solid-State Circuits*, vol. 42, no. 1, pp. 123-133, Jan. 2007.
[2] W. Yang and R. Yuste, "*In vivo* imaging of neural activity," *Nature Methods*, vol. 14, no. 4, pp. 349-359, April 2017.
[3] S. A. Wirdatmadja, et al., "Wireless optogenetic neural dust for deep brain stimulation," *IEEE 18th International Conference on e-Health Networking, Applications and Services (Healthcom)*, Munich, pp. 1-6, 2016.
[4] D. Seo, et al., "Wireless Recording in the Peripheral Nervous System with Ultrasonic Neural Dust," *Neuron*, vol. 91, pp. 529–539, Aug. 2016.
[5] I. Haydaroglu and S. Mutlu, "Optical Power Delivery and Data Transmission in a Wireless and Batteryless Microsystem Using a Single Light Emitting Diode," *Journal of Microelectromechanical Systems*, vol. 24, no. 1, pp. 155-165, Feb. 2015.
[6] H. Hemmati, et al., "Deep-Space Optical Communications: Future Perspectives and Applications," *Proc. IEEE*, vol. 99, no. 11, pp. 2020-2039, Nov. 2011.
[7] K. Podgorski and G. Ranganathan, "Brain heating induced by near-infrared lasers during multiphoton microscopy," *Journal of Neurophysiology*, vol. 116, pp. 1012–1023, 2016.

 978-1-5090-4940-0/18/$31.00

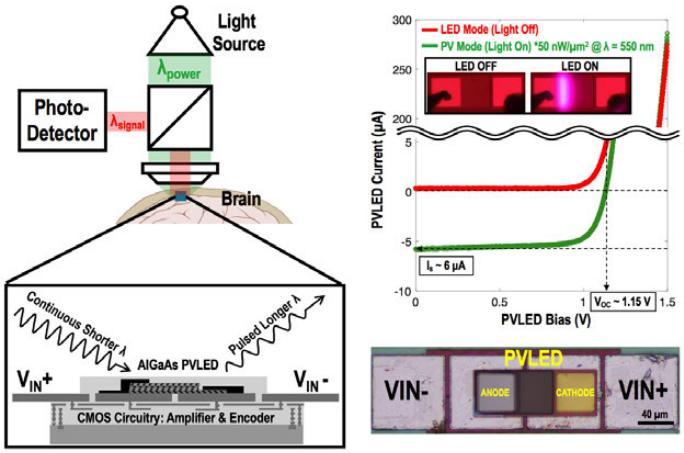


Figure 17.7.1: Left: an envisioned implementation of the system. Top-right: characterization of a micro-scale AlGaAs diode device utilized as PV as well as LED. Bottom-right: a micrograph of an assembled system.

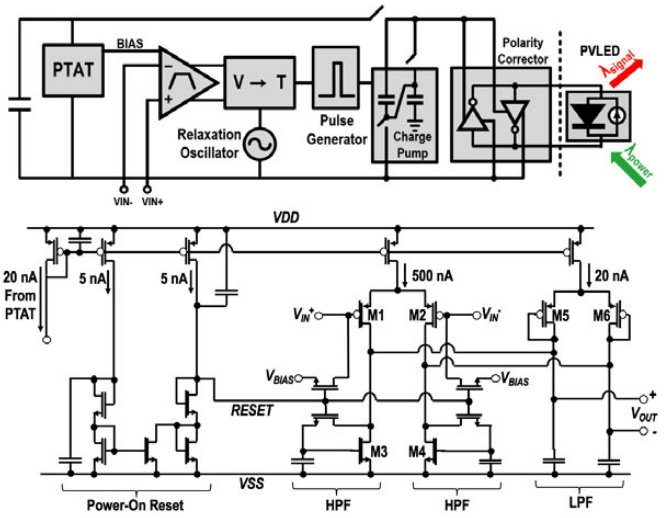


Figure 17.7.2: Top: block diagram of the proposed system. Bottom: schematic of the amplifier used, including the startup circuit and filtering.

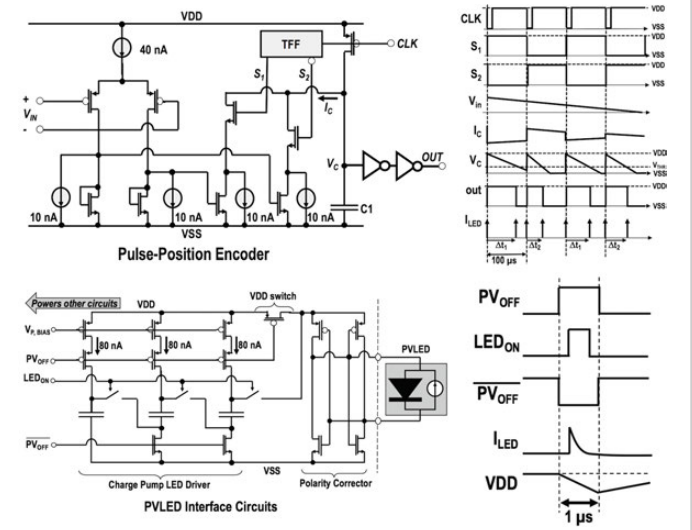


Figure 17.7.3: Top: the PPM encoder and associated timing diagram. Bottom: the pulse generator and its associated timing diagram.

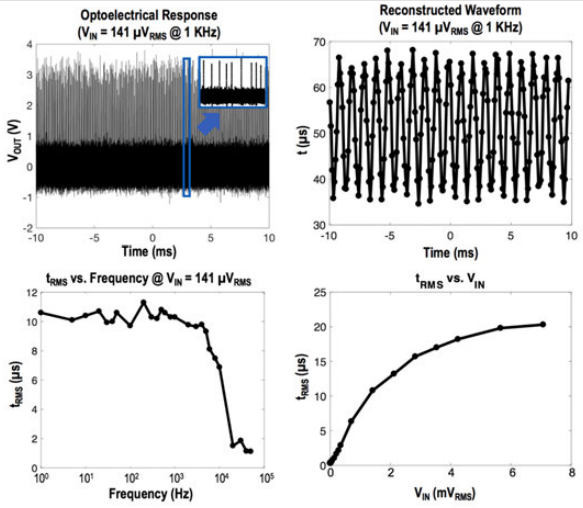


Figure 17.7.4: Example measurement of optical output pulse (top-left), and associated reconstructed 1KHz waveform (top right). Bottom: signal gain as a function of frequency (left) and amplitude (right).

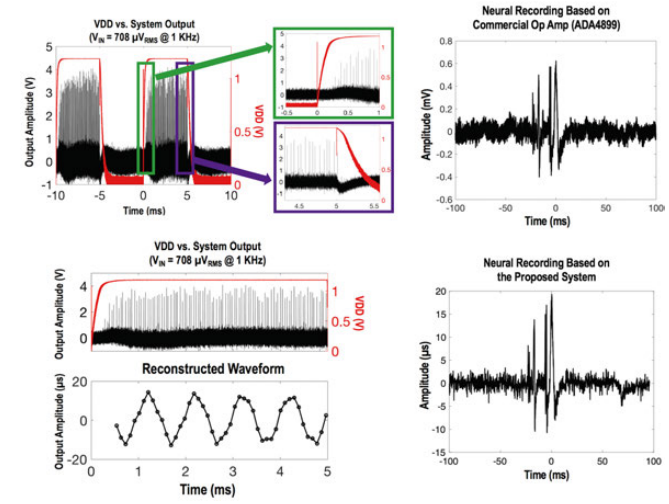
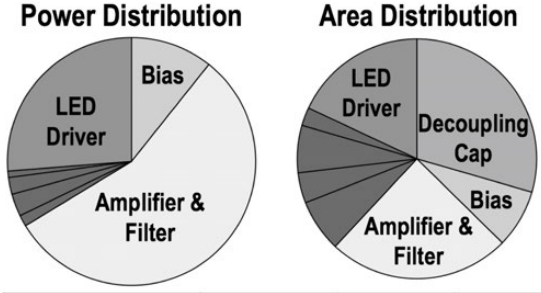


Figure 17.7.5: Left: start-up, showing optical pulses (top) and decoded signal (bottom). Right: neural recording on an earthworm ventral nerve using a commercial amplifier (top) and the presented system (bottom).



	Integrated Neural Interface [1]	Neural Dust [4]	This Work
Wireless Power Source	RF Coil	Ultrasonic	Optical
Gain (dB)	60	N/A	24
Bandwidth (KHz)	5	> 30	10
Noise Floor (μV_{RMS})	5.1	180	42
No. of Channels	100	1	1
Power per Channel	135 μW	0	< 1 μW
Volume per Channel (mm^3)	> 10	2.4	< 0.01

Figure 17.7.6: Top shows the power and area breakdown of the presented work. Bottom tabulates the comparison against prior art.

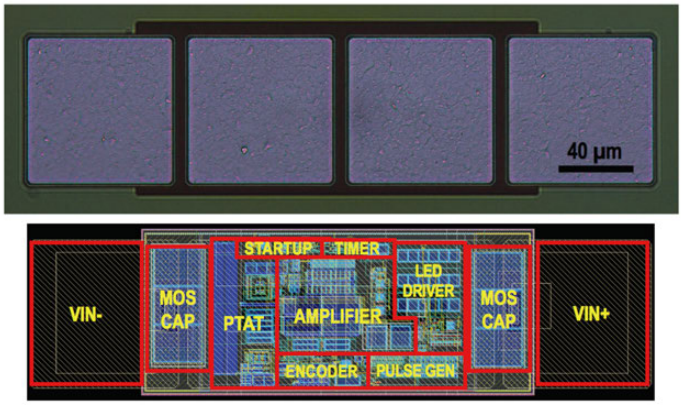


Figure 17.7.7: Top: die micrograph of the 180nm CMOS die, bottom: corresponding layout with circuitry annotated.

978-1-5090-4940-0/18/$31.00

17.8 A 665µW Silicon Photomultiplier-Based NIRS/EEG/EIT Monitoring ASIC for Wearable Functional Brain Imaging

Jiawei Xu[1], Mario Konijnenburg[1], Budi Lukita[1], Shuang Song[2], Hyunsoo Ha[1], Roland van Wegberg[1], Erfan Sheikhi[1], Massimo Mazzillo[3], Giorgio Fallica[3], Walter De Raedt[2], Chris Van Hoof[2,4], Nick Van Helleputte[2]

[1]imec - Holst Centre, Eindhoven, The Netherlands; [2]imec, Leuven, Belgium
[3]STMicroelectronics, Catania, Italy; [4]KU Leuven, Leuven, Belgium

Functional brain imaging is considered a powerful and practical solution for understanding the brain and neurological diseases. While EEG is an established method for non-invasive electrical activity, electrical-impedance tomography (EIT) and near-infrared spectroscopy (NIRS) can additionally measure impedance changes and hemodynamic processes. To facilitate long-term multi-channel brain imaging in a wearable form factor without cabling overhead, there is a need for low-power local amplifiers [1] to support all these modalities. The main principle of optical hemodynamic measurements is to send light pulses into the tissue and measure the reflected light, which is modulated by the oxygen levels in the blood (Fig. 17.8.1). State-of-the-art NIRS ICs typically consume a few mW, primarily for the LEDs to meet the required light sensitivity at the photodiodes (PDs). Silicon photomultipliers (SiPMs) are promising alternatives because they have excellent low-light detection capabilities, speed of response and higher detection efficiency in both visible and near infrared range [2]. Hence, SiPMs allow deeper brain sensing depth and the possibility to sample consistent cerebral regions with larger inter-optode distance. This benefit would significantly reduce the number of NIRS channels and the associated power for a wearable NIRS device. Although SiPMs require a higher bias voltage (~30V) than PDs, they achieve similar NIRS responses with a few hundred times less LED current. This results in a low-power NIRS ASIC and an overall power-efficient system. Existing optical sensing ICs are not suitable for a SiPM because of its large and variable output current. Trimming-based calibration methods [3] suffer from drift over time. Auto-zeroing by swapping an integrator capacitor [4][5] compensates ambient light at the cost of the integrator's headroom. Apart from ambient light, the dynamic range (DR) of the amplifier is also limited by a large NIRS signal, leading to a power-hungry readout.

This paper presents a low-power active ASIC for multimodal (NIRS, EEG, EIT) wearable functional brain imaging (Fig. 17.8.1). The NIRS channel supports both SiPMs and PDs. To cope with 10-to-200µA SiPM output while retaining a sufficient DR with low power, the NIRS channel utilizes ambient light calibration and digital DC servo loops (DSLs). Both techniques work in the background to improve the channel DR. The ASIC is co-integrated with electrodes and/or optodes as an active sensor. Up to 16 ASICs can be distributed locally on the scalp for multi-channel recording. An I²C or SPI bus connects all the ASICs in a daisy chain, which significantly simplifies system connection [1].

The main challenge for wearable NIRS sensing is to design a power-efficient readout with a large DR of >80dB. This DR allows blood oxygen saturation (SpO_2) measurements and to handle ambient light interference. For highly sensitive SiPMs, both ambient light and the NIRS signal can saturate the amplifier or severely reduce its SNR and linearity. In addition, the DC component of the NIRS signal is also critical for SpO_2 detection. Instead of a power-hungry high-resolution readout, we propose a transimpedance amplifier (TIA) with variable gain of 1k-to-100kΩ, followed by a low-power 12-bit SAR ADC (Fig. 17.8.2). A single readout channel records from 2 LEDs (with different wavelengths) in a time-interleaved manner. The TIA is equipped with an ambient light calibration loop and two digital DSLs (one for each wavelength) to improve the DR. When the LEDs are off, a SAR-controlled 7-bit current DAC (I-DAC) coarsely senses and compensates the ambient light current in 10ms. The I-DAC code is held until the next calibration, which is repeated at a programmable interval. The SAR-based calibration is fast and not limited by the DR of the TIA. After calibration, the readout samples residual ambient light, which is then subtracted from the NIRS samples in the next phase when the LEDs are pulsed. In this phase, the digital DSLs extract the DC components of the sampled NIRS signals, respectively. Two FIR based 0.4Hz LPFs control two I-DACs to compensate the NIRS current. These I-DACs are synchronously pulsed with the LEDs to save power. Utilizing the SAR-based calibration and the digital DSL improves the channel's DR to 87dB. Both NIRS and ambient light signals are reconstructed by combining the outputs of the ADC and the I-DACs.

One active sensor-based EEG channel consists of two ASICs. The EEG is sampled at 500sps with 14.2b ENOB. The chopper instrumentation amplifier (IA) (fc=2kHz) (Fig. 17.8.1) contains a gm-C based analog DSL to reject max. 300mV electrode offset. Active shielding boosts the input impedance through a built-in buffer. This buffer is reused for fast settling by charging C_{ext}. At the system level, a common-mode feedforward (CMFF) [1] between multiple ASICs improves their CMRR to maximum 100dB by driving each IA's analog DSL with a buffered CM input.

Prior EIT ASICs suffer from limited input impedance and 1/*f* noise of the current generator (CG) [4]. The input impedance is often reduced by chopping or high-pass filtering. Hence, this work moves the EIT input modulator to the output of the TC stage (Fig. 17.8.1), while high-pass filters are replaced by an analog DSL. To relax the BW of the TI stages and associated power, the up-modulated EIT input signal (1kHz-to-1MHz) is demodulated to fc=2kHz prior to the TI stages [6]. 1/*f* noise of the CG can be mitigated by dynamic element matching (DEM), where all current mirrors are chopped sequentially [6]. However, 1/f noise of the CG's reference current is not compensated. In this work, both the CG and the bandgap reference employ DEM to reduce 1/*f* noise (Fig. 17.8.3), leading to 3mΩ/√(Hz) sensitivity including the CG noise. The EIT channel can also be reconfigured (via SW_ETI in Fig.1) to measure electrode tissue impedance (ETI) to monitor the lead connection quality and/or for motion artifact reduction.

The IC is fabricated in 0.18µm CMOS (Fig. 17.8.7). In simultaneous NIRS/ EEG/EIT recording mode, the ASIC consumes 665µW (including LED power). In Fig. 17.8.4, the first plot shows the NIRS channel output during ambient light calibration, which is reached in 7 steps. The top-right plot shows that enabling the digital DSL prevents channel saturation when a sinusoidal input current (20µA@2Hz) is superimposed on a 60µA DC current. The next two plots on the right show the I-DAC code of the DSL and the reconstructed output signal. In the bottom-left plots, the EEG channel (obtained from 2 ASICs) has 120nV/√(Hz) input-referred noise and 720MΩ input impedance at 50Hz. In the bottom-right plot, the NIRS channel shows 0.2-to-0.9nA/√(Hz) input current noise at 10Hz. Figure 17.8.5 shows a recording with a SiPM and two-wavelength (735/850nm) LEDs from the forehead clearly showing heart rate. Blood oxygenation saturation reduces when the subject was holding breath for about 40 seconds. The bottom plot shows EEG alpha waves measured with polymer (dry) electrodes. Figure 17.8.6 compares the IC performance. This is the first sub-mW ASIC supporting SiPMs for wearable NIRS sensing, where the NIRS channel consumes 287µW including 2 LEDs at 75µW each. Even including the measured SiPM power of 1.5mW, the proposed NIRS channel is still 3× more power efficient than prior art. Thanks to the SAR based calibration and digital DSLs, the NIRS channel realizes 87dB DR and is able to sense both ambient light (40µA max) and NIRS (400µA max).

Acknowledgements:
This project has received funding from the European Research Council (ERC) under the European Union's Horizon 2020 research and innovation programme (grant agreement No 692470)

References:
[1] J. Xu, et al., "A 15-channel digital active electrode system for multi-parameter biopotential measurement", *IEEE J. Solid-State Circuits*, pp. 2090-2100, Sept. 2015.
[2] R. Pagano, et al., "Improvement of sensitivity in continuous wave near infrared spectroscopy systems by using silicon photomultipliers." *Biomedical Optics Express 7(4)*, pp. 1183–1192, 2016.
[3] U. Ha, et al., "A wearable EEG-HEG-HRV multimodal system with simultaneous monitoring of tES for mental health management," *IEEE Trans. on Biomedical Circuits and Systems,* pp. 758-766, Dec. 2015.
[4] M. Konijnenburg, et al., "A multi(bio)sensor acquisition system with integrated processor, power management, 8 x 8 LED drivers, and simultaneously synchronized ECG, BIO-Z, GSR, and two PPG readouts," *IEEE J. Solid-State Circuits*, pp. 2584-2595, Nov. 2016.
[5] P. Schönle, et al., "A power-efficient multi-channel PPG ASIC with 112dB receiver DR for pulse oximetry and NIRS," *IEEE Custom Integrated Circuits Conference,* pp. 1-4, 2017.
[6] H. Ha, et al., "A bio-impedance readout IC with frequency sweeping from 1k-to-1MHz for electrical impedance tomography," *IEEE Symposium on VLSI Circuits*, pp. 174-175, 2017.

 978-1-5090-4940-0/18/$31.00

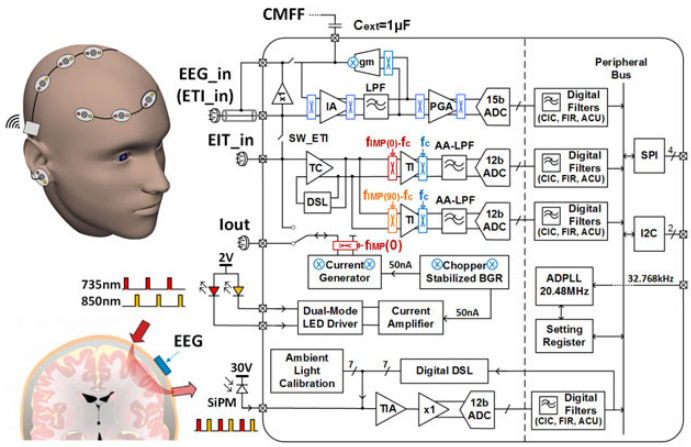


Figure 17.8.1: Active NIRS/EEG/EIT sensor and ASIC block diagram.

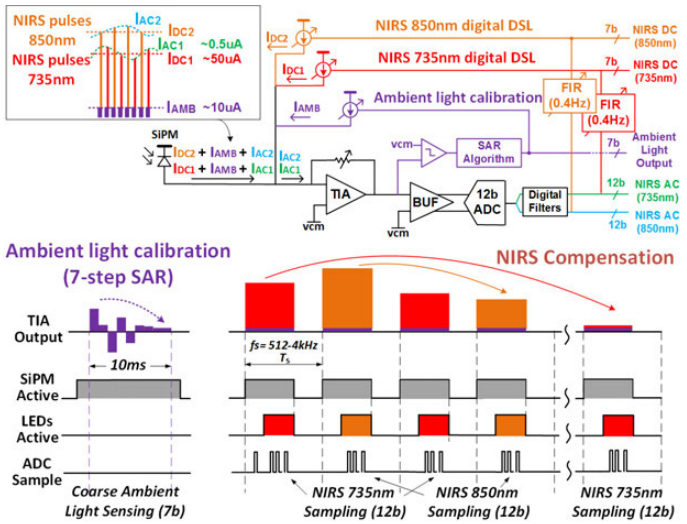


Figure 17.8.2: Ambient light calibration and digital DSLs.

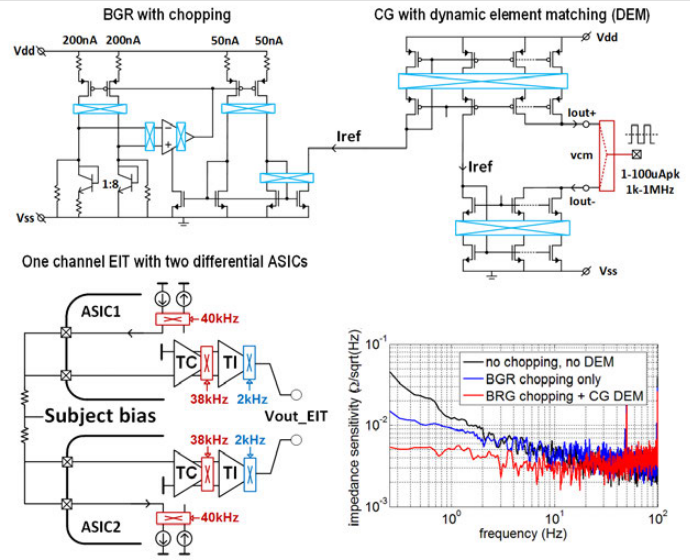


Figure 17.8.3: Low noise bandgap reference (BGR) and current generator (CG).

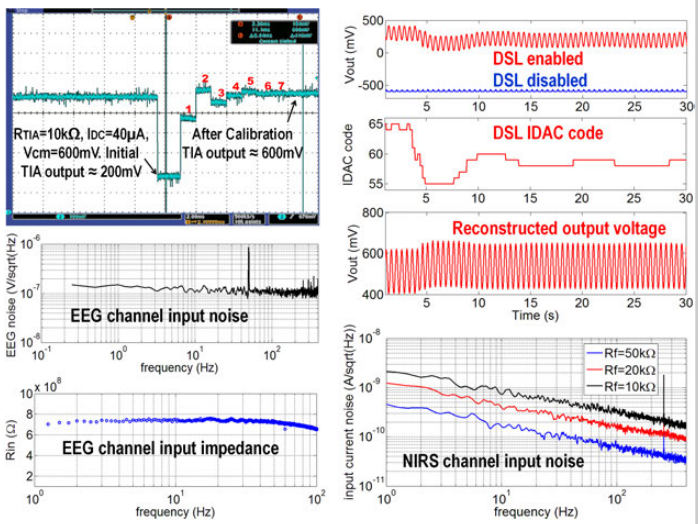


Figure 17.8.4: IC measurement results.

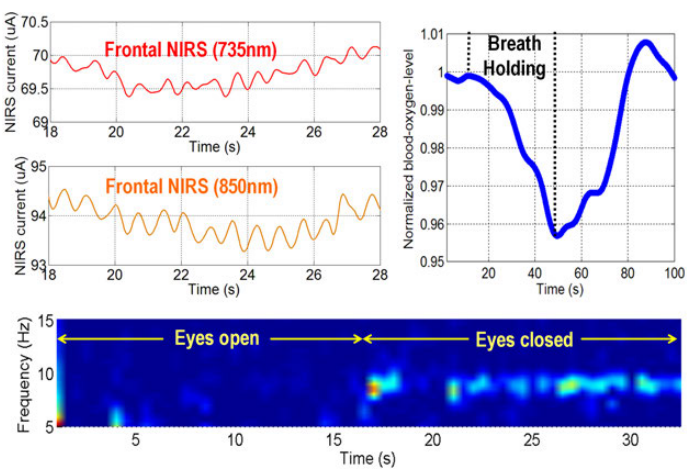
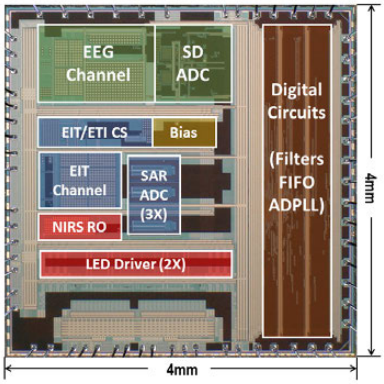


Figure 17.8.5: Biomedical measurement results.

Parameters		[3]	[5]	[4]	This work
Supply		1.2V/3.3V	1.2V/3.3V	1.2V/3.3V	1.2V/3.3V
NIRS optode		PD only	PD only	PD only	SiPM, PD
Ambient light compensation		6 bit 12.8nA max	2.6 bit 10μA max	5 bit 10μA max	7 bit 40μA max
NIRS compensation		no	yes, trimming	no	yes (400μA max) background
NIRS sampling rate		500-4kSps	100-2kSps	128Sps	2-512Sps
NIRS dynamic range		--	112dB	89dB	87dB
LED current	Peak	5mA	50mA	0-160mA	280uA
	Average	n/a	2.2mA	2mA	75uA
NIRS power (with LEDs)		--	5.65mW	6.14mW	282μW
EEG noise (0.5-100Hz)		1.2μVrms	--	0.6μVrms	1.2μVrms
EEG input impedance		1GΩ@60Hz	--	500MΩ@50Hz	720MΩ@50Hz
CMRR		132dB	--	110dB	100dB
Electrode offset rejection		200mV	--	300mV	300mV
EEG ADC		11b	--	13.5 ENOB	14.2 ENOB
EEG power		8.8μW (IA)	--	49μW	43μW
EIT noise		--	--	3mV/√(Hz)	3mV/√(Hz)@3Hz with CG noise
EIT power		--	--	46μW	45μW

Figure 17.8.6: Comparison table.



ASIC Power Consumption Breakdown

	Current (μA)	Voltage (V)	Power (μW)
NIRS RO	110	1.2	132
NIRS DACs	~39.3		47
EEG RO	35.8		43
EIT RO	37.3		45
EIT CG	20		24
Bias	16.2		20
VDD Analog	259	1.2	311
VDD Digital	170	1.2	204
VDD LEDs	75	2.0	150
ASIC Power			665
SiPM Power*	~50	30	1500

*measured in normal environment
LEDs and SiPM shielded by black tape

Figure 17.8.7: Chip micrograph (4mm×4mm).

978-1-5090-4940-0/18/$31.00

17.9 A Recursive-Memory Brain-State Classifier with 32-Channel Track-and-Zoom $\Delta^2\Sigma$ ADCs and Charge-Balanced Programmable Waveform Neurostimulators

Gerard O'Leary[1], M. Reza Pazhouhandeh[1], Michael Chang[1], David Groppe[2], Taufik A. Valiante[3], Naveen Verma[4], Roman Genov[1]

[1]University of Toronto, Toronto, Canada
[2]Krembil Neuroscience Center, Toronto, Canada
[3]Toronto Western Hospital, Toronto, Canada
[4]Princeton University, Princeton, NJ

The advancement of closed-loop neuromodulation for treating neurological disorders demands: (1) analog circuits monitoring the brain activity uninterruptedly even during neurostimulation, (2) energy-efficient high-efficacy processors for responsive, adaptive, personalized neurostimulation, and (3) safe neurostimulation paradigms with rich spatio-temporal stimuli for controlling the brain's complex dynamics. This paper presents an implantable neural interface processor (NURIP) that addresses these requirements - it performs brain state classification for reliable seizure prediction and contingent seizure abortion.

As shown in Fig. 17.9.1 (center) NURIP includes 32 bidirectional channels, each with an arbitrary-waveform neurostimulator and an input-tracking $\Delta^2\Sigma$-based analog-to-digital converter. The ADC automatically detects any sharp transitions in the intracranial electroencephalogram (iEEG), such as those due to a stimulation artifact, and shifts its high-resolution input range to zoom to the input signal, anywhere within the power rails, as depicted in Fig. 17.9.1 (top, left). Compared to conventional amplifiers, it experiences no blind intervals caused by sharp input transitions. The input digital stage features an autoencoder neural network for both iEEG spatial filtering and dimensionality reduction. Dedicated feature extraction blocks are implemented for univariate (signal-band energy, SE) and multivariate (phase locking value, PLV, and cross frequency coupling, CFC) neural signal processing. The proceeding support vector machine (SVM) accelerator employs these features for brain state classification. A general-purpose CPU facilitates additional custom feature extraction and system control. In response to the detection of a pathological brain state, an appropriate modulation waveform is generated to control the operation of the current-mode neurostimulator.

Figure 17.9.2 (top) shows the block-diagram of the $\Delta^2\Sigma$ neural recording channel with the input range moving to track the input signal. In the event of a sudden change in the input due to stimulation artifacts, the $\Delta^2\Sigma$ modulator saturates and outputs either many ones or many zeros due to the inability of the in-channel DAC to quickly ramp up/down, to follow the artifact signal. A digital integrator accumulates the output bit-stream of the $\Delta^2\Sigma$ modulator. Upon crossing predefined window levels, REF_H or REF_L, the in-channel feedback multi-bit DAC increments the step size up or down, by a factor of two. This procedure continues until the in-channel multi-bit DAC catches up with the input signal of the $\Delta^2\Sigma$ modulator. Next, the saturation-detection block and the in-channel multi-bit DAC are reset back to the minimum step size – the ADC zooms in at a new DC offset. Figure 17.9.2 also illustrates the output voltage of the in-channel DAC during input tracking (bottom, left) and an experimentally measured output waveform of the in-channel DAC with a 1kHz sinusoidal signal superimposed on a step waveform (bottom, middle). Figure 2 (bottom, right) shows an experimentally measured output FFT and an *in vivo* mouse recording. An SNDR of 70dB and ENOB of 11.3b at an OSR of 1000 have been measured. The current-mode stimulator DAC is reused during the recording phase, and comes at no extra cost in area.

Figure 17.9.3 illustrates both spatial and spectral filtering of the input neural signals. An auto-encoder neural network performs spatial filtering and dimensionality reduction from 32 recording channels to 4 weighted combinations (such as in principal component analysis), reducing the processing requirements by 8×. Sixteen hardware-based circular buffers (including the shown 4) enable online processing of neural recording streams, with a 256-sample window. They are mapped to 8kB of address space within 64kB of global SRAM. Incoming samples are mapped to varying physical addresses whereas the corresponding virtual addresses used by the system are fixed. The output stream is band-pass filtered using a global configurable FIR filter, which utilizes coefficients symmetry to halve the number of MACs.

A subsequent array of three configurable neural signal feature extractors, shown in Fig. 17.9.4 (top), enables custom patient-specific processing to maximize classifier performance. The absolute output value of each bandpass filter is taken as a measure of signal energy. As shown in Fig. 17.9.4 (top, left), a specific power signature in the δ, θ, α, β and γ iEEG bands is indicative of a seizure. The phase locking value (PLV) extractor shown in Fig. 17.9.4 (top, middle) detects phase difference precursors of an upcoming seizure onset. An analytic signal is obtained using a global Hilbert FIR filter along with the dual-core CORDIC block to extract the phase difference between two input channels. Efficient resource reuse results in an overall area reduction of 9× versus the state of the art [1]. Cross-frequency coupling (CFC) is a key mechanism in neuronal computation and its abnormal appearance can serve as a spatial biomarker for seizure detection. Low-frequency brain rhythms modulate high-frequency activity and the resulting envelope is extracted with re-use of PLV hardware. CFC is then computed as the synchrony between the extracted envelope and a low-frequency modulating signal [5]. The ensemble of these three biomarkers yields a uniquely high-dimensional feature space for the classifier.

In the case of seizure prediction, onset biomarkers are subtle and can occur minutes before seizure onset. This presents a challenge in processing and memory requirements for implantable devices. NURIP introduces the Exponentially Decaying-Memory Support Vector Machine (EDM-SVM) accelerator for efficient classification of long-term temporal patterns. The EDM-SVM input stage (Fig. 17.9.4 (bottom, right)) recursively captures a feature's history across multiple timescales, up to multiple minutes, using a combination of memory decay rates to enable the learning of temporal relationships. An efficient implementation using shift and add operations is implemented by constraining decay coefficients to powers-of-two. The proceeding SVM accelerator core allows the selection of linear, polynomial and radial-basis function (RBF) kernels to trade off between performance, energy and memory usage. As the EDM is updated every sample, classification can be performed continuously to minimize detection latency.

Upon the detection of a seizure, the integrated digitally charge-balanced neurostimulation waveform generator responds as demonstrated in Fig. 17.9.5. To prevent electrode and tissue damage due to stimulus charge imbalance, binary exponential charge recovery (BECR) ensures the net stimulus integral is zero after arbitrary waveform stimulation or when safe limits have been exceeded. A 15kSps function generator is efficiently implemented with the reuse of the dual-core CORDIC and MAC blocks, enabling the generation of sums or products of sinusoids. Arbitrary waveform replay is supported by 3MSps streaming from on-chip SRAM.

Figure 17.9.6 outlines NURIP's classifier performance using the EU intracranial EEG database [6]. The extracted feature space used for classification consists of 125 dimensions derived through offline feature selection and is constrained to <200 support vectors by the on-chip SRAM. A sensitivity of 100% and a false positive rate (FPR) of 0.81 per hour have been achieved. A classification rate of 4Hz requires a power consumption of 674.4 µW with a nominal voltage of 1.2V and an operational frequency of 10MHz. The SoC micrograph and the channel floorplan are shown in Fig. 17.9.7 (top). The chip is compared with the state of the art both in terms of the channel performance and digital processing performance in Fig. 17.9.7 (bottom).

Acknowledgements:
The authors thank the Canadian Microelectronics Corporation (CMC) for IC fabrication and Jintao Zhang for his assistance. The first two authors contributed equally (Gerard O'Leary, NURIP, and M. Reza Pazhouhandeh, AFE).

References:
[1] H. Kassiri, et al., "All-wireless 64-channel 0.013 mm 2/ch closed-loop neurostimulator with rail-to-rail DC offset removal," *ISSCC Digest Tech. Papers*, pp. 452–453, Feb. 2017.
[2] K. H. Lee and N. Verma, "A Low-Power Processor With Configurable Embedded Machine-Learning Accelerators for High-Order and Adaptive Analysis of Medical-Sensor Signals," *IEEE JSSC*, vol. 48, no. 7, pp. 1625–1637, 2013.
[3] M. A. Bin Altaf, et al.,"A 16-Channel Patient-Specific Seizure Onset and Termination Detection SoC With Impedance-Adaptive Transcranial Electrical Stimulator," *IEEE JSSC*, vol. 50, no. 11, pp. 2728–2740, Nov. 2015.
[4] W. M. Chen, et al., "A Fully Integrated 8-Channel Closed-Loop Neural-Prosthetic CMOS SoC for Real-Time Epileptic Seizure Control," *IEEE JSSC*, vol. 49, no. 1, pp. 232–247, Jan. 2014.
[5] G. O'Leary, et al., "Low-latency VLSI architecture for neural cross-frequency coupling analysis," *IEEE Engineering in Medicine and Biology Conference*, pp. 2247–2250, 2017.
[6] M. Ihle, et al., "EPILEPSIAE – A European epilepsy database," *Computer Methods and Programs in Biomedicine*, vol. 106, no. 3, pp. 127–138, June 2012.

 978-1-5090-4940-0/18/$31.00

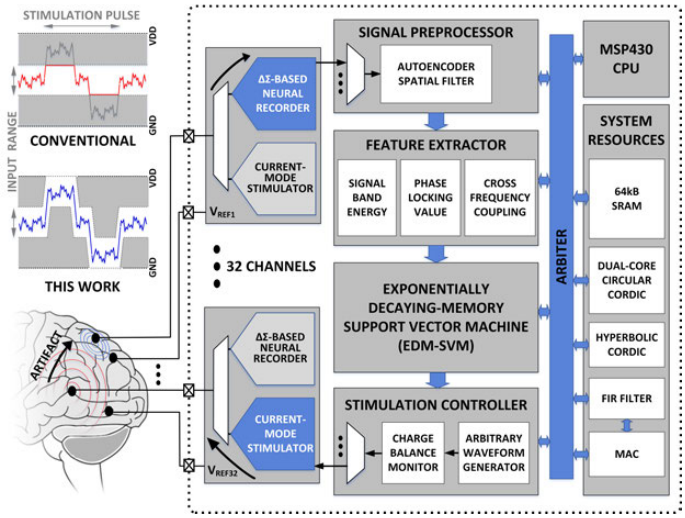


Figure 17.9.1: Neural interface processor (NURIP) architecture for responsive neuromodulation.

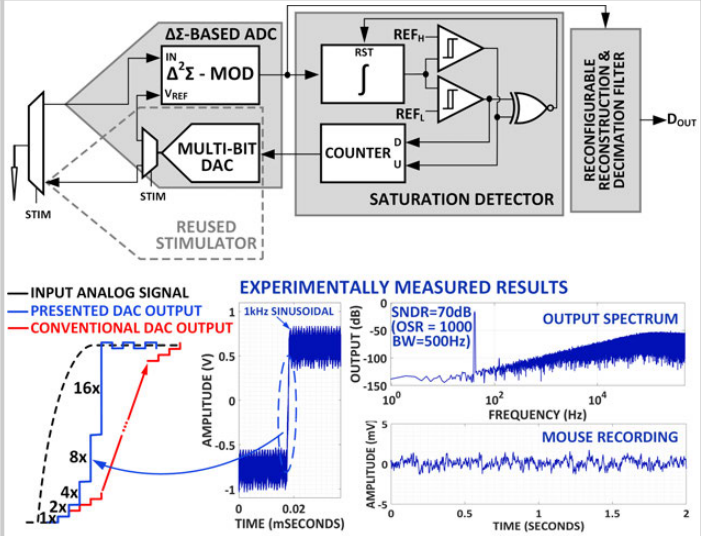


Figure 17.9.2: Adaptive-input-range analog front end and its experimentally measured results.

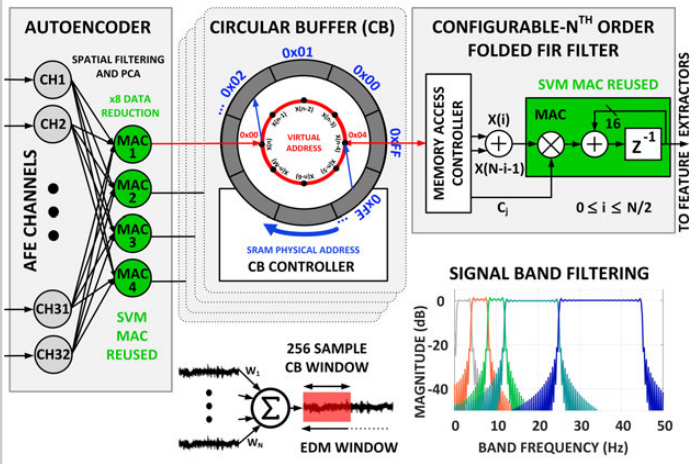


Figure 17.9.3: Autoencoder neural network for spatial filtering and FIR spectral filtering.

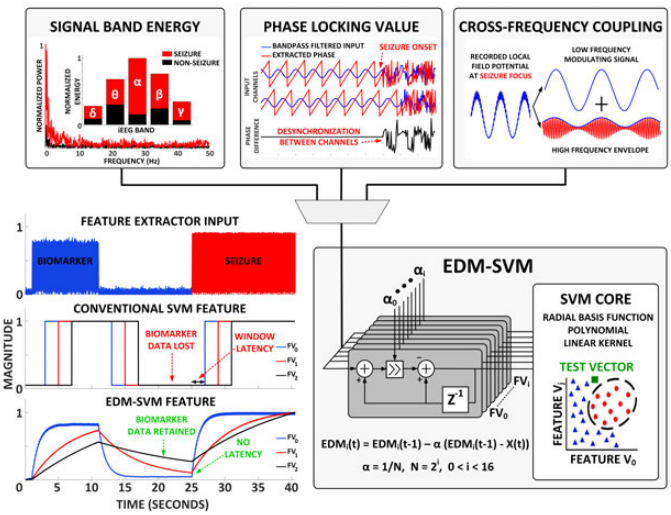


Figure 17.9.4: Feature extraction cores and Exponentially Decaying Memory SVM (EDM-SVM).

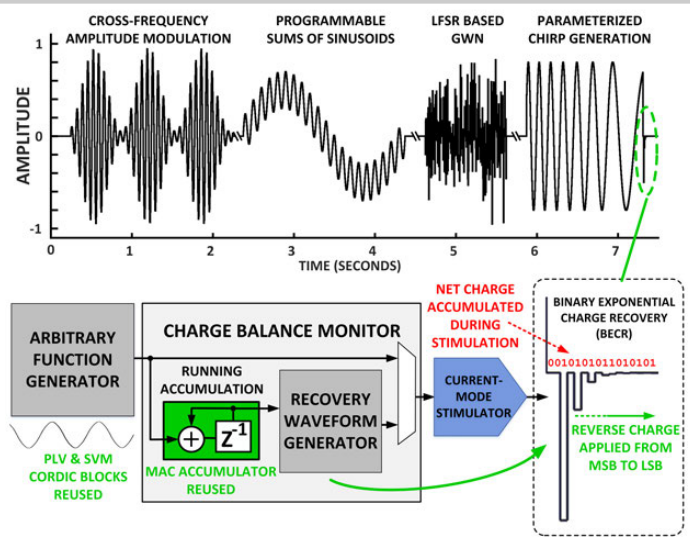


Figure 17.9.5: Experimentally measured arbitrary waveform generation and digital charge balancing.

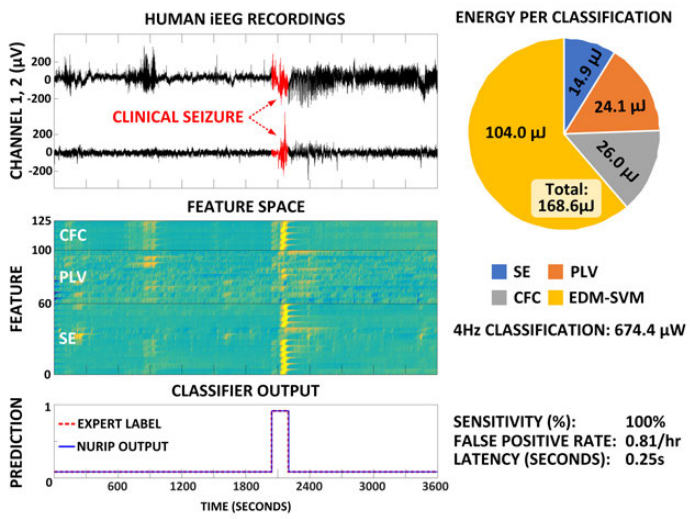
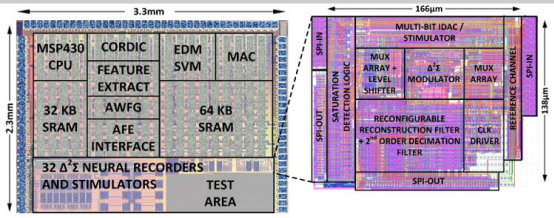


Figure 17.9.6: Experimentally computed feature space and EDM-SVM classifier output in offline human iEEG recordings.



	JSSC'13 [2]	JSSC'14 [4]	JSSC'15 [3]	ISSCC'17 [1]	THIS WORK
TECHNOLOGY (µm)	-	0.18	0.18	0.13	0.13
FEATURE EXTRACTION	CPU	FFT, Entropy	SE	PLV	**PLV, CFC, SE, CPU**
CLASSIFIER	SVM	LLS	D²A-LSVM	Threshold	**EDM-SVM**
SAMPLE MEMORY	6s	96 samples	3s	64 samples	**∞ (EDM)**
LATENCY (s)	2	0.8	1	-	**<0.1**
MEMORY (kB)	64	0	64	0	**96**
ENERGY/CLASS. (µJ)	273	77.91	2.73	-	168.6
STIM. ARTIFACT TOL.		NO	NO	NO	**YES**
INPUT TRACKING	-	NO	NO	LINEAR	**EXP.**
BLIND INTERVAL	-	YES	YES	mSEC	**µSEC**
BANDWIDTH (Hz)	-	0.1-7k	0.5-100	0.01-500	**0.01-1000**
POWER/CH (µW)	-	53.7	1.62	1.26[1]	1.26[2]
AREA/CH (mm²)	-	-	0.7	0.013	0.023
NOISE BW (Hz)	-	0.5-7k	0.5-100	0.1-500	0.1-500
IRN (μV_{RMS})	-	5.23	0.90	1.6	1.6
NEF	-	1.77	3.29	2.86	2.86
MAX OFFSET (mV)	-	AC-Coupled	AC-Coupled	Rail-to-Rail	Rail-to-Rail
# STIMULATORS	-	1	1	64	32
MAX STIM (mA)	-	0.03	-	1.35	**3**
WAVEFORM GEN.	-	Bi-phasic	Bi-phasic	-	**AWG**
CHARGE BALANCING	-	-	PVTES	-	**BECR**

[1]Power without digital filter blocks.
[2]Power for LFPs without saturation detector and digital filter blocks

Figure 17.9.7: NURIP SoC micrograph and comparison table.

978-1-5090-4940-0/18/$31.00 ©2018 IEEE

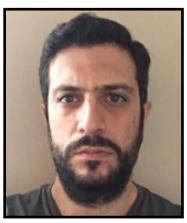
Session 18 Overview:
Adaptive Circuits and Digital Regulators
DIGITAL CIRCUIT TECHNIQUES SUBCOMMITTEE

Session Chair:
Dennis Sylvester
University of Michigan, Ann Arbor, MI

Associate Chair:
Koji Hirairi
Sony LSI Design, Kanagawa, Japan

Subcommittee Chair: ***Edith Beigné,*** CEA-LETI, Grenoble, France

This session focuses on circuits that detect and adapt to various types of process-voltage-temperature-aging (PVTA) variations, along with the latest progress in digital low-dropout (LDO) voltage regulators. The presented adaptive circuits demonstrate rapid and accurate voltage-droop detection to perform instruction throttling, clock period stretching via use of a critical-path replica within the phase-locked loop (PLL) and (buck) voltage regulator, and closed-loop continuous-body-bias for temperature, process, and aging compensation. There are five digital LDO papers that introduce a range of new techniques, such as an analog-assisted NMOS power transistor, a beat-frequency-based adaptive-sampling scheme, and the replacement of traditional power-transistor devices with a switched-capacitor resistor. These papers also describe a distributed LDO array that improves IR drop and response time, as well as an LDO that is designed to minimize pass-through of voltage ripple from a preceding switching regulator.



8:30 AM

18.1 Droop Mitigation Using Critical-Path Sensors and an On-Chip Distributed Power Supply Estimation Engine in the z14™ Enterprise Processor

C. Vezyrtzis, IBM Research, Yorktown Heights, NY

In Paper 18.1, IBM describes a new set of techniques to improve latency in voltage-droop mitigation, including slope calculation and unit-to-unit communication, that double the performance gains over their previous-generation design. They also describe a new method to estimate voltage noise in a large processor using local activity detectors. These techniques are validated on their latest 14nm enterprise class processor.

9:00 AM

18.2 A Combined All-Digital PLL-Buck Slack Regulation System with Autonomous CCM/DCM Transition Control and 82% Average Voltage-Margin Reduction in a 0.6-to-1.0V Cortex-M0 Processor

X. Sun, University of Washington, Seattle, WA

In Paper 18.2, the University of Washington presents a combined buck converter and PLL design in 65nm that uses a programmable canary-based oscillator to reduce the impact of supply droop on timing margin. Combined with all-digital autonomous switching between continuous- and discontinuous-conduction modes, the design reduces margin by 90% and offers 95% peak converter efficiency across 0.6-1.0V.

 978-1-5090-4940-0/18/$31.00

9:30 AM

18.3 A 2.5µW 0.0067mm² Automatic Back-Biasing Compensation Unit Achieving 50% Leakage Reduction in FDSOI 28nm over 0.35-to-1V V_{DD} Range

A. Quelen, CEA-LETI-MINATEC, Grenoble, France

In Paper 18.3, CEA-LETI and STMicroelectronics describe a low-overhead compensation unit that provides continuous and body-bias values to reduce leakage at fixed frequency across process/temperature/aging variation in 28nm FDSOI technology. The unit is 92× smaller and 4× lower power than prior work in body-bias generation, while offering 100ms response time across a wide supply voltage range of 0.35-1V.

9:45 AM

18.4 A 0.4V 430nA Quiescent Current NMOS Digital LDO with NAND-Based Analog-Assisted Loop in 28nm CMOS

X. Ma, University of Macau, Macau, China and University of Electronic Science and Technology of China, Chengdu, China

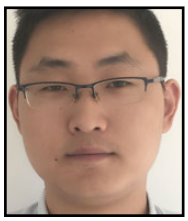
In Paper 18.4, the University of Macau presents a digital LDO design that exploits an NMOS power device for intrinsic response to output voltage droops, and employs a high-pass analog path with a voltage-doubled NAND gate to increase conductance during load transients. Together, these techniques enable a 5.1fs speed FoM in a 0.0055mm² 28nm LDO that provides 20mA output current with 50mV minimum dropout.

10:15 AM

18.5 A Fully Integrated 40pF Output Capacitor Beat-Frequency-Quantizer-Based Digital LDO with Built-In Adaptive Sampling and Active Voltage Positioning

S. Kundu, University of Minnesota, Minneapolis, MN

In Paper 18.5, the University of Minnesota and Cisco describe a new approach to balancing the demands of response time (speed) and quiescent current (power) in a digital LDO. An adaptive sampling frequency is generated via a beat-frequency quantizer. The resulting 0.0374mm² design in 65nm improves settling time by 25× and voltage droop by 5× over a fixed sampling frequency design and uses a small 40pF output capacitor.



10:45 AM

18.6 A 500mA Analog-Assisted Digital-LDO-Based On-Chip Distributed Power Delivery Grid with Cooperative Regulation and IR-Drop Reduction in 65nm CMOS

Y. Lu, HKUST, Hong Kong, China

In Paper 18.6, the Hong Kong University of Science and Technology presents a 3×3 mesh of analog-assisted LDOs that work together to supply 500mA to an unevenly distributed load. Using interleaved phases for clocking, the 9 LDOs work together to improve response speed. The design is fabricated in 65nm CMOS process and employs a 0.9nF on-chip capacitor.

11:15 AM

18.7 A Sub-1.55mV-Accuracy 36.9ps-FOM Digital-Low-Dropout Regulator Employing Switched-Capacitor Resistance

L. G. Salem, University of California, San Diego, La Jolla, CA

In Paper 18.7, the University of California San Diego describes a fully digital LDO that uses a switched-capacitor resistance to replace the PMOS switch array in conventional digital LDOs. The 65nm CMOS design achieves 99.3% peak current efficiency at 3mA I_{LOAD}, 34.3ps speed FoM and less than 1.5mV steady state error.

11:45 AM

18.8 A High-Efficiency and Fast-Transient Digital-Low-Dropout Regulator with the Burst Mode Corresponding to the Power-Saving Modes of DC-DC Switching Converters

Y.-S. Ma, National Chiao Tung University, Hsinchu, Taiwan

In Paper 18.8, National Chiao Tung University presents a fully digital LDO in 40nm using a non-linear switch control technique to improve current efficiency by decreasing quiescent current and reducing switching power consumption with variable-switching frequency control. It achieves peak current efficiency of 99.8% at 20mA I_{LOAD}, under 10uA quiescent current, 6mV voltage ripple, and 40mV droop (1.3µs transient response time) on a 1-to-20mA load step.

18.1 Droop Mitigation Using Critical-Path Sensors and an On-Chip Distributed Power Supply Estimation Engine in the z14™ Enterprise Processor

Christos Vezyrtzis[1], Thomas Strach[2], Pierce I-Jen Chuang[1], Preetham Lobo[3], Richard Rizzolo[4], Tobias Webel[2], Pawel Owczarczyk[4], Alper Buyuktosunoglu[1], Ramon Bertran[1], David Hui[4], Susan M. Eickhoff[4], Michael Floyd[5], Gerard Salem[6], Sean Carey[4], Stelios G. Tsapepas[4], Phillip J. Restle[1]

[1]IBM Research, Yorktown Heights, NY; [2]IBM STG, Boeblingen, Germany
[3]IBM STG, Bangalore, India; [4]IBM STG, Poughkeepsie, NY; [5]IBM STG, Austin, TX
[6]IBM STG, Essex Junction, VT

Enterprise server processor designs, which operate at extreme high frequencies and power envelopes, depend critically on power supply noise mitigation techniques. With supply voltage scaling, very high current draws, and broad usage of clock gating, advanced solutions are needed for next-generation products to minimize droop mitigation response time, which can be defined as the latency from when a dangerous droop begins until a countermeasure is effective.

In this paper, we present a suite of new noise mitigation techniques that have been implemented in the 14nm SOI z14™ (die photograph shown in Fig. 18.1.7) enterprise processor to more quickly and accurately sense, predict, and react to droop events as compared with prior product generations. These new techniques use information from critical-path-monitor (CPM) [1] sensors in each core or from an on-chip distributed real-time estimator engine of the power-supply network. We call this the cycle-by-cycle-activity (CBCA) estimator engine [2] since power is tracked by core activity counters to determine local current. To optimize bandwidth and memory latency, the cores operate synchronously with the cache / multi-processor fabric ("nest"), utilizing a single high-frequency clock domain across the chip. With this design, adaptive clocking [3] is not practical and instruction throttling is the countermeasure of choice in terms of actuation [4] in conjunction with sensors.

Each z14™ processor core and nest region includes several CPM sensors. Each CPM outputs a 12b thermometer code representing the local value of the power supply voltage, where higher values indicate higher supply values and thus higher timing margin. The CPM schematic is shown in Fig. 18.1.1. The macro includes a programmable delay to calibrate each CPM across a range of process and supply conditions, consisting of a coarse delay (MUX stage) and a fine delay (programmable-strength inverters). The edge signal, adjusted by the programmable delay, is captured by a delay line. The delay line is sampled by two sets of flip-flops, as shown in Fig. 18.1.1. The two sets of FFs receive clocks which are skewed by a delay, calibrated to one half of the delay line's stage through programmable capacitive loads switched on by pass gates. This double sampling effectively doubles the resolution of the CPM, which can sense supply voltage changes of less than 1%.

In the previous generation z13™ CP chip, the minimum of all CPM values in each core was collected and filtered at a central location to control throttling in each core. When the filtered value dropped below a simple threshold, throttling signals were sent out to several units. The throttling reduced the rate of instruction processing, lowering the activity in the core logic and the current demand in the core, thus reducing the voltage droop. When the voltage recovered, the instruction throttling is ramped back to unthrottled operation, closing the mitigation loop. The response time was slowed by the need to combine the signals from all CPMs in a core, and then redistribute the throttling signal across the core. In the new design, three new CPM-based methods have been added.

The first new technique is to use the CPMs to measure the slope (dV/dt) of the droop to help to predict when a dangerous droop is starting. The slope of the supply voltage is measured by counting the number of clock cycles between specific CPM value changes. In the original technique, throttling could be initiated when a filtered core CPM value dropped below a certain threshold. Slope triggering adds the ability to initiate throttling if the CPM value drops between two higher values in less than N clock cycles; effectively, this allows droops to be sensed and stopped sooner to minimize required margins. As shown in Fig. 18.1.2, a different slope threshold can be set for each of three pairs of CPM values. Threshold-based throttling is still used to mitigate lower-frequency droops.

The second new technique is to add multiple faster local loops within each core as shown in Fig. 18.1.3. In the original technique, all CPM outputs were combined in a single power-management unit and a global per-core throttling signal was sent to the targeted units for instruction throttling. In this design, each CPM still sends its output to the per-core power-management unit, but is also involved in a local loop. The local loop only uses a single CPM and a local copy of the power-management logic; it uses a process similar to an original loop, but with different filter, threshold, and slope parameters. Each local loop throttles a single unit in the core, but has a much faster response time. The local throttling signals are merged appropriately with the full core throttling loop signals to avoid excessive throttling.

The third new technique is to use CPM information from neighboring cores. Measurements of the previous generation modules revealed that power supply noise generated in one core propagates to neighboring cores approximately 5ns later [4]. CPM droop information can be sent to neighboring cores faster than this noise propagation to help to predict droops earlier. When a core decides to throttle, it forces its nearest neighbors to throttle as shown in Fig. 18.1.2 via the usage of "WARNING" signals communicated only between neighboring cores. Similarly, as shown in Fig. 18.1.3, non-local information is used within each core: when a local CPM determines a droop is beginning, it notifies cores via similar "WARNING" signals, which immediately initiate throttling for a small number of cycles.

As an alternative to measuring the supply through an electrical sensor such as the CPM, we also constructed a digital estimator engine for the power delivery network (PDN) distributed across the chip. The estimator engine is shown in Fig. 18.1.4; it consists of 15 node blocks (one node for each core and 5 nodes representing similar sized regions of the nest) as well as 22 connectors and 15 edges. The core nodes estimate a local transient current by summing up weighted activity counters inside the core that represent major power-hungry events (e.g. different issue types). In a subsequent operation, the nodes calculate a local voltage resulting from the local transient current, as well as transfer currents incoming from neighboring nodes and edges. The connector passes current information between cores and simulates the retarding effects of inter-node power-grid resistance and inductance. The edge simulates a node's power grid, external decoupling capacitance, power-network and C4 inductance and resistance. At every cycle, each node outputs a value "Q" for each core, which represents the charge consumed by the core at each cycle; higher Q values result in smaller power supply voltage values. Estimator results, including voltage and dV/dt slope, can then begin to initiate instruction throttling at any core, as an alternative to CPM methods.

Figure 18.1.5 shows measurement data using the CPM as the droop sensor. Adding slope-based thresholds, we mitigate more than half of the worst-case droop, with the drawback of more frequent throttling and thus increased risk of performance impact for extreme workloads. It is important to find the optimal settings for all the loop parameters to maximize droop margin reduction, while guaranteeing negligible performance loss for any realistic workload. This optimization is performed by hardware system testing using many real and artificial workloads.

Figure 18.1.6 shows how the charge, Q_{TOTAL}, calculated from the CBCA-based digital estimator engine and the CPM values (recorded in a single core) correlate. As the core alternates between high- and low-activity periods during this experiment, the CBCA "Q" output accurately tracks the measured CPM values. Lower/higher supply voltage values, caused by high/low core activity, lead to a large/small Q estimation; during these times, the CPM senses the grid rail supply voltage and reports a smaller/larger value. The Q-to-CPM correlation is highly accurate for different core activity phases. This shows that the CBCA-based estimator engine can augment or replace the CPM sensors to control noise mitigation. These noise mitigation features collectively yield a 4% improvement in z14™ product frequency with no loss in reliability margins, doubling the benefit seen in the previous generation from centralized threshold-based throttling alone.

References:

[1] A. Drake, et al., "A Distributed Critical-Path Timing Monitor for a 65nm High-Performance Microprocessor," *ISSCC*, pp. 398-399, 2007.
[2] T. Webel, et al., "Robust Power Management in the IBM z13™," *IBM JRD*, vol. 59, Issue 5, pp. 16:1-16:13, 2015.
[3] K. Bowman, et al., "A 16nm Auto-Calibrating Dynamically Adaptive Clock Distribution for Maximizing Supply-Voltage-Droop Tolerance Across a Wide Operating Range," *ISSCC*, pp. 152-153, 2015.
[4] P. Chuang, et al., "Power Supply Noise in a 22nm z13™ Microprocessor," *ISSCC*, pp. 438-439, 2017.

 978-1-5090-4940-0/18/$31.00

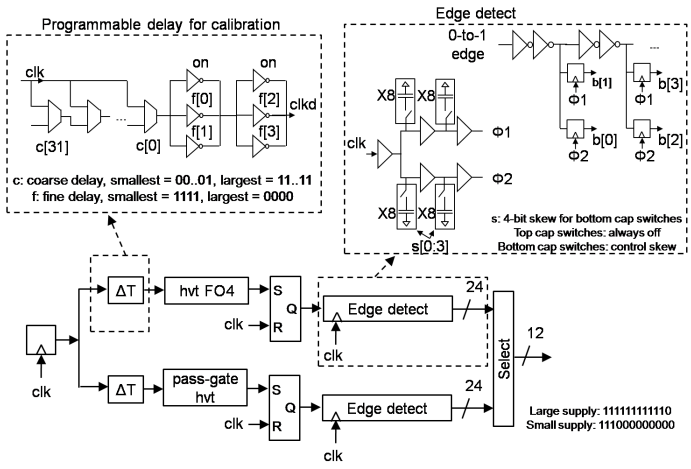


Figure 18.1.1: Design of the critical-path-monitor (CPM) macro with increased resolution using Φ1, Φ2 skewed clocks.

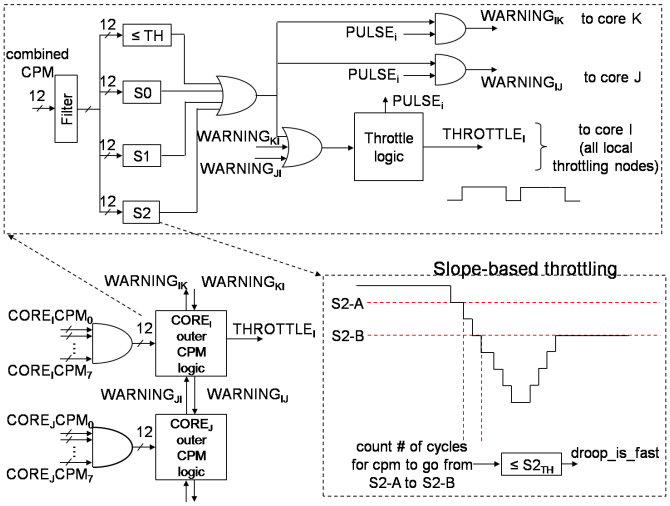


Figure 18.1.2: Centralized CPM processing in each core based on local CPM value, slope, or nearest neighbor throttling. WARNING_KI denotes warning from core K throttling core I.

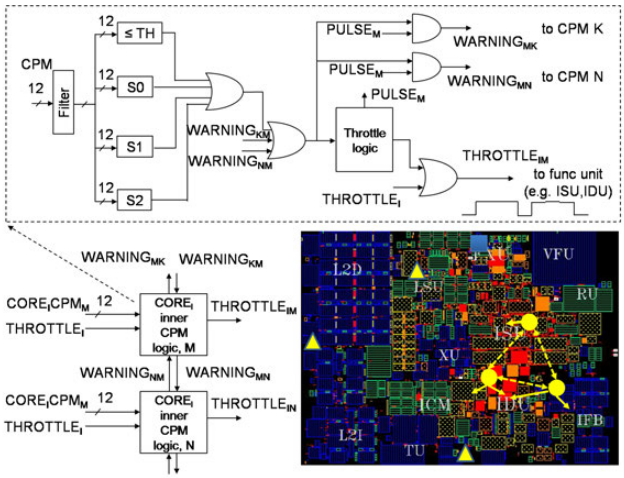


Figure 18.1.3: CPM processing logic within each core. Each inner loop triggers throttling based on one CPM (circle) value or slope, and warns neighboring inner loops within each core. Other CPMs without inner loops (triangle) are also shown.

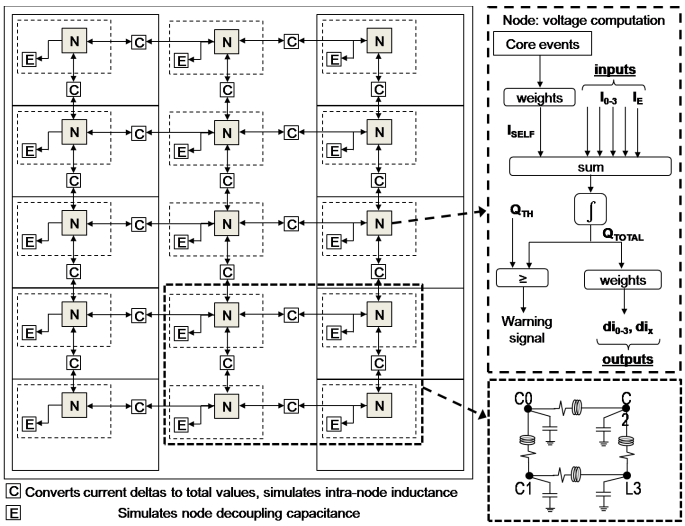


Figure 18.1.4: The on-chip CBCA estimator engine showing 10 core nodes and 5 nest nodes (middle column), where core activity determines currents.

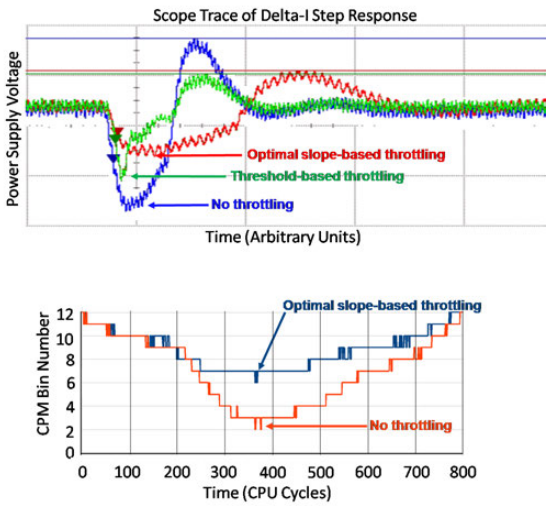


Figure 18.1.5: Measured power-supply voltage waveforms (top) and CPM tracing (bottom) through a droop, where slope-based throttling reduces droop by 4 CPM bins.

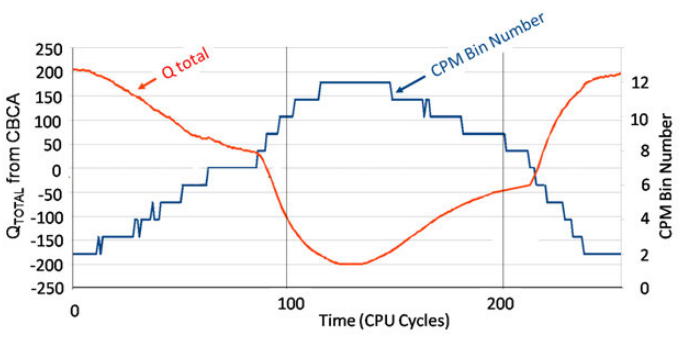


Figure 18.1.6: Measured CPM traces versus CBCA Q_{TOTAL} values for a core during a periodic droop-stress workload.

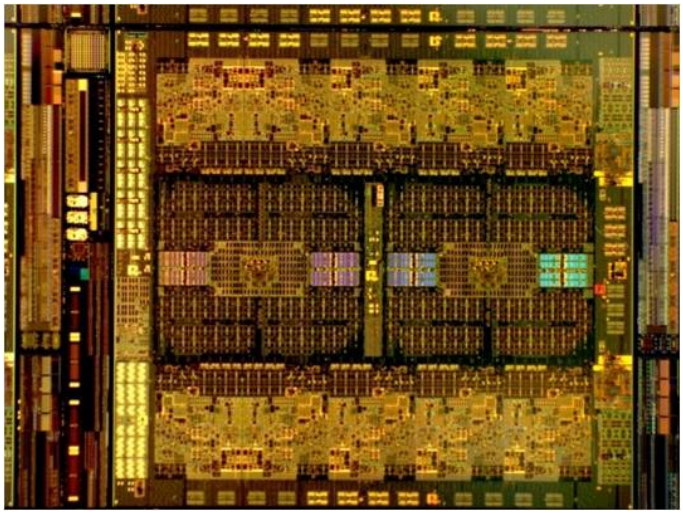


Figure 18.1.7: z14™ chip die photograph.

978-1-5090-4940-0/18/$31.00

18.2 A Combined All-Digital PLL-Buck Slack Regulation System with Autonomous CCM/DCM Transition Control and 82% Average Voltage-Margin Reduction in a 0.6-to-1.0V Cortex-M0 Processor

Xun Sun, Sung Kim, Fahim ur Rahman, Venkata Rajesh Pamula, Xi Li, Naveen John, Visvesh S. Sathe

University of Washington, Seattle, WA

Integrated Voltage Regulation (IVR) using buck converters enables efficient, fine-grained supply-voltage control in modern SoC domains [1]. However, existing IVR implementations face several challenges. As voltage domains continue to shrink, reduced per-domain decoupling capacitance requires rapid IVR transient response, leading to unfavorable efficiency and supply droop margin trade-offs. Additionally, digital domains exhibit a wide load current (I_{load}) range, requiring capabilities for autonomous transition between Continuous Conduction Mode (CCM) and Discontinuous Conduction Mode (DCM). All-digital IVR solutions are particularly desirable for ease of integration in SoCs.

Several techniques have been proposed to address these IVR challenges. Autonomous CCM-DCM transition is proposed in [2], but requires analog comparators and does not account for bridge driver delays for zero current switching, adversely affecting IVR efficiency. Adaptive clocking techniques that inject load-domain (V_{dd}) supply noise into phase-locked loop (PLL) oscillators to modulate the clock period (T_{clk}) and maintain timing slack have been proposed [3], [4]. However, benefits observed in [3] are limited by V_{dd}-delay sensitivity mismatch between critical paths and the PLL oscillator, and by the undesirable phase tracking mechanism of conventional PLLs. A fused Low-Dropout Regulator (LDO) and PLL system addresses these concerns [4], but the technique is restricted for use with a specific LDO regulator topology – no general unified clock and power architecture has been demonstrated to date. Importantly, existing adaptive clocking techniques are unable to completely restore cycles lost or gained during V_{dd} transients, a highly desirable feature for inter-domain data communication and real-time applications.

This paper presents a unified clock and power (UniCaP) architecture that exploits joint supply-voltage and phase/frequency control to aggressively reduce dissipative V_{dd} margins arising from supply-noise and temperature variation. In addition, UniCaP enables complete recovery of any cycles gained or lost during supply noise events. The key idea behind this architecture is the use of a V_{dd}-powered digitally tunable replica oscillator (TRO) to guarantee timing slack in the presence of supply noise, while incorporating voltage regulation into the PLL loop to allow the TRO to track the reference clock (REFCLK). Unifying the clock and voltage regulator subsystems allows for all-digital construction, voltage-reference-free implementation, and the ability to effectively reject timing degradation due to supply droop and temperature variation. We also present an all-digital DLL-based technique for autonomous CCM-DCM transition and Zero Voltage Switching (ZVS).

Figure 18.2.1 compares the proposed UniCaP architecture with conventional IVRs that regulate V_{dd} based on a voltage-reference (V_{ref}). Voltage regulation requires margins for Process, Voltage and Temperature (PVT) variation to avoid timing failure. In contrast, UniCaP employs a configurable TRO to match the delay and V_{dd}-sensitivity (k_{TRO}) of the critical path. Any supply droop or ripple suitably modulates T_{clk}, compensating for droop-induced logic delay degradation. The buck converter is incorporated into the PLL control loop to regulate the system operating frequency through V_{dd} control. Locking the TRO to REFCLK using a wide-range Time-to-Digital Converter (TDC) ensures by construction, that UniCaP completely recovers any cycles lost or gained during supply transients. Relying on the TRO to guarantee timing slack, and adjusting V_{dd} to regulate the system operating frequency also allows UniCaP to continuously track temperature variation, severely reducing temperature-induced V_{dd} margins.

Figure 18.2.2 shows the structure of the UniCaP architecture implemented as a joint PLL-Buck system. Use of a V_{dd}-powered TRO poses a significant tracking challenge. Large supply noise voltages and high TRO voltage-sensitivity cause phase errors that readily exceed the limited 2π phase-detection range of conventional Phase-Frequency Detectors (PFD). This excursion results in cycle-slipping and the inability to recover cycles lost (gained) during a V_{dd} droop (surge). To allow wide-range acquisition, the PLL-buck employs a coarse-grained TDC, which relies on counting TRO clocks in each REFCLK cycle to digitally measure f_{clk}, and computationally derive the phase error. The PLL-buck is designed to track phase excursions of up to 16π, sufficient for this application. A DLL-enabled tracking loop enables precise timing control for ZVS and DCM, both of which are critical for IVR applications that use lower inductor values. A clocked comparator captures the voltage polarity across M_n on the rising edge of *n*, allowing the DLL to adjust t_{dead} (the dead-time), and align the M_n turn-on event with V_x=0V for ZVS. The comparator similarly tracks the direction of the inductor current (I_L) at the falling edge of *n*. This I_L-direction indicator allows the controller to tune M_n and M_p on-times during DCM, aligning M_n turn-off with the I_L=0 event.

Joint compensation of the PLL-buck system in UniCaP is more involved than compensation for individual PLLs or buck converters due to the interaction of both sub-systems within a single loop (Fig. 18.2.3). The presence of 2 discrete buck converter poles close to unity, in addition to the two unity-poles required for zero steady-state phase error offer a very limited locus of stable operation, making implementation of traditional control techniques infeasible. The PLL-buck employs composite control for stable compensation, tracking frequency (and therefore V_{dd}) at the nominal rate (T_{REFCLK}) and phase at 10× lower rate. Composite control, using two sufficiently different time scales yields an effective system transfer function that is much easier to compensate for phase lock. CCM-DCM transition and control is governed by a finite state-machine (FSM). The DLL relies on clocked-comparator-based detection of the polarity of $I_{L,min}$, the minimum inductor current to detect the DCM condition, and adjusts on-times for M_n and M_p to ensure zero current switching. A 3-cycle wait in each direction of the CCM-DCM transition introduces the required hysteresis in CCM-DCM transition control.

A 65nm CMOS test-chip (die-photograph in Fig. 18.2.7) was implemented to demonstrate and characterize key components of the UniCaP architecture. An ARM Cortex M0 processor is used to evaluate the impact of the PLL-buck on digital systems. All reported data points pass standard M0 f_{max} and *speed-indicative* benchmarks. A programmable synthetic load was used to emulate larger, more sophisticated digital systems. Fig. 18.2.4 shows DVFS operation, performed by changing the PLL divider ratio to assert a new target frequency, allowing V_{dd} to be automatically determined by system based on PVT conditions. f_{max} experiments of the processor running benchmark traces are performed under *baseline* configuration (no additional supply noise), and with additional supply noise from a 90mA, 1ns current step under *conventional* (no elastic TRO) and PLL-buck (elastic TRO) configurations at 20°C. The PLL-buck achieves a peak 95% droop margin reduction across its operating frequency range. Temperature sweeps demonstrate the PLL-buck system's ability to track temperature variation through V_{dd} control. The temperature-induced V_{dd} margin reduction across a -10-to-100°C range was measured to be 55mV. At V_{dd}=1.0V, peak buck converter efficiency was measured to be 96.3% with 15mA of load current.

Figure 18.2.5 shows measured V_{dd} droop in response to a 90mA, 1ns I_{load} step-up and step-down, under the default PLL-lock configuration, and under frequency-only lock (with the phase-control path disabled). Both waveforms indicate stable operation in the presence of significant V_{dd} disturbance. The PLL-buck system temporarily drives V_{dd} above its target value in response to a droop, adjusting f_{clk} to recover the phase lost during the V_{dd} droop. Corresponding behavior is observed for a V_{dd} surge. Also shown are measured V_{dd} waveforms indicating autonomous CCM-DCM transitions depending on changes in I_{load} caused by transitioning between program test patterns.

Figure 18.2.6 summarizes the performance and key features of the UniCaP PLL-buck implementation in comparison to relevant IVR and supply droop management techniques. Test-chip measurements of the proposed all-digital PLL-buck system demonstrate an average supply margin reduction of 82% due to voltage droop and 55mV due to temperature variation. Any cycles gained or lost due to supply noise are fully recovered.

Acknowledgments:
The authors thank John Uehlin, Daniel Zindel for their assistance during system test, Carlos Tokunaga, Sanjay Pant and Arijit Raychowdhury for valuable discussions, and ARM for providing processor IP. This work is partly funded by SRC under task 2712.006

References:
[1] H. K. Krishnamurthy, et al., "A Digitally Controlled Fully Integrated Voltage Regulator With On-Die Solenoid Inductor with Planar Magnetic Core in 14nm tri-gate CMOS," *ISSCC*, pp. 336-337, 2017.
[2] X. Zhang, et al., "A 0.6 V Input CCM/DCM Operating Digital Buck Converter in 40 nm CMOS," *IEEE JSSC*, vol. 40, no. 11, pp. 2377-2386, Nov. 2014.
[3] N. Kurd, et al., "Next Generation Intel Core Micro-Architecture (Nehalem) Clocking," *IEEE JSSC*, vol. 44, no. 4, pp. 1121-1129, 2009.
[4] S. Gangopadhyay, et al., "UVFR: A Unified Voltage and Frequency Regulator with 500MHz/0.84V to 100KHz/0.27V Operating Range, 99.4% Current Efficiency and 27% Supply Guardband Reduction", *ESSCIRC*, pp. 321-324, 2016.
[5] C. Huang, et al., "An 82.4% Efficiency Package-Bondwire-Based Four-Phase Fully Integrated Buck Converter with Flying Capacitor for Area Reduction," *ISSCC*, pp. 362-363, 2013.
[6] M. Kar, et al., "An All-Digital Fully Integrated Inductive Buck Regulator with a 250-MHz Multi-Sampled Compensator and a Lightweight Auto-Tuner in 130-nm CMOS," *IEEE JSSC*, vol. 52, no. 7, pp. 1825-1835, 2017.

 978-1-5090-4940-0/18/$31.00

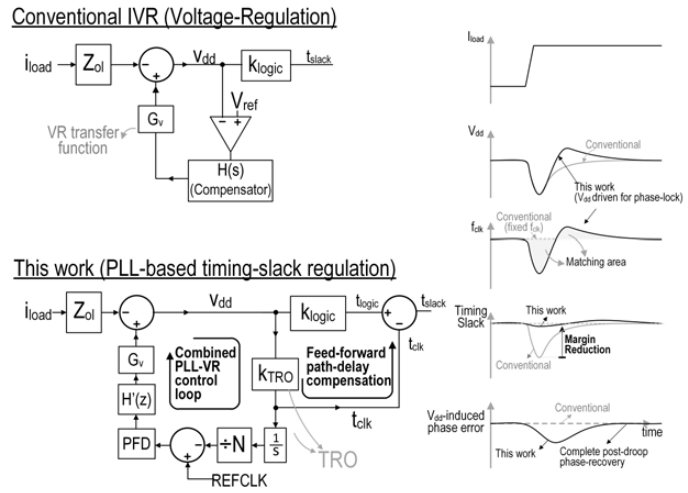


Figure 18.2.1: Comparison of conventional IVR, which regulates to a target voltage, to the proposed UniCaP architecture, regulating operating frequency through V_{dd} control.

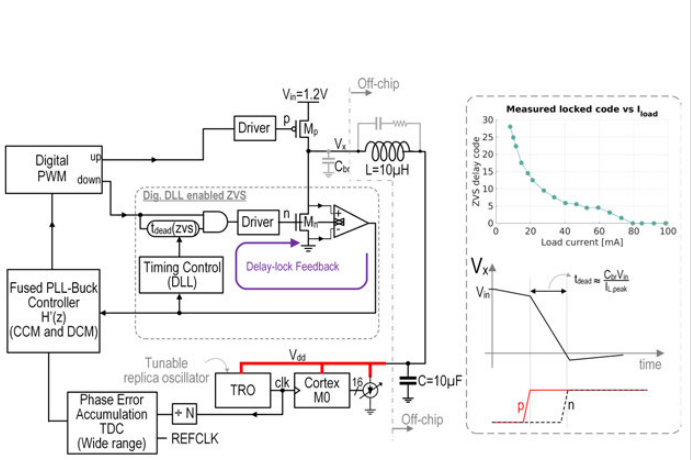


Figure 18.2.2: Architecture overview of unified PLL-buck regulator with DLL-enabled ZVS and autonomous CCM/DCM transition.

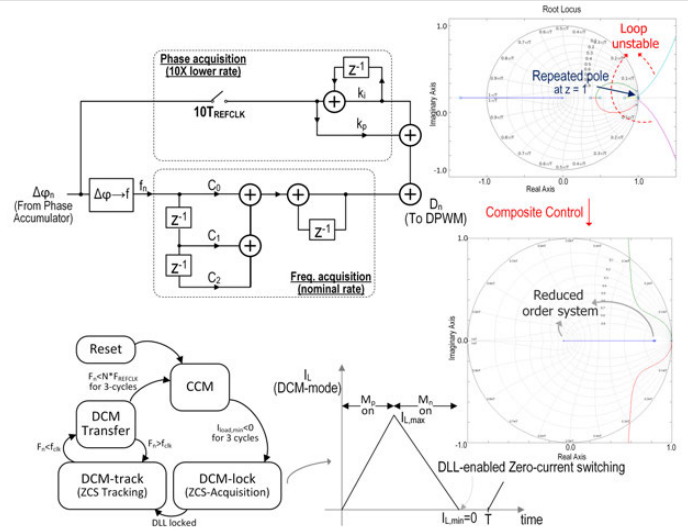


Figure 18.2.3: Composite control allows for stable PLL-buck compensation (top); CCM/DCM transition control finite state machine (bottom).

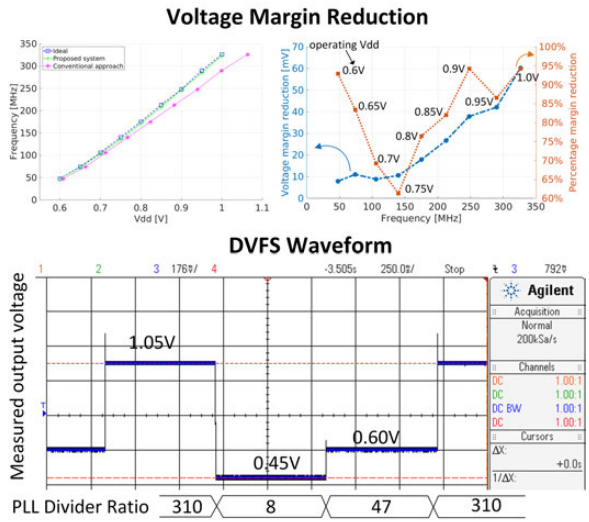


Figure 18.2.4: Measured Cortex M0 fmax vs. V_{dd}, and V_{dd}-margin reduction (top); Oscilloscope trace of DVFS transitions effected by varying N (Fig. 18.2.2).

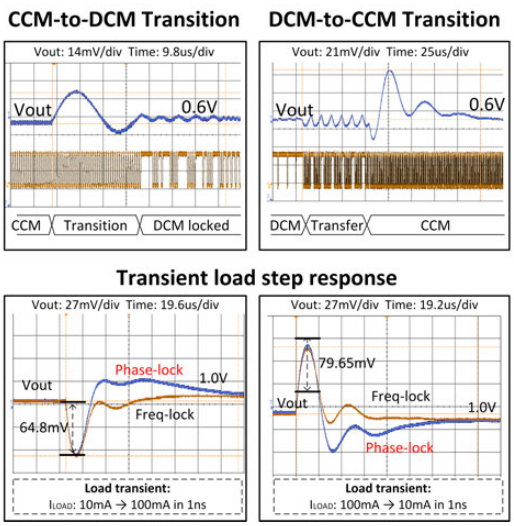


Figure 18.2.5: Measured waveforms of CCM/DCM transition (top); transient voltage step response under phase and frequency lock configurations (bottom).

	[2]	[5]	[6]	[4]	This work
Technology	CMOS 40nm	CMOS 130nm	CMOS 130nm	CMOS 130nm	CMOS 65nm
Topology	Buck	Buck	Buck	LDO	Buck
All-digital	No	No	No	Yes	Yes
Input voltage [V]	0.6 – 1.1	1.2	1.2	0.6 – 1.0	1.2
Output voltage [V]	0.3 – 0.55	0.6 – 1.05	0.45 – 1.05	0.38 – 0.81	0.6 – 1.0
Frequency / Phase Tracking	No / No	No / No	No / No	Yes / No	Yes / Yes
Margin reduction (Droop / Temp)	No	No	No	Yes	Yes
General applicability	NA	NA	NA	No (LDO only)	Yes
Inductor / Capacitor	220uH/N.R.	8nH/3.73nF	11.8nH/3.2nF	NA	10uH/10uF
Operation mode	CCM/DCM	CCM/DCM	CCM/DCM	NA	CCM/DCM
Autonomous CCM/DCM transition	Yes	No	No	NA	Yes
Switching frequency [MHz]	0.1	100	125/250	NA	1
Peak efficiency	94%	82.4%	71%	99.4%	96.3%
I_{load} [mA]	0.05 – 10	4 – 1200	5 – 65	0.1 – 6	1 – 100

Figure 18.2.6: Performance summary and comparison with related works implementing voltage regulation and supply droop management techniques.

18

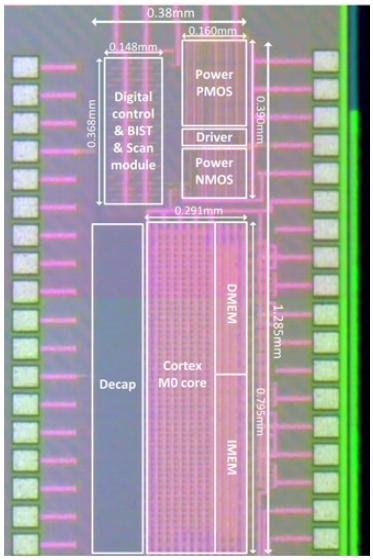


Figure 18.2.7: Die micrograph.

978-1-5090-4940-0/18/$31.00

18.3 A 2.5µW 0.0067mm² Automatic Back-Biasing Compensation Unit Achieving 50% Leakage Reduction in FDSOI 28nm over 0.35-to-1V V_{DD} Range

Anthony Quelen[1], Gael Pillonnet[1], Philippe Flatresse[2], Edith Beigné[1]

[1]CEA-LETI-MINATEC, Grenoble, France,
[2]STMicroelectronics, Crolles, France

Worst-case design and post-silicon tuning are well established digital design practices reducing timing violations in presence of process, temperature, aging and voltage variations, but they suffer from extra power consumption due to overdesign [1]. Adaptive voltage scaling (AVS) [2] and body bias modulation [1] are well-known strategies to dynamically ensure that the digital core can operate at a targeted frequency, even in the presence of delay degradation due to variations. In a multiple voltage islands context, AVS requires many integrated supply generators, such as switched capacitor converters that need to be controlled accurately. Also, for fine-grained compensation, level shifters are required, impacting circuit performance. As FDSOI technology offers the ability to adjust transistor speed through high sensitivity (85mV/V_{BB}) V_{TH} tuning by acting on buried Nwell (NW) and Pwell (PW) voltages, back-biasing generators have been investigated [3-5]. However, they require an external controller to reach the optimal Back Bias (BB) voltages (no self-adjustment) ([3-4] and [5]), imposing a non-negligible area overhead for a sub-mm² digital core having a narrow compensation range limited to 0.35-0.45V V_{DD}. We therefore propose a variation-aware BB compensation unit (BBC), which dynamically self-adjusts the N- and PMOS transistors' BB voltages to maintain the target frequency with low-latency tuning (100µs) across a wide range of supply voltage (0.35-1V) and temperature (-40-125°C). The low reported area of 0.0067mm² makes it affordable for a small digital core area (0.1-2mm²). Requiring only a reference frequency signal F_{TGT}, the self-operating BBC exhibits 2.5µW quiescent current without any external components. Compared to a worst-case design strategy, the BBC unit brings up to 50% leakage reduction @0.45V_{DD}, 120°C and reduces the energy per cycle up to 32% compared to worst-case design. By providing continuous BB voltage adjustment (continuous V_{TH} tuning), the target frequency is maintained within ±3.5% accuracy.

As shown Fig. 18.3.1, the BBC unit tracks the digital frequency F_{DIG} of an on-chip critical path replica (CPR). F_{SENSE}, divided from F_{DIG}, is compared to the target frequency F_{TGT} by using a phase-frequency detector (PFD), sending up or down NW voltage commands to an NW driver. A second compensation circuit ensures symmetrical biasing between NW and PW with the use of a middle sensor (MS) and a fully-integrated negative switched-capacitor converter able to provide -1.5V in 0.6mm² digital core area (C_{FLY}=8pF). The chip is composed of the BBC unit compensating a 0.6mm² LVT digital core consuming around 50mW. The pull-up/down drivers' resistors are set to allow positive and negative 60mV/µs slew rates and limit the BBC loop gain. The BB driver is able to handle driving NW and PW equivalent capacitors (1nF each) and a 1.2nF PW/NW coupling capacitor. The BBC unit sets a digital flag BB_{OK} when the target frequency is reached by the digital core. The internal sequencing is done by a state machine using an F_{SENSE} clock.

Figure 18.3.2 illustrates the three main compensation configurations, where V_{PW} and V_{NW} are adjusted sequentially according to CPR and MS, respectively. First, an N counter from a fractional frequency of F_{DIG} ensures comparison with the divided target frequency F_{TGT2}. If N+DN periods are counted before $T_{TGT2}/2$, V_{NW} is decreased by activating pull-down resistors in NW drivers until the next falling edge of F_{TGT2}. Conversely, NW is decreased until $N\times T_{SENSE}$-$T_{TGT2}/2$, if the PFD counts less than N periods from F_{SENSE} before $T_{TGT2}/2$. During the last quarter of the F_{TGT2} period, MS is enabled to compare V_{NW} and V_{PW}, and then V_{PW} is adjusted to be equal to $-V_{NW}$. When a steady state is reached, the BBC unit senses F_{DIG} at the F_{TGT} rate and eventually compensates the drift voltage due to the body leakage (0.28µA/mm²@125°C, 1.4V_{NW}) by activating an UP or DOWN signal during T_{SENSE}. The middle sensor (MS) block, as detailed in Fig. 18.3.3, ensures an optimal symmetrical well polarization within 20mV V_{MID}. Hysteresis is realized by mismatching the R and R' resistance values. MS is only activated by enabling EN_{MID} during the last quarter of the F_{TGT2} period to decrease the average BBC unit quiescent current.

Figure 18.3.4 depicts the frequency compensation for temperature w/ or w/o BBC. When both well voltages are set to 0V@125°C, F_{DIG} decreases from 9MHz@125°C to 6MHz@-40°C. When the BBC unit is enabled, V_{NW} increases from 0V@125°C to 0.5V@-40°C to maintain the 9MHz target frequency (M=8). The well-voltage resolution is theoretically unlimited as BBC unit drivers are analog-controlled. Fig. 18.3.4 also shows the start-up phase when the BBC unit is turned on. The steady state is reached in less than 200µs and the BB_{OK} flag is set to be used at system-level. Despite that the BBC unit is not intended to mitigate power supply droops, Fig. 18.3.4 shows the dynamics of BB voltages for an abrupt 100mV V_{DD} step. The NW ramps up from 0.92 to 1.23V in less than 90µs to compensate for the frequency decrease.

Figure 18.3.5 explores the BBC unit's energy-saving ability by showing the normalized leakage current with and without BBC over -40-125°C. By self-adapting V_{NW} from 1.2V@-40°C to 0.9V@125°C (and PW accordingly), while meeting the frequency requirement, the leakage current is reduced by 50%@125°C compared to a scenario without BBC, at near-threshold operation (0.45V_{DD}). The energy dissipation per cycle is also reduced by 32%@0.45V_{DD} with a 5% activity factor compared to worst-case design. F_{SENSE} can be regulated from the F_{TGT} set value over 5.4× frequency range at near-threshold operation (0.375V) and 1.4×@1V. Indeed, the back-biasing effect is proportional to the V_{TH} to V_{DD} ratio. The ratio between F_{TGT} and F_{SENSE} is maintained within two F_{SENSE} cycles which represents a precision of 3.5%.

Figure 18.3.6 compares our variation-aware back-biasing system to the most relevant published works. The main benefits are the self-control loop capability, refreshing back biasing at the F_{TGT} rate, over a full V_{DD} range (0.35-1V). In addition, the BBC unit has negligible power (2.5µW@0.375V_{DD}), small 0.7%/mm² area overhead and dynamically compensates variations (process, temperature, aging) because of its 100µs time constant. Though only forward body bias (FBB) is shown in this paper, as LVT devices were used, the system is compatible with reverse BB (RBB) polarization with RVT devices. The demonstrated BBC unit is capable of compensation at near-threshold operation for a small-area digital core (as would be used for IoT applications), or at full supply voltage for core-to-core variations in large digital SoC with a sub-mm²-scale grain to compensate within-die device parameter variations. The BBC unit also reduces the need for challenging power supply scaling and/or low efficiency worst-case design strategies.

References:

[1] J. W. Tschanz, et al., "Adaptive Body Bias for Reducing Impacts of Die-to-Die and Within-Die Parameter Variations on Microprocessor Frequency And Leakage," *IEEE JSSC*, vol. 37, no. 11, pp. 1396–1402, 2002.

[2] M. Cho, et al., "Post-Silicon Voltage-Guard-Band Reduction in a 22nm Graphics Execution Core Using Adaptive Voltage Scaling and Dynamic Power Gating," *ISSCC*, pp. 152–153, 2016.

[3] N. Kamae, et al., "A Body Bias Generator with Wide Supply-Range Down to Threshold Voltage for Within-Die Variability Compensation," *IEEE ASSCC*, pp 53-56, 2014.

[4] M. Blagojević, et al., "A Fast, Flexible, Positive and Negative Adaptive Body-Bias Generator in 28nm FDSOI," *IEEE Symp. VLSI Circuits*, 2016.

[5] S. Clerc, et al., "A 0.33V/-40°C Process/Temperature Closed-Loop Compensation SoC Embedding All-Digital Clock Multiplier and DC-DC Converter Exploiting FDSOI 28nm Back-Gate Biasing," *ISSCC*, pp. 150-151, 2015.

 978-1-5090-4940-0/18/$31.00

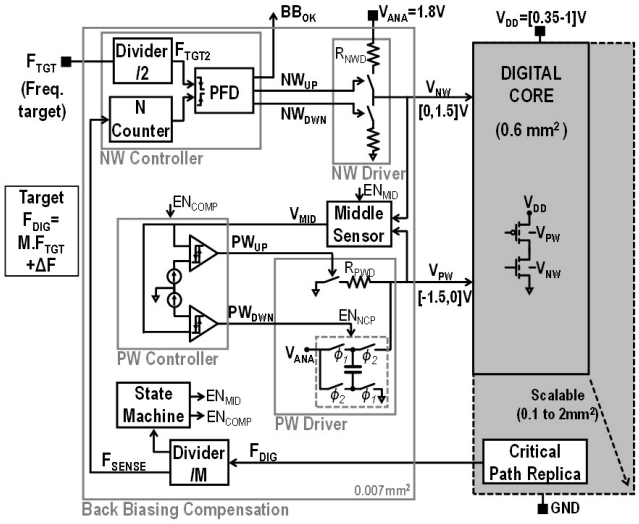


Figure 18.3.1: Main building blocks of back-biasing compensation unit embedded with a sub-mm² digital core.

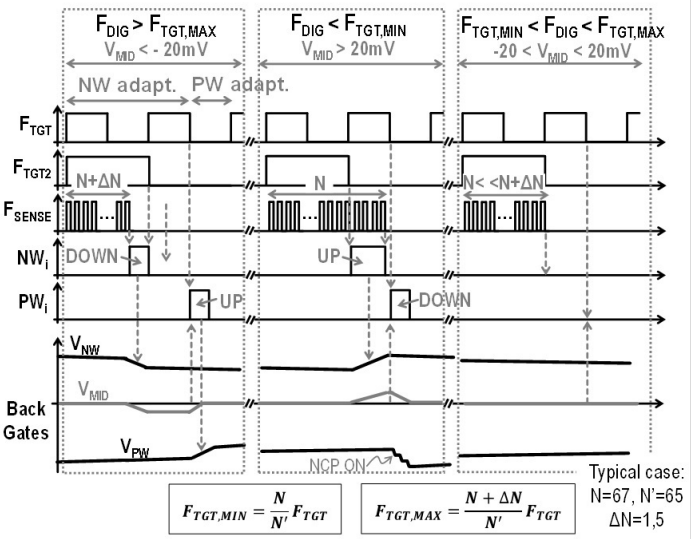


Figure 18.3.2: Chronogram in three main biasing configurations.

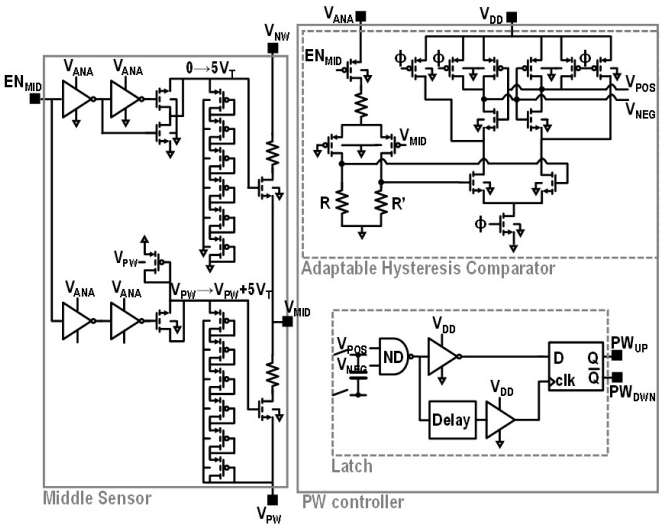


Figure 18.3.3: Schematic of middle sensor (MD) including duty-cycled V_{MID} resistive divider and latched hysteresis comparator.

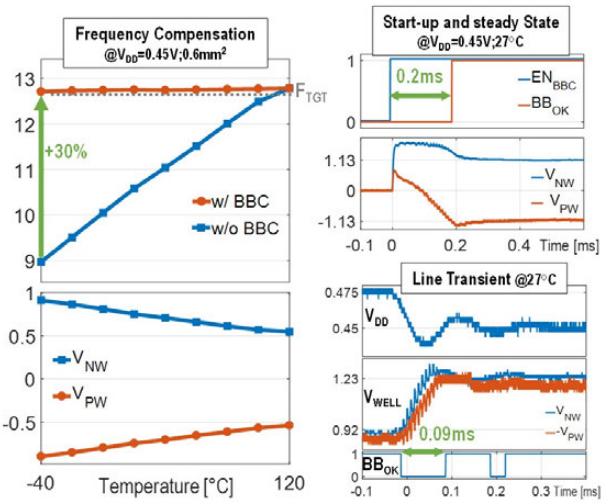


Figure 18.3.4: Back-biasing profile vs. temperature to maintain targeted frequency; start-up phase timing and biasing dynamics during abrupt power supply transient.

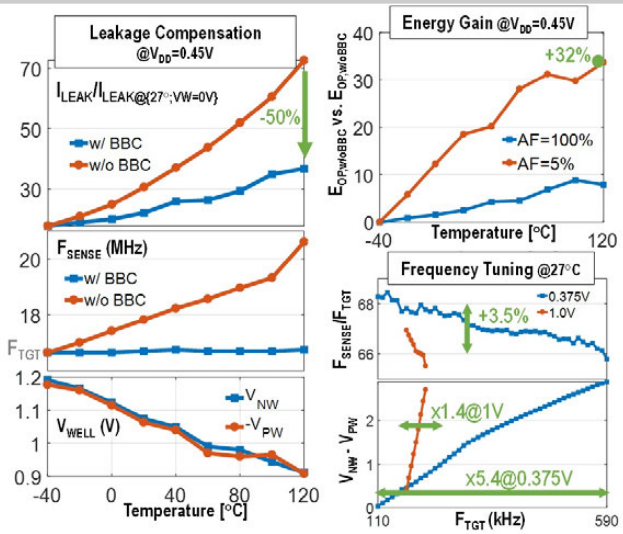


Figure 18.3.5: Leakage and energy reduction due to the back-biasing compensation unit, frequency compensation and accuracy vs. voltage supply.

Conditions	ASSCC'14 [3]	VLSI'16 [4]	ISSCC'15 [5]	This Work	Unit
Technology	65nm Bulk	28nm FDSOI	28nm FDSOI	28nm FDSOI	-
Onchip compensation	No	No	Yes	Yes	-
Input target	Voltage	Voltage	Freq.	Freq.	-
Frequency accuracy	N.A	N.A	-	3.5	%
Core supply range	0.5-1.2	0.76-0.97	0.33-0.45	**0.35-1**	V
Temperature range	-	-	-40 / 40	**-40 /125**	° C
Quiescent power	600	10	-	**2.5**	µW
Loop time constant	N.A	N.A	-	**0.1**	ms
V_{WELL} min. step	19	58	100	**continuous**	mV
Nwell Pwell	-0.6 / 0.6 -0.6 / 0.6	0 / 1.8 -1.4 / 0	0 / 1.4 -1.4 / 0	**0 / 1.8** **-1.5 / 0**	V
Drivers area	0.0052	0.012	-	**0.0037**	mm²
System area	N.A	N.A	0.62*	**0.0067**	mm²
Overhead area	2.3	1.2	-	**0.35**@2mm²	%

** external capacitance needed (1µF)*

Figure 18.3.6: Comparison to state-of-art back biasing generator with or without feedback loop.

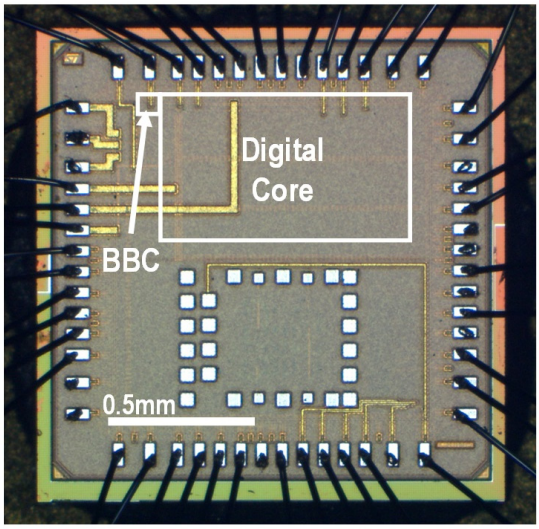


Figure 18.3.7: Die micrograph including the BBC unit (0.0067mm²) and digital core (0.6mm²) on a 1.2×1.2mm² die.

978-1-5090-4940-0/18/$31.00

18.4 A 0.4V 430nA Quiescent Current NMOS Digital LDO with NAND-Based Analog-Assisted Loop in 28nm CMOS

Xiaofei Ma[1,2], Yan Lu[1], Rui P. Martins[1,3], Qiang Li[2]

[1]University of Macau, Macau, China
[2]University of Electronic Science and Technology of China, Chengdu, China
[3]Instituto Superior Tecnico/University of Lisboa, Lisbon, Portugal

Ultra-low-power fully-integrated voltage regulators with fast load-transient performance are highly attractive for low-power systems-on-a-chip (SoCs). In such systems, the digital units working in the subthreshold region are more sensitive to supply variations. The digital low-dropout regulator (DLDO) is more suitable for low-supply-voltage operation, as compared to an analog LDO regulator. But, traditional DLDOs are either slow or power hungry, and need a large output capacitor (consumes area) to survive a fast load transient. When a higher clock frequency is used for faster response, both the current efficiency and the loop stability are degraded [1]. An analog-assisted (AA) loop was used in [2] to provide a high-pass loop in parallel with the slow digital loop for fast response. However, a large coupling capacitor (100pF) was still needed, trading off area with power and speed. An NMOS power stage as a source follower is sometimes used in replica LDOs and cascaded LDOs [3] for its intrinsic response to load transient; the NMOS source follower naturally provides more output current when V_{OUT} drops. To improve upon the power-speed-area tradeoffs, this paper presents a DLDO using NMOS power switches, and employs a NAND-gate-based high-pass analog path (NAP) to assist the slow low-power digital loop. With these two techniques, nearly two orders of better FoM is achieved relative to the state-of-the-art.

Figure 18.4.1 shows the DLDO with the NMOS intrinsic response and NAP, which combined, achieve a fast response. As the NMOS switch array needs to be driven by a high voltage, a small 2× charge pump (CP) is employed. Because the CP only supplies the tiny quiescent current of the gate drivers and the level shifters (LSs), a 12pF total capacitor (including $C_{F1,2}$=3pF and C_B=6pF) can meet the current demand sufficiently. The dynamic part of the switch-driving current will be filtered by C_B.

For the PMOS switch array and the AA loop in [2], the g_m of the switch array is proportional to the number of turned-on PMOS switches. When there are only a few switches turned on in light load, the g_m is too small to compensate a large load transient, and consequently, the effectiveness of the AA loop decays. For our NMOS switch array, we add the NAP to the two highest bits (being off in light load) of the switch array, breaking the tie between the effective g_m and the number of switches turned on. As shown in Fig. 18.4.2, the output voltage is AC coupled to V_{CP}, which is DC-biased to $2\times V_{DD}$ by R_1. Then, V_{CP} is connected to a modified NAND gate of which the corresponding PMOS M_1 has large size to amplify the coupled signal to the power NMOS's gate when the output has an undershoot. Since the C_C is connected to the NAND buffer, instead of the power NMOS, only a 12pF C_C is used for the NAP.

Upon a load transient event, the NMOS power switch responds first to increase/decrease the output current. Then, the loop-1 starts the action of increasing the output current if the output has an undershoot. At the next clock rising edge, the loop-2 starts to adjust the output voltage to the preset value. Fig. 18.4.2 shows the simulated transient responses of the traditional PMOS DLDO, NMOS DLDO without NAP, and NMOS DLDO with NAP, when the load current jumps from 0.5mA to 20.5mA with 3ns edge time and zero C_L. The traditional PMOS DLDO would have a 426mV undershoot, while the NMOS DLDO's undershoot is 244mV benefiting from the NMOS intrinsic response. The NMOS DLDO with NAP loop (this work) has only 96mV of undershoot – a superior transient-response capability. Fig. 18.4.2 also exhibits the I-V characteristics of the turned-on NMOS and PMOS switches, when both the PMOS and NMOS have an I_{OUT}=11.5mA at V_{OUT}=450mV; the NMOS power stage can manage 19% higher current than the PMOS one with the same V_{OUT} variation of 20mV. Combined with the dead-zone control structure, an NMOS DLDO has a higher probability of working in the dead-zone region than a PMOS DLDO, reducing its dynamic power.

Figure 18.4.3 shows the overall architecture of the NMOS DLDO with NAP. The digital control loop uses a weighted shift register (SR) and a coarse-fine tuning structure to get higher DC accuracy and smaller recovery time. The weights of the SR in the fine and coarse loops are optimized for small glitches, while reducing the number of registers. The fine loop contains an 8b SR, controlling eight 1×-size power switches. The coarse loop uses a 4b low SR and a 16b high SR with carry in/out between the two SRs. The low SR controls four 8×-size power switches and the high SR controls sixteen 32×-size power switches in the coarse loop. All the shift-register outputs go to the LSs to control the power stage in the $2\times V_{DD}$ domain.

Figure 18.4.4 shows the simulated voltage waveforms during a load transient. Due to the control-loop logic-delay mismatches, large glitches, which may enlarge the voltage undershoot/overshoot, will happen during the carry operation between the different weighted switch arrays [2]. Here, we designed the low and high SRs to have different logic delays so voltage glitches are opposite in direction to the undershoot/overshoot. In the proposed architecture, the mode-selecting multiplexer of the coarse loop, which is the clock-gating stage for the low SR, is implemented with ultra-high-threshold-voltage (UHVT) devices; and, the NOR gate connected to the high SR for clock gating uses regular-threshold-voltage (RVT) devices. As such, edges of CLK_H arrive earlier than CLK_L, as shown in the zoomed-in region. So, the weighted shift-register structure will experience a carry for the high SR first, and then reset the low SR, turning the glitch polarity to the desired direction.

The double-tail comparator [4] with an extra 'valid' output is used for the three comparators. The 'valid' signal will be high after the comparator finishes the comparison and will be low during its reset period. Thus, the fine and coarse loops are triggered by the 'valid' signal, instead of being triggered by the global clock, to avoid toggling the SR incorrectly. The rationale for this is that it is challenging to maintain the right clock timing for all the SRs and comparators, due to the PVT variations, especially under a 400mV sub-threshold input voltage. In addition, the comparator in the fine loop stops working during coarse tuning and freeze mode to reduce the dynamic current consumption.

The test chip was fabricated in a 28nm bulk CMOS process. The proposed NMOS DLDO has a 50mV dropout voltage and can deliver 20mA with 0.35-to-0.5V output, 0.4-to-0.55V input, and 4MHz clock. Fig. 18.4.5 shows the measured load transient response. With V_{OUT}=0.45V and V_{IN}=0.5V and a zero-output capacitor, when the load current changes from 0.5mA to 20.5mA with 3ns edge time, the measured undershoot and overshoot voltages are 117mV and 49mV, respectively. And, with V_{OUT}=0.35V, V_{IN}=0.4V, zero-output capacitor, and 1MHz clock, when the load current changes from 1mA to 16mA with 3ns edge time, the measured undershoot and overshoot voltages are 111mV and 46mV, respectively. For dynamic voltage scaling, the measured reference up- and down-tracking speeds are 238mV/μs and 91mV/μs, respectively. The comparison table in Fig. 18.4.6 shows that this design has two orders of better FoM as compared to the state-of-the-art, with the low quiescent current and small required capacitance. Fig. 18.4.7 shows the chip micrograph.

Acknowledgments:
This work is supported in part by the Macao Science and Technology Development Fund (FDCT) 093/2016/A and SKL-AMSV-2017-2019, and the Research Committee of University of Macau, and in part by National Natural Science Foundation of China (NSFC) under grant 61534002.

References:
[1] S. B. Nasir, et al., "A 0.13μm Fully Digital Low-Dropout Regulator with Adaptive Control and Reduced Dynamic Stability for Ultra-Wide Dynamic Range," *ISSCC*, pp. 98–99, 2015.
[2] M. Huang, et al., "An Output-Capacitor-Free Analog-Assisted Digital Low-Dropout Regulator with Tri-Loop Control," *ISSCC*, pp. 342–343, 2017.
[3] Y. Lu, et al., "An NMOS-LDO Regulated Switched-Capacitor DC–DC Converter With Fast-Response Adaptive-Phase Digital Control," *IEEE Trans. Power Electron.*, vol. 31, no. 2, pp. 1294–1303, 2016.
[4] D. Schinkel, et al., "A Double-Tail Latch-Type Voltage Sense Amplifier with 18ps Setup+Hold Time," *ISSCC*, pp. 314–605, 2017.
[5] L. G. Salem, et al., "A 100nA-to-2mA Successive-Approximation Digital LDO with PD Compensation and Sub-LSB Duty Control Achieving a 15.1ns Response Time at 0.5V," *ISSCC*, pp. 340–341, 2017.
[6] Y. J. Lee, et al., "A 200-mA Digital Low Drop-Out Regulator With Coarse-Fine Dual Loop in Mobile Application Processor," *IEEE JSSC*, vol. 52, no. 1, pp. 64–76, Jan. 2017.
[7] F. Yang and P. K. T. Mok, "A Nanosecond-Transient Fine-Grained Digital LDO With Multi-Step Switching Scheme and Asynchronous Adaptive Pipeline Control," *IEEE JSSC*, vol. 52, no. 9, pp. 2463-2474, 2017.

 978-1-5090-4940-0/18/$31.00

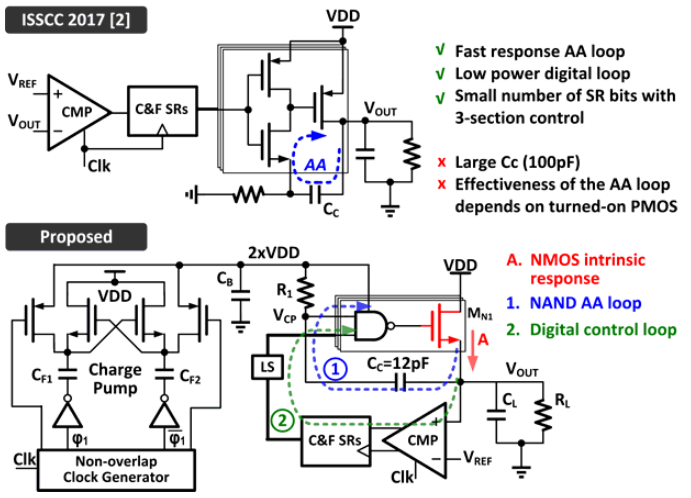


Figure 18.4.1: The AA-DLDO scheme (top); the proposed NMOS DLDO with NAND-gate-based high-pass analog path (bottom).

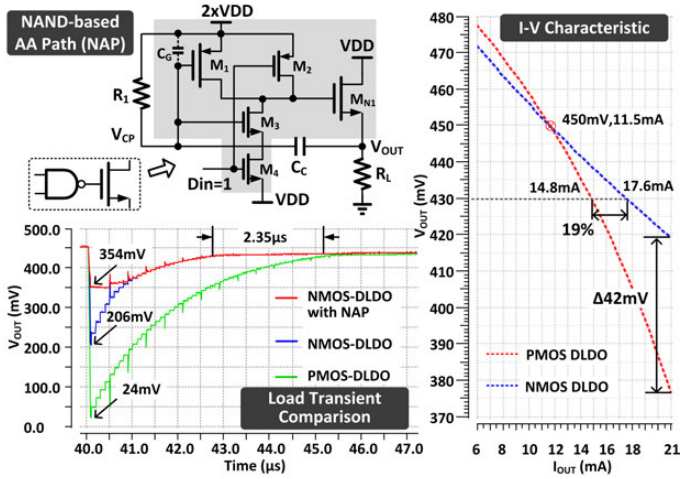


Figure 18.4.2: Circuit implementation with NAND-gate-based high-pass analog path; the transient waveforms of the traditional PMOS DLDO, NMOS DLDO without NAP, and NMOS DLDO with NAP; I-V characteristics of the turned-on NMOS and PMOS power switches.

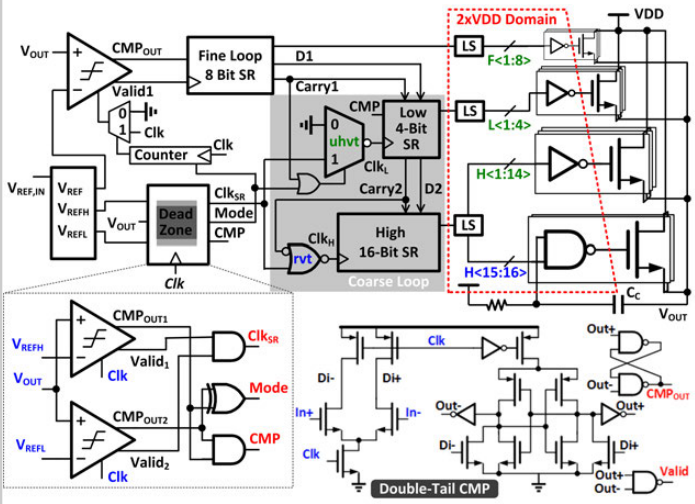


Figure 18.4.3: Overall architecture of the proposed NMOS NAP-DLDO; schematics of the dead-zone comparator and double-tail comparator.

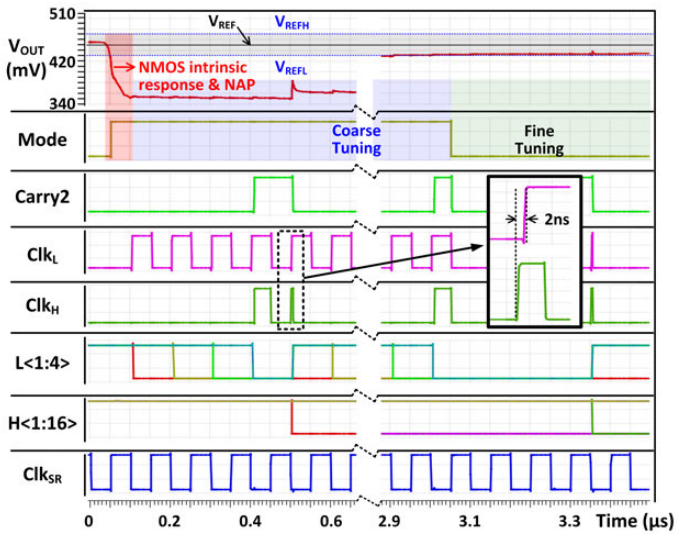


Figure 18.4.4: Timing diagram of the NMOS NAP-DLDO.

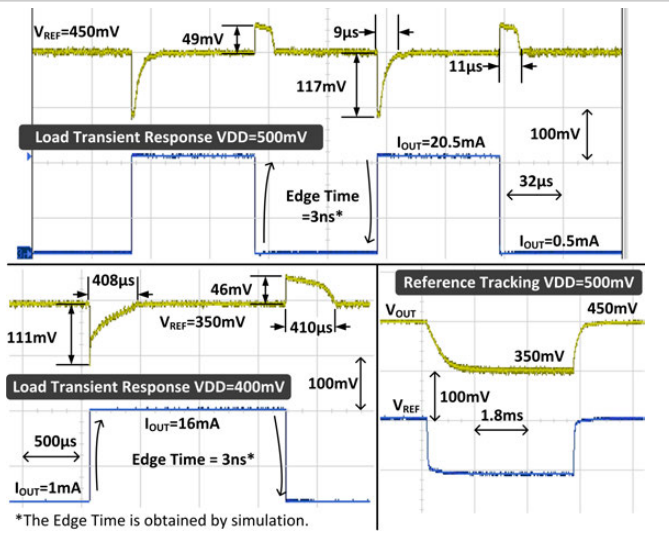


Figure 18.4.5: Measured load transient response; and reference tracking waveforms.

	This Work		[2] ISSCC'17	[1] ISSCC'15	[5] ISSCC'17	[6] JSSC'17	[7] JSSC'17
Process	28nm		65nm	130nm	65nm	28nm	65nm
Area [mm^2]	0.0055		0.03	0.114	0.0023	0.021	0.158
Type	Digital		Digital	Digital	Digital	Digital	Digital
Achitecture	SR/NMOS/NAP		SR/AA	SR/RDS	SR/PD/PWM	SR/ADC	Async.
V_{IN} [V]	0.4-0.55	0.4	0.5-1	0.5-1.2	0.5-1	1.1	0.6-1
V_{OUT} [V]	0.35-0.5	0.35	0.45-0.95	0.45-1.14	0.3-0.45	0.9	0.55-0.95
F_{CLK} [MHz]	4	1	10	400	240	N.A.	N.A.
I_{Q_MIN} [µA]	0.81	0.43	3.2	24-221	14	110	300
C_{TOTAL} [pF]	24		100	1000	400	23500	1500
ΔV_{OUT} [mV]	117	111	105	90	40	120	50
ΔI_{LOAD} /T_{EDGE}	20mA /3ns	15mA /3ns	10mA /1ns	1.4mA /N.A.	1.06mA /1ns	180mA /4µs	100m /2ns
FOM*[ps]	0.0057	0.0051	0.23	8600	199	7.75	2.3

* $FOM = \frac{C_{TOTAL} \times \Delta V_{OUT} \times I_Q}{\Delta I_{LOAD}^2}$

Figure 18.4.6: Comparison with the state-of-the-art.

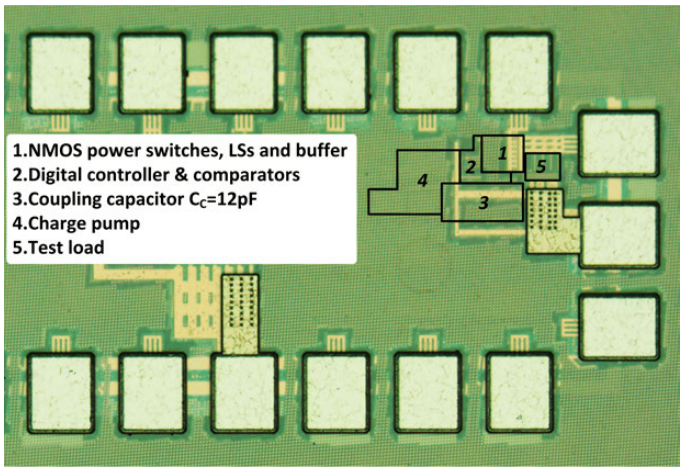


Figure 18.4.7: Chip micrograph of the NMOS NAP-DLDO.

978-1-5090-4940-0/18/$31.00

18.5 A Fully Integrated 40pF Output Capacitor Beat-Frequency-Quantizer-Based Digital LDO with Built-In Adaptive Sampling and Active Voltage Positioning

Somnath Kundu[1], Muqing Liu[1], Richard Wong[2], Shi-Jie Wen[2], Chris H. Kim[1]

[1]University of Minnesota, Minneapolis, MN
[2]Cisco Systems, San Jose, CA

Integrated voltage regulators with a wide output current/voltage dynamic range are required to support fast dynamic voltage and frequency scaling (DVFS). Low Dropout Regulators (LDOs) based on digital-intensive circuits have been gaining popularity [1]–[4] due to their compactness, process scalability, high immunity to process-voltage-temperature (PVT) variations and easy programmability for design optimization. Conventional digital LDOs utilizing a comparator and shift-registers [1] suffer from a slow response time during a large/fast change in load current (I_{LOAD}). Higher sampling frequency (f_S) improves the response time, but at the cost of increased power consumption and reduced loop stability. Multi-bit quantizers utilizing ADCs [2-4] can reduce the settling time, however, the presence of a high resolution ADC and the control logic increases the design complexity. Moreover, the ADC resolution limits the maximum f_S. In order to overcome the trade-off between speed and power, adaptive sampling techniques were incorporated in [1], [4]. But the overhead of multiple VCOs operating simultaneously and a separate overshoot/droop detection circuitry [1], or an event-driven controller with 7b ADC [4], increase the complexity and power consumption. Furthermore, none of the previous designs incorporated active voltage positioning (AVP), a popular ripple-suppression technique, whereby the LDO output is set slightly above (in low-activity state) or below (in high-activity state) the reference voltage depending on the processor workload conditions [5].

In this work, a beat-frequency quantizer-based digital LDO (BF-LDO) is described, where a pair of ring VCOs and simple digital blocks are used to generate an adaptive-sampling clock. It has several critical benefits: 1) Time quantization utilizing the VCO and counter provides a highly digital and tunable ADC design solution. The VCO phase quantization provides 1st-order noise shaping achieving high resolution. 2) Dynamically adaptive f_S proportional to the output voltage error reduces droop/overshoot and settling time by increasing f_S during transient ripples. Low f_S near steady state improves the quantizer resolution, LDO power efficiency and stability margin. 3) Inherent AVP reduces the transient ripple further by dynamically controlling the steady-state voltage. 4) The design is very robust to PVT variations, as the quantizer output is a function of the ratio of two VCO frequencies, cancelling frequency variations due to common-mode effects.

Figure 18.5.1 shows the basic operation of a time-based digital LDO. The VCO pair converts the reference and the output voltages (V_{REF}, V_{LDO_OUT}) to clock pulses (CK_{REF}, CK_{OUT}) of proportional frequency (f_{REF}, f_{OUT}). The time quantizer calculates the frequency difference, $f_{REF} - f_{OUT}$ to generate a digital code N_{OUT}. By comparing N_{OUT} with a predefined N, the digital controller adjusts the number of PMOS switches to keep V_{LDO_OUT} constant for a given I_{LOAD} range. A time quantizer is conventionally implemented by counting CK_{OUT} edges in a fixed sampling period (CK_s), which is generated by dividing CK_{REF} by a factor N. A higher N improves the quantizer resolution and the loop stability margin, but at the cost of slow response time and large transient ripple. To overcome this limitation, an adaptive sampling technique is proposed that utilizes a D-flip-flop (DFF) as a digital frequency subtractor, also known as a BF quantizer [6]. Using the beat frequency for loop sampling makes f_S proportional to $|V_{REF}–V_{LDO\_OUT}|$. During I_{LOAD} transients, V_{LDO_OUT} experiences a large ripple, increasing f_S for faster recovery. On the other hand, near steady state, f_S reduces as $|V_{REF}–V_{LDO\_OUT}|$ is very small. This improves the quantizer resolution to set V_{LDO_OUT} very precisely, and at the same time reduces the switching power dissipation with excellent loop stability. The counter counts the CK_{REF} edges in a sampling period generating $N_{OUT}=f_{REF}/|f_{REF}-f_{OUT}|$. Once steady state is reached, $N_{OUT}=N$ makes $f_S=f_{REF}/N$, causing a fixed offset at V_{LDO_OUT}. This inherent offset enables AVP, which is addressed later. In addition, since the VCOs are operating continuously without phase reset, the quantization noise is 1st-order high-pass filtered, increasing the resolution [6].

The voltage offset, i.e. $|V_{REF}-V_{LDO\_OUT}|$, can also be written as $f_{REF}/K_{VCO}\cdot N_{OUT}$, where K_{VCO} is the VCO voltage-to-frequency conversion gain. It reaches a minimum value of $f_{REF}/K_{VCO}\cdot N$ during steady state. In theory, N can be set to infinity in order to cancel this steady-state offset. However, the maximum value of N is defined by the size of the BF counter, the digital control logic, as well as the number of PMOS switches. In this implementation, for 10b PMOS control, N is nominally set to 64, introducing an offset of 16mV for the VCO pair operating at 250MHz frequency with K_{VCO} of 250MHz/V.

Figure 18.5.2 illustrates the BF-LDO architecture. It comprises the proposed BF quantizer and the conventional linear quantizer as a baseline for performance comparison. SEL_Q selects the desired quantizer. Since the BF quantizer detects the absolute voltage difference between V_{REF} and V_{LDO_OUT}, a simple 1b comparator is introduced for polarity detection to keep the loop in a negative feedback configuration at all times. The BF quantizer operation is explained in the timing diagram. The quantizer output is subtracted from the external code N and the difference goes to a 10b proportional-integral (PI) control to tune 1024 PMOS switches. The PI control parameters, K_F, K_P and K_I are fully programmable. I_{LOAD} is generated from an NMOS array driven by a test VCO that triggers a series of DFFs. The wide programmability in the VCO frequency, as well as in the number of DFFs enable an I_{LOAD} range of 0-400mA and a rise/fall time range of 16ns–9µs with a resolution of 1mA and 1.3ns, respectively.

AVP as illustrated in Fig. 18.5.3 (left), compromises the DC voltage regulation to reduce the transient ripple significantly [5]. In the proposed BF-LDO, a positive step in I_{LOAD} causes a droop in V_{LDO_OUT} and it settles at an offset, $f_{REF}/K_{VCO}N$ below V_{REF}. Similarly, a negative I_{LOAD} step brings V_{LDO_OUT} above V_{REF} with the same offset achieving built-in AVP. Although the offset is negligible for a large N providing good DC regulation, it can be easily increased for ripple reduction by reducing N. Measured waveforms in Fig. 18.5.3 show the ripple reduction from 260mV to 120mV by reducing N from 64 (=weak AVP) to 16 (=strong AVP). The PMOS switch array and the VCO implementations are shown in Fig. 18.5.3 (right). A switch driver with dummy buffers and PMOS branches keep balanced loading for the 10b binary PMOS switches, as shown in the example for 3b. This removes glitches due to rise/fall-time imbalance during code transition. Uniformly distributed layout also reduces any mismatch among PMOS branches. The VCOs have both coarse and fine frequency tuning to achieve wide frequency range and to compensate any frequency offset between them.

Figure 18.5.4 shows the measured transient response from a 65nm test-chip. I_{LOAD} steps between 20mA and 70mA in 0.8µs cause a 108mV droop and a 148mV overshoot with a settling time of 1.24µs and 1.13µs, respectively. The steady state f_S is 3.9MHz, but it increases to a much higher value during I_{LOAD} transition. The baseline LDO with a fixed 3.9MHz f_S experiences a 564mV (i.e. 5×) droop (ΔV) and a 30.8µs (i.e. 25×) settling time (Ts). The BF-LDO also achieves a fast response in V_{REF} step (Fig. 18.5.4, bottom right), which is critical for fast DVFS systems. Fig. 18.5.5 (top) compares ΔV and Ts of the proposed BF-LDO with the baseline for different rise-time, ΔT and load step, ΔI_{LOAD} showing a significant benefit of adaptive sampling. The line and load regulation are shown in Fig. 18.5.5 bottom verifying the BF-LDO functionality over wide operating conditions. The current efficiency plot in Fig. 18.5.6 shows >93% efficiency over 10× variation in I_{LOAD} with a peak value of 99.5%. Utilizing the frequency ratio of the two VCOs makes the quantizer insensitive to supply variation as evident from the measured low frequency PSRR of -38dB. The comparison table in Fig. 18.5.6 shows comparable ΔV and Ts with much lower C_{LOAD} achieving the best FOM. The active area is 0.0374mm^2 including C_{LOAD}, as shown in Fig. 18.5.7.

References:

[1] B. Nasir, et al., "0.13µm Fully Digital Low-Dropout Regulator with Adaptive Control and Reduced Dynamic Stability for Ultra-Wide Dynamic Range," *ISSCC*, pp. 98-99, 2015.
[2] Y. J. Lee, et al., "A 200mA Digital Low-Drop-Out Regulator with Coarse-Fine Dual Loop in Mobile Application Processors," *ISSCC*, pp. 148-149, 2016.
[3] L. G. Salem, et al., "A 100nA-to-2mA Successive-Approximation Digital LDO with PD Compensation and Sub-LSB Duty Control Achieving a 15.1ns Response Time at 0.5V," *ISSCC*, pp. 340-341, 2017.
[4] D. Kim, et al., "A 0.5V-VIN 1.44mA-Class Event-Driven Digital LDO with a Fully Integrated 100pF Output Capacitor," *ISSCC*, pp. 148-149, 2017.
[5] A. Paul, et al., "System-Level Power Analysis of a Multicore Multipower Domain Processor with ON-Chip Voltage Regulators," *IEEE TVLSI,* vol. 24, no. 12, pp. 1-4, 2016.
[6] S. Kundu, et al., "Two-step Beat Frequency Quantizer Based ADC with Adaptive Reference Control for Low Swing Bio-potential Signals," *CICC*, pp. 1-4, 2015.

 978-1-5090-4940-0/18/$31.00

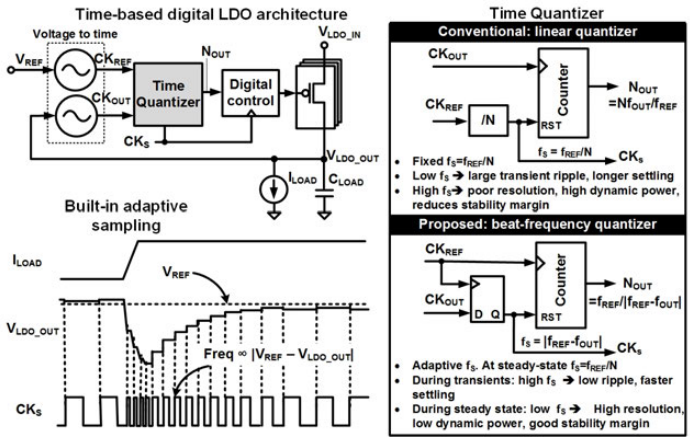


Figure 18.5.1: (Upper left) Basic operation of a time-based digital LDO. (Lower left) Adaptive sampling utilizing beat-frequency. (Right) Linear and beat-frequency time quantizer circuits.

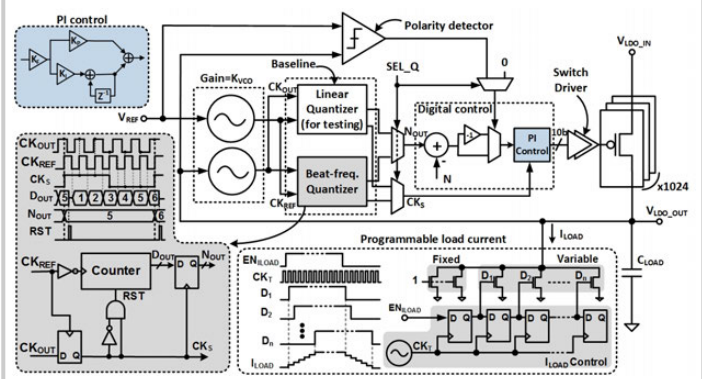


Figure 18.5.2: Implementation of beat-frequency quantizer-based digital LDO. The conventional linear quantizer was also implemented for performance comparison.

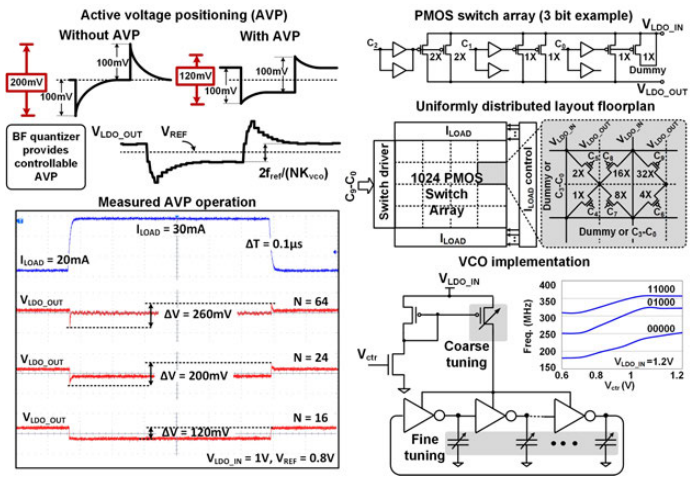


Figure 18.5.3: (Left) BF quantizer-based digital LDO provides inherent AVP. (Upper right) Implementation of the PMOS switch array. (Lower right) Implementation of the VCO.

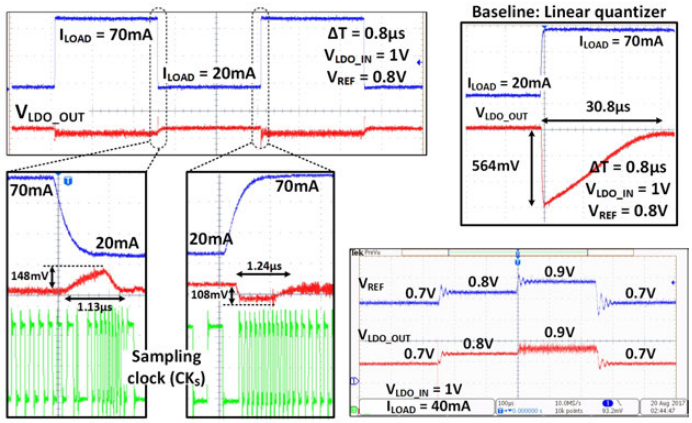


Figure 18.5.4: Measured transient response of the BF-LDO. The built-in adaptive sampling in BF-LDO provides 25× faster settling and 5× lower droop compared to the baseline.

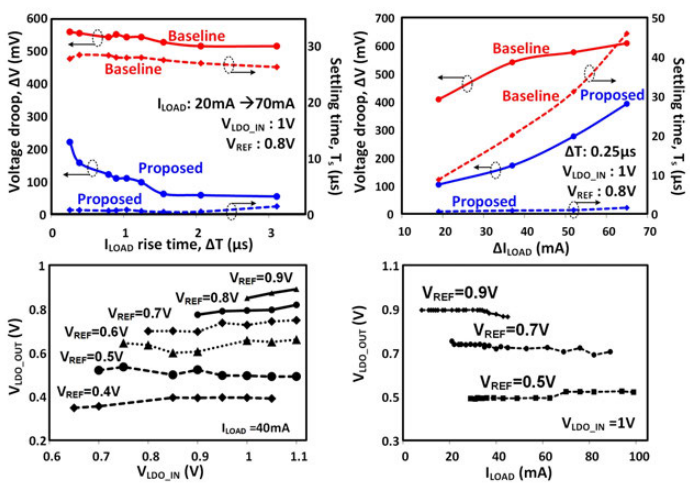
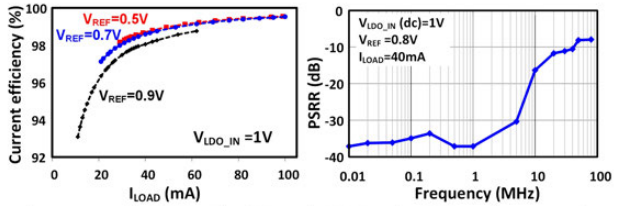


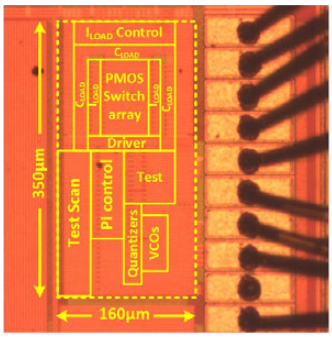
Figure 18.5.5: (Top) Measured voltage droop and settling time for different I_{LOAD} rise time and ΔI_{LOAD}. (Bottom) Measured line and load regulation for different V_{REF}.



	[1] Nasir, ISSCC'15	[2] Lee, ISSCC'16	[3] Salem, ISSCC'17	[4] Kim, ISSCC'17	This Work
Process	130nm	28nm	65nm	65nm	65nm
Architecture	Shift Reg	Shift Reg+ADC	SAR ADC	Event-driven ADC	VCO based ADC
Adaptive sampling/ AVP	Yes / No	No / No	No / No	Yes / No	Yes / Yes
V_{LDO_IN} (V)	0.5 – 1.2	1.1	0.5 – 1	0.45 – 1	0.6 – 1.2
V_{LDO_OUT} (V)	0.45 – 1.14	0.9	0.3 – 0.45	0.4 – 0.95	0.4 – 1.1
Max I_{LOAD} (mA)	4.6	200	2	3.36	100
I_Q (mA)	0.024 – 0.22	0.11	0.014	0.08 – 0.26	0.1 – 1.07
C_{LOAD} (nF)	1	23.5	0.4 (Integrated)	0.1 (Integrated)	0.04 (Integrated)
Max current efficiency (%)	98.3	99.94	99.8	99.2	99.5
ΔV, Ts @ ΔI_{LOAD}	90mV,1.1μs @ 1.4mA	120mV,40μs# @180mA	40mV,0.1μs @1.06mA	34mV,11.2μs @1.44mA	108mV,1.24μs @50mA
Area (mm²)	0.355	0.021	0.0023*	0.03*	0.0374
FOM (ps)**	10100	9.57	199	20	1.38

From oscilloscope waveform *Without C_{LOAD} **FOM=$C_{LOAD}\Delta V * I_Q/(\Delta I_{LOAD})^2$. Smaller FOMs are better

Figure 18.5.6: (Top) Measured current efficiency and PSRR. (Bottom) performance comparison with state-of-the-art digital LDOs.



	Proposed	Baseline
Process	65nm CMOS	
AVP	Yes	No
Steady-state f_S	3.9MHz @ V_{LDO_IN}=1V, V_{REF}=0.9V	
I_{LOAD}	8-100mA	
I_Q	0.8mA @ V_{LDO_IN}=1V,V_{REF}=0.9V (VCOs: 0.6mA (sim.), Switching: 0.2mA)	
C_{LOAD}	40pF	
ΔV, Ts @ ΔI_{LOAD}=50mA	108mV, 1.24µs	564mV, 30.8µs
Area (mm^2)	Core: 0.0374 (VCOs, quantizer, dig:0.015, Switch:0.008, C_{LOAD}:0.01) Test circuits: 0.0186 Total: 0.056	
FOM	1.38ps	7.21ps

Figure 18.5.7: Chip microphotograph and results summary.

978-1-5090-4940-0/18/$31.00

18.6 A 500mA Analog-Assisted Digital-LDO-Based On-Chip Distributed Power Delivery Grid with Cooperative Regulation and IR-Drop Reduction in 65nm CMOS

Yasu Lu[1], Fan Yang[2], Feng Chen[1], Philip K. T. Mok[1]

[1]Hong Kong University of Science and Technology, Hong Kong, China
[2]Qualcomm, Singapore

With the die area of modern processors growing larger and larger, the IR drop across the power supply rail due to its parasitic resistance becomes considerable. There is an urgent demand for local power regulation to reduce the IR drop and to enhance transient response. A distributed power-delivery grid (DPDG) is an attractive solution for large area power-supply applications. The dual-loop distributed micro-regulator in [1] achieves a tight regulation and fast response, but suffers from large ripple and high power consumption due to the comparator-based regulator. Digital low-dropout regulators (DLDOs) [2] can be used as local micro-regulators to implement a DPDG, due to their low-voltage operation and process scalability. Adaptive control [3], asynchronous 3D pipeline control [4], analog-assisted tri-loop control [5], and event-driven PI control [6] are proposed to enhance the transient response speed. However, digital LDOs suffer from the intrinsic limitations of large output ripple and narrow current range. This paper presents an on-chip DPDG with cooperative regulation based on an analog-assisted digital LDO (AADLDO), which inherits the merits of low output ripple and sub-LSB current supply ability from the analog control, and the advantage of low supply voltage operation and adaptive fast response from the digital control.

Figure 18.6.1 shows the topology of the AADLDO-based DPDG. The outputs of each AADLDO are connected as a 3×3 mesh and all regulators respond to the local voltage in the local vicinity of the grid. Thus, the DPDG provides point-of-load tight local regulation in the steady state. $AADLDO_{1,3,7,9}$, $AADLDO_{2,4,6,8}$ and $AADLDO_5$ have two, three and four neighbors, respectively. The current load is typically unevenly distributed on chip (as shown in Fig. 18.6.1) and once a load suddenly increases, the voltage at that node will drop and the current will flow from the neighbors to that node to help handle the transient. With the cooperative regulation, dramatically improved response speed during a transient is possible. The clock of each AADLDO uses one of the nine interleaved phases of a 9-stage current-starved ring oscillator, the load change detection speed is thus faster with the DPDG. With the help of a DPDG, the maximum average current drive capacity of each local AADLDO can be designed to be smaller than the peak load current, if the load only drains such a large current occasionally.

Figure 18.6.2 shows the topology of $AADLDO_5$ in the proposed 3×3 grid with a simplified finite state machine of the controller and an operation timing diagram. $AADLDO_5$ has four neighbors and the other 8 AADLDOs have the same topology, but fewer neighbors. 128 power MOS and their buffers are divided into 16 groups, with 8 power MOS and buffers in each group sharing the same V_{BFH} and V_{BFL}, which can be switched between V_{DD}/GND and analog control voltage V_{ANA}. By connecting V_{BFH} and V_{BFL} to V_{ANA}, the gate of the selected power MOS is then controlled by an analog voltage instead of the digital V_{DD}/GND. The regulator can then provide sub-LSB current and thus, the output ripple is reduced. The controller is specified in Verilog and synthesized via a digital design flow. The controller can operate in digital burst mode (DBM), digital/analog transition mode (DATM) and analog steady mode (ASM). Two comparators with an added inner offset compare V_{OUT} with $V_{REF}+\Delta$ and $V_{REF}-\Delta$. The comparison result Q_5[1:0] indicates whether the voltage is above, inside or below the boundary of 2Δ. When V_{OUT} is not inside the boundary (InB=0), the controller works in DBM to achieve a fast response and DAS[15:0] selects the V_{BFH} and V_{BFL} of all the groups to connect to V_{DD} and GND, respectively. The 128b adaptive-gain shift register (AGSR) can shift left/right to fully turn on/off the power MOS in an adaptive step size. The barrel gain of the shift register can be adjusted from -5 to +5 according to the local comparator results Q_5[1:0] and the states of four neighboring AADLDOs observed from $Q_{2,4,6,8}$[1:0]. If the V_{OUT} enters the boundary (InB=1), the controller will work in DATM, the AGSR stops shifting and IVG[23:0] controls the initial voltage generator to generate the initial voltage to prepare for the coming smooth transition from digital to analog. DAS[15:0] will select three (or two) groups of power MOS to be controlled by the analog loop after the analog voltage V_{ANA} is switched to initial voltage $V_{INITIAL}$ by the AIS. When all four neighbors enter the boundary (NbrsInB=1), the controller will go into ASM and the AIS will choose the amplifier to take over the control to reduce output ripple and improve regulation. The clock frequency is 16MHz if all $V_{OUT1\text{-}9}$ are inside the boundary and is increased to 100MHz immediately if any one goes outside the boundary.

The operation of the digital/analog selector is shown in Fig. 18.6.3 (top). LOC[15:0] is generated by the XOR of the headers of each group in the AGSR, indicating the location of the boundary of '1' and '0' in the shift register. Once the '1'/'0' boundary is identified in group[i], the group and neighboring group[i-1] and group[i+1] will be switched to the analog control loop by DAS[15:0], and the gate voltage of all the power MOS in these three groups will be controlled by analog voltage V_{ANA}. Only two groups will be controlled by the analog loop if the '1'/'0' boundary is in group[0] and group[15]. The initial voltage is generated according to the relative location of the '1'/'0' boundary in the 8b data of the previously-mentioned group.

A 1×3 and 3×3 DPDG were fabricated in a 1.2V low-leakage 65nm CMOS process. The 1×3 DPDG can be reconfigured as a single LDO, 1×2 and 1×3 grid for testing. A single AADLDO is designed to deliver 56mA current (comprising a total of 500mA with the entire DPDG). Fig. 18.6.4 shows the load transient measurement results. The load transient speed is 3× faster with AGSR compared to a conventional shift register and the analog-assisted scheme provides better regulation compared to the dead-zone scheme, both measured with a single AADLDO. Three waveforms are measured with a single AADLDO, 1×2 grid and 1×3 grid, which has zero, one and two neighbors, respectively. With the help of two neighboring regulators, the droop voltage has a 66% reduction. When $Load_5$ in the 3×3 grid changes between 3mA to 53mA, the four neighbors of $AADLDO_5$ will help to regulate, while $AADLDO_2$ only has three neighbors to help when $Load_2$ has the same change. With one more neighbor to help, the droop voltage is reduced from 90mV to 55mV. When all the loads change between 3mA to 53mA with a 20ns edge time, all nine voltages in the grid have a 125mV droop and 250ns recovery time.

Figure 18.6.5 shows that neighbors can provide more help if their load current is large, but this ability will saturate when their load reaches a critical point of 30mA – close to their maximum current of 56mA. A tunable inter-gridpoint resistance (R_{IGP}) is intentionally added between the neighboring power delivery grid points to emulate the size of the grid. A larger grid with a larger R_{IGP} offers less help during the transient period. The DC characteristic of the output voltage vs. the load current is measured with 0.6-to-1.2V V_{IN} range with a 50mV dropout voltage. With a 50mV dropout voltage, the current efficiency increases when the load increases and the current efficiency at low input voltage is slightly higher, due to less quiescent current in low voltage.

Figure 18.6.6 summarizes and compares the performance of the design presented with state-of-the-art designs. With AADLDO-based DPDG, this work achieves a fast 0.28ps FOM with a 0.9nF on-chip capacitor. Fig. 18.6.7 shows the micrograph of the AADLDO-based 1×3 and 3×3 DPDG.

Acknowledgements:
This work was supported by the Research Grant Council of Hong Kong SAR Government, China, under project No. 16206615. The authors would like to thank Mr. S. F. Luk for his technical support.

References:
[1] J. F. Bulzacchelli, et al., "Dual-Loop System of Distributed Microregulators With High DC Accuracy, Load Response Time Below 500 ps, and 85-mV Dropout Voltage," *JSSC*, vol. 47, no. 4, pp. 863-874, 2012.
[2] Y. Okuma, et al., "0.5-V Input Digital LDO with 98.7% Current Efficiency and 2.7-µA Quiescent Current in 65nm CMOS," *CICC*, pp. 1-4, 2010.
[3] S. B. Nasir, et al., "A 0.13µm Fully Digital Low-dropout Regulator with Adaptive Control and Reduced Dynamic Stability for Ultra-Wide Dynamic Range," *ISSCC*, pp. 98-99, 2015.
[4] F. Yang, et al., "A Nanosecond-Transient Fine-Grained Digital LDO With Multi-Step Switching Scheme and Asynchronous Adaptive Pipeline Control," *JSSC*, vol. 52, no. 9, pp.2463-2474, 2017.
[5] M. Huang, et al., "An Output-Capacitor-Free Analog-Assisted Digital Low-Dropout Regulator with Tri-Loop Control," *ISSCC*, pp. 342-343, 2017.
[6] D. Kim, et al., "Fully Integrated Low-Drop-Out Regulator Based on Event-Driven PI Control," *ISSCC*, pp.148-149, 2016.

 978-1-5090-4940-0/18/$31.00

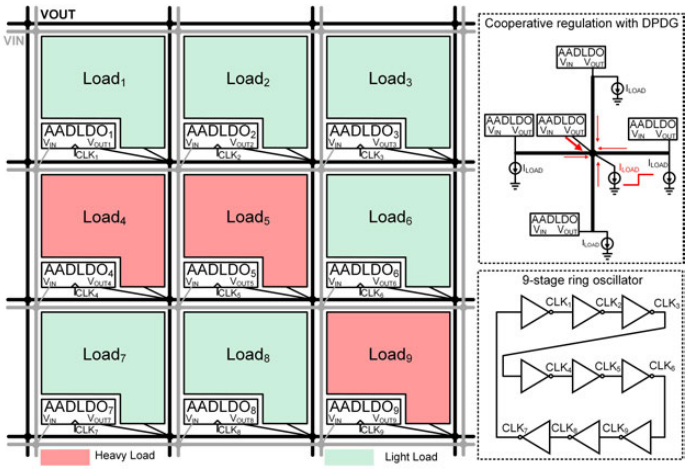


Figure 18.6.1: AADLDO-based 3×3 distributed power delivery grid.

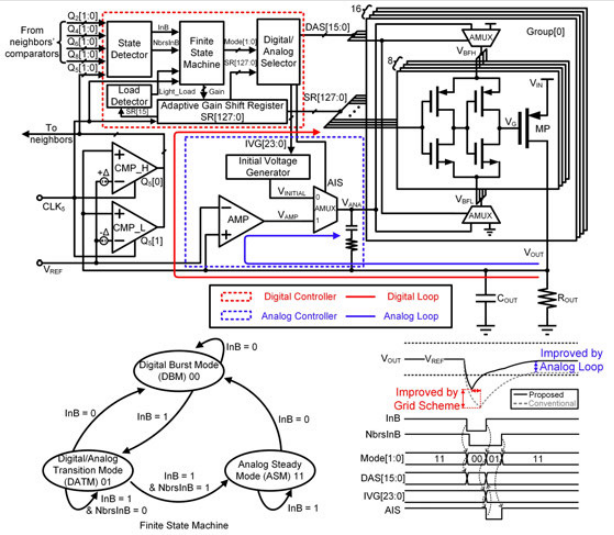


Figure 18.6.2: Topology of AADLDO$_5$, simplified FSM of the digital controller and operation timing diagram.

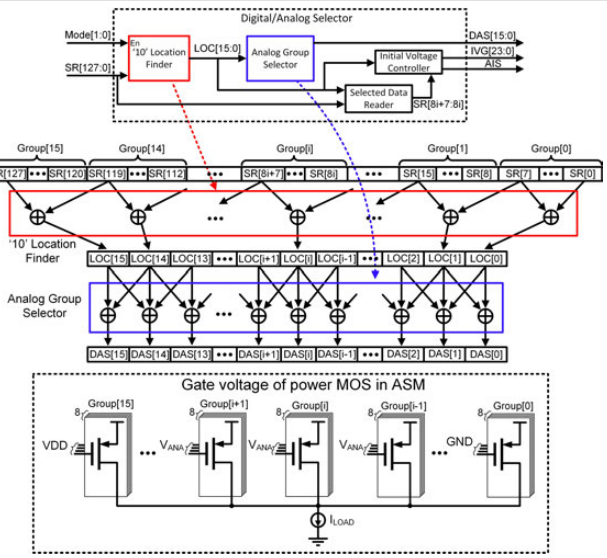


Figure 18.6.3: Digital/analog selector and gate voltage of power MOS in ASM.

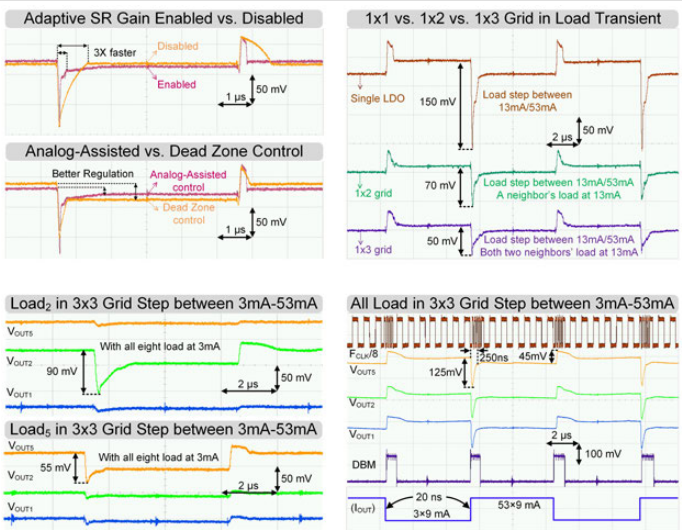


Figure 18.6.4: Measured load transient response of the DPDG under different conditions.

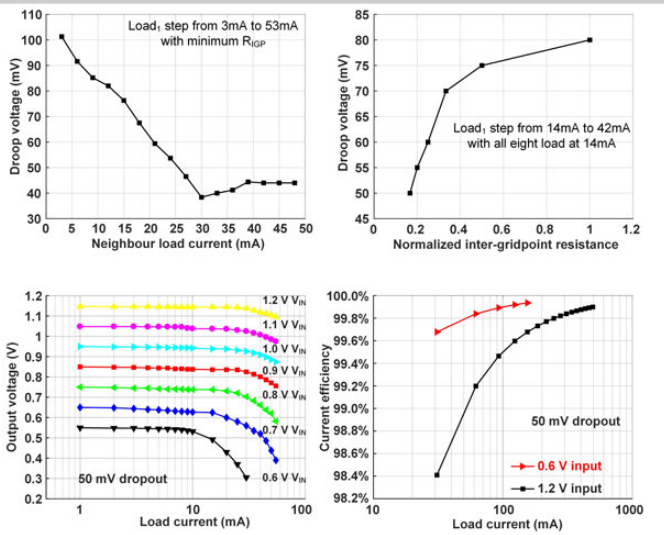


Figure 18.6.5: Measured droop voltage vs. neighbors´ load current and normalized R_{IGP}, DC output voltage and current efficiency versus I_{LOAD}.

	[1] JSSC2012	[3] ISSCC2015	[4] JSSC2017	[5] ISSCC2017	[6] ISSCC2016	This Work
Process	45 nm SOI	130 nm	65 nm	65 nm	65nm	65 nm
Type	Distributed	Centralized	Centralized	Centralized	Centralized	Distributed
Architecture	ALDO	DLDO	DLDO	AADLDO	DLDO	AADLDO
V_{IN} [V]	1.179-1.625	0.5-1.2	0.6-1	0.5-1	0.5-1	0.6-1.2
V_{OUT} [V]	0.9-1.1	0.45-1.14	0.55-0.95	0.45-0.95	0.45-0.95	0.55-1.15
$F_{SW,MAX}$ [MHz]	-	400	-	10	200	100
C_{OUT} [nF]	1.46	1	1.5	0.1	0.4	0.9
$Area_{active}$[mm²]	0.075	0.114	0.158	0.01	0.029	0.7758
$I_{LOAD,MAX}$ [mA]	42	4.6	500	12	3.511	500
I_Q [μA]	12000	24-221	300	3.2	12.5-216	500
Ripples [mV]	12.5	< 20 *	5	< 10 *	~ 3 *	4
ΔV_{OUT} [mV]@ $\Delta I_{OUT}/T_{EDGE}$	7.6@ 4.5mA/0.309ns	90@ 1.4mA/N.A.	50@ 100mA/2ns	105@ 10mA/1ns	40@ 0.4mA/N.A.	125@ 450mA/20ns
FOM [ps]**	62.4	76.5	2.3	0.34	1.11	0.28

* Observed from the figures

** $FOM = \frac{C\Delta V_{OUT}}{\Delta I_{MAX}} \times \frac{I_Q}{\Delta I_{MAX}}$

Figure 18.6.6: Comparison with the state-of-the-art.

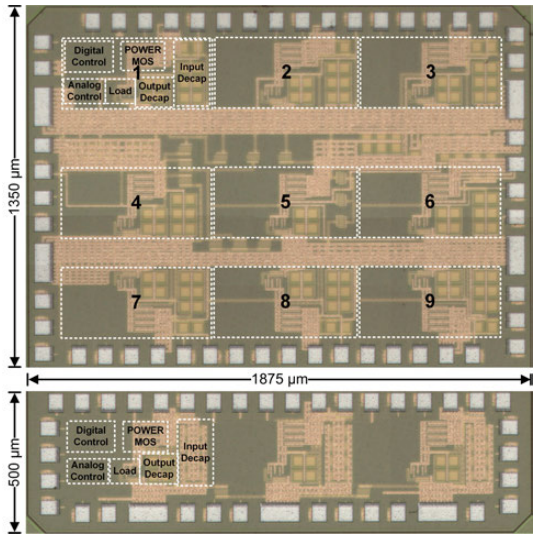


Figure 18.6.7: Chip micrograph of the proposed 3×3 and 1×3 AADLDO-based DPDG.

978-1-5090-4940-0/18/$31.00

18.7 A Sub-1.55mV-Accuracy 36.9ps-FOM Digital-Low-Dropout Regulator Employing Switched-Capacitor Resistance

Loai G. Salem, Patrick P. Mercier

University of California, San Diego, La Jolla, CA

Modern DVFS-enabled SoCs require nimble supply regulators that rapidly respond to abrupt load changes and offer fine resolution (e.g., 12.5mV in [1], 10mV in [2]) over large voltage and current dynamic ranges. Switch-array digital LDOs (SA-DLDOs) are a potentially attractive regulation option due to their ability to operate with low input voltages and in part to their modular digital nature and scalability. SA-DLDOs employ 2^N unary- [3] or binary-weighted [4] PMOS arrays that are modulated through a 1b or multi-bit ADCs to maintain the output voltage (V_{out}) at the desired level (V_{ref}), as shown in Fig. 18.7.1 (top left). Unfortunately, while array conductance in SA-DLDOs *linearly* increases with *equal step size* ($_G$LSB) as the code is increased, the output voltage step, $_v$LSB, does not; in fact, $_v$LSB is nonlinear: $\sim G_L V_{out} \times {}_G$LSB. Thus, SA-DLDOs achieve a nonlinear steady-state error, $e_{ss} = V_{ref} - V_{out} \approx \pm {}_G\text{LSB}/G_L \times V_{drop}$, as shown in Fig. 18.7.1 (bottom left), that deteriorates at large dropout voltages, $V_{drop} = V_{in} - V_{out}$, and at small loads, G_L. As a result, the required supply step of 10mV (with ±15% typical accuracy) to perform per-core DVFS over a typical 100× I_L dynamic range requires an impractical 16b PMOS array resolution. Even with limit-cycle oscillations, the load range that can achieve ±1.5mV accuracy is provably limited to $2^{N-6.7}$ at $V_{ref}=V_{in}/2$ (Fig. 18.7.2, top left), which would still require a 14b array resolution that, even if it were feasible to build, would come with linearly (for binary search) or exponentially (for linear search) increased response time (T_R), quiescent power (I_Q), and area.

To enable industry-compliant digital replacement to analog LDOs, this work replaces the PMOS array in a SA-DLDO with a switched-capacitor resistance (SCR) that is created by switching the DLDO output capacitor C_o as shown in Fig. 18.7.1 (top right). The SCR is then frequency modulated through a hysteretic comparator to regulate V_{out} at V_{ref}. Using two non-overlapping clocks, the top (H) and bottom (L) terminals of C_o are alternately connected to (V_{in}, V_{out}) in ϕ_a and (V_{out}, V_{in}) in ϕ_b to charge and discharge C_o by $2\times V_{drop}$, which maximizes the charge delivered per unit capacitance. Importantly, a bilinear instead of series SCR is utilized to ensure 100% of C_o is always connected to V_{out} despite switch commutation. In Fig. 18.7.1, the charge transferred, and hence the SCR equivalent conductance, G_{SCR}, increases linearly with f_{sw} in the slow-switching limit (SSL) region, $G_{SCR,SSL}= 4C_of_{sw}$, until saturating in the fast-switching limit (FSL) region, $G_{SCR,FSL}$, to its maximum value of $G = 1/(2R_{on})$ when T_{SW} is near the SC time-constant, $\tau = C_o/G$, where R_{on} is the switch equivalent on-resistance. Since G_{SCR} can be made arbitrarily small, the SCR-DLDO can, unlike SA-DLDOs, regulate down to arbitrarily low I_L.

Interestingly, placing the SCR connected to the load in feedback with a hysteretic comparator establishes a V_{ref}-controlled relaxation oscillator which accumulates the difference $\Delta V = V_{ref}-V_{out}$ to determine the oscillator period, T_{SW} (=$1/f_{sw}$), that realizes $V_{out} = V_{ref}$. While both SA-DLDOs and the proposed SCR-DLDO follow a search (i.e., integration) control law of the control variable, the hysteretic oscillator is nonlinear and hence abruptly finds the target f_{sw} (i.e., with ~0 acquisition time), enabling a response time that is limited only by comparator latency. Compared to an RLDO [4] that achieves $T_R=NT_{CLK}$ (Fig. 18.7.2, bottom right), the SCR-DLDO achieves $T_R<T_{CLK}$ for the same ΔI_L and C_o, where V_{drop} is typically larger than V_{droop} and hence the charge storage capacity of C_o is amplified by the actively produced voltage swing, $2V_{drop}$. This, along with the efficient SCR-DLDO architecture, enables provably 2N and 2^{2N+1} better FOM over binary- [4] and linear-search [3] SA-DLDOs, respectively.

Using a clocked comparator, T_{SW} can only take on integer multiples of the comparator sampling clock (i.e., $T_{SW} = kT_{CLK}$). In this case, the minimum G_{SCR} step within the SSL is $\sim G_{SCR,SSL}$ f[k]/f_{CLK}. Thus, the achievable V_{out} resolution is $_v$LSB = V_{out} f[k]/f_{CLK}, and hence e_{ss}, unlike in SA-DLDOs, actually enhances with smaller I_L or larger V_{drop} due to a decreasing f[k]. In the FSL region, the SCR $_G$LSB and $_v$LSB values can be found from the G_{SCR} expression in Fig. 18.7.1 (bottom right). Throughout the SSL and FSL regions, e_{ss} is below 1mV (ENOB>10b) across the entire V_{out} and I_L ranges when employing an f_{CLK} of only $4/\tau$ (Fig. 18.7.2, top right). This outperforms the accuracy of a 10b SA-DLDO that not only suffers from e_{ss}=138mV (ENOB=2.9b), but that also suffers from a limited 0.65-to-0.95 V_{out} range at $100\times R_{L,min}$. Above all, increasing the SCR resolution via a finer T_{CLK} actually improves T_R, unlike SA-DLDOs.

It can be shown that, for the same C_o and for the maximum achievable conductance in an LDO (G), the switching frequency, f_{sw}, of the SCR-DLDO is at least 20× lower than the required sampling clock, f_{CLK}, in a SA-DLDO, making the switches' gate-drive losses insignificant in the overall current efficiency, η. Unlike SA-DLDOs, f_{sw} scales with I_L in proportion to the time constant of the load $(C_L+C_o)/G_L$, which exponentially scales the control overhead and the gate drive losses as shown in Fig. 18.7.2 (bottom left), vastly improving efficiency at light loads. Since the SCR-DLDO accuracy exponentially enhances with lower I_L, the comparator external sampling clock (f_{CLK}), and thus its I_Q, can be linearly scaled with I_L by providing f_{CLK} directly from the frequency-scaled clock of the underlying load, as shown in the measurements in Fig. 18.7.2 (bottom left).

Unlike SA-DLDOs, which enter limit-cycle oscillations and suffer from periodic 20-to-70mV ripple [3], output ripple, ΔV_{pp}, in the SCR-DLDO can be reduced by a factor M by time-interleaving M unit SCR cells as shown in Fig. 18.7.3 (top right) for M=4. Since ΔV_{pp} nominally increases linearly with V_{drop} as described by relation (1) in Fig. 18.7.3 (top left), the amount of capacitance that takes part in charge transfer can be scaled by a similar factor to the V_{drop} increase, P, thereby canceling each other per relation (2), without affecting the I_L handling capability, $4C_oV_{drop}f_{sw}$. The proposed binary-ripple-control (BRC) scheme, shown in Fig. 18.7.3 (bottom right), divides the capacitance and conductance of each of the 4 interleaved phases into 5 binary-weighted banks that are enabled by EN[4:0] and a redundant always-on LSB bank, where EN[4:0] can be provided from an existing battery state-of-charge monitoring circuit or the switching regulator supplying V_{in}. The SCR-DLDO 1× unit cell, non-overlap circuit, and comparator schematics are shown in Fig. 18.7.3 (bottom).

The proposed SCR-DLDO is fully integrated in 0.00137mm² core area in 65nm CMOS with C_o=200pF and C_L=165pF that mimics the inherent capacitance of a 3mA digital load with 5% activity. Steady-state measurement results in Fig. 18.7.4 demonstrate the efficacy of the SCR-DLDO in realizing high accuracy: a steady-state error of at most ±1.55mV is measured across all desired V_{out} values between 0.3-0.8V over V_{in} corners of 0.5V and 0.9V, all over a 10µA-3mA (300×) dynamic range (Fig. 18.7.4 left and top right). For V_{in}=0.9V, f_{CLK} is set to 1GHz to enable a T_{SW} LSB of $\tau/4$, which would theoretically achieve <±1mV accuracy as in Fig. 18.7.2. Due to the comparator frequency-dependent s-shaped offset (Fig. 18.7.4 left), e_{ss} grows to ±1.55mV, which still enables up to a 174.7× accuracy improvement over a simulated 10b shifter-based DLDO in Fig. 18.7.2 (top left). Since e_{ss} enhances with lower I_L, f_{CLK} is safely linearly scaled from 1GHz down to 4MHz at 10µA to scale comparator I_Q (Fig. 18.7.2 bottom, left), which improves η and I_L-range by 50.3% and 10×, respectively (Fig. 18.7.4, bottom right).

Compared to the RLDO in [4], at V_{in}=0.5V the SCR-DLDO reduces the measured error by 2.4-4.3× for V_{out} ranging between 0.3-0.45V (Fig. 18.7.4, top right), demonstrating the accuracy advantage of SCR-DLDOs over SA-DLDOs. The SCR-DLDO achieves a peak η of 99.3%, and at V_{out}=0.3V, operates from 1.5µA-1.75mA with η>70% (a 1,167× dynamic range), exceeding [4] by 5× and improving light-load efficiency by 37% (Fig. 18.7.4, bottom right). With I_Q=48.4µA (f_{CLK}=1GHz), the SCR-LDO achieves a measured T_R=2.48ns with V_{droop}=20.5mV in response to periodic 50µA-to-3.3mA on-chip load swings occurring within 200ps. Thus, the achieved FOM is 36.9ps, for a 5.4× improvement over prior-art (Fig. 18.7.6). BRC operation is demonstrated to reduce ripple from 161.3mV to 21.7mV at the worst-case voltage corner (V_{in}=0.9V to V_{out}=0.3V) across a 300× I_L range (Fig. 18.7.5, bottom left and right) and at the worst-case current corner (V_{in}=0.9V with I_L=10µA) for V_{out} between 0.3-0.8V (bottom middle). A die photo is shown in Fig. 18.7.7.

References:

[1] Z. Toprak, et al., "Distributed System of Digitally Controlled Microregulators Enabling Per-Core DVFS for the Power8 Microprocessor," *ISSCC*, pp. 98-99, 2014.

[2] Texas Instruments TPS659037, "Power Management IC (PMIC) for ARM Cortex A15 Processors," Available Online: http://www.ti.com.

[3] S. B. Nasir, et al., "A 0.13µm Fully Digital Low-Dropout Regulator with Adaptive Control and Reduced Dynamic Stability for Ultra-Wide Dynamic Range," *ISSCC*, pp. 98-99, 2015.

[4] L. G. Salem, et al., "A 100nA-to-2mA Successive-Approximation Aigital LDO with PD Compensation and Sub-LSB Duty Control Achieving a 15.1ns Response Time at 0.5V," *ISSCC*, pp. 340-341, 2017.

 978-1-5090-4940-0/18/$31.00

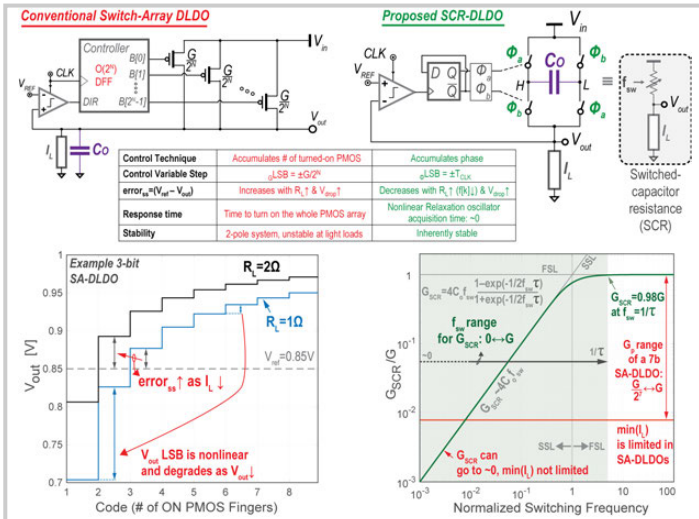


Figure 18.7.1: A conventional switch-array DLDO (left) and its accuracy problem (bottom left); proposed SCR-DLDO using a switched-capacitor resistance (right).

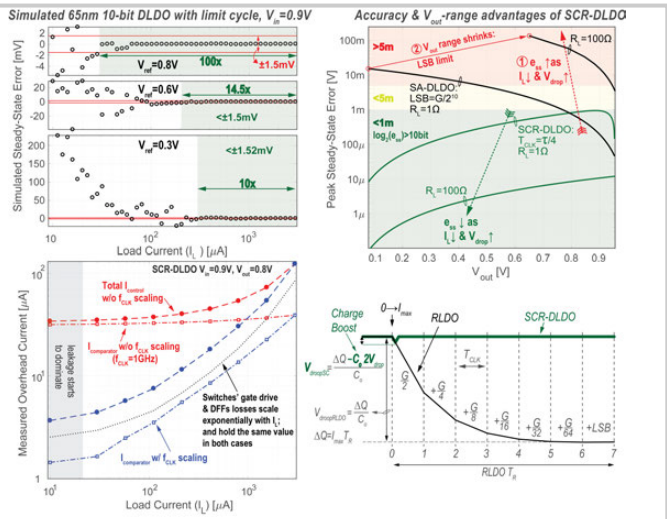


Figure 18.7.2: Accuracy advantage of a 1V SCR-DLDO over SA-DLDOs (top); overhead current reduction with f_{CLK} scaling and the fast response time advantage of the proposed SCR-DLDO (bottom).

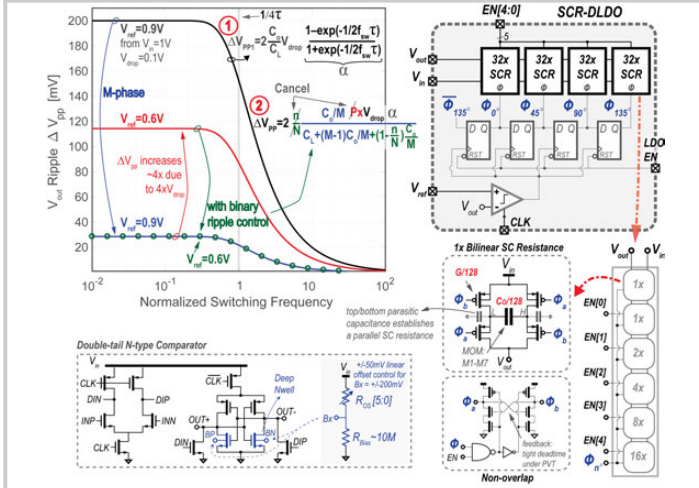


Figure 18.7.3: Ripple reduction via the proposed scheme (top-left); top-level block diagram of the SCR-DLDO (top-right); cell partitioning, and the schematics of the SCR 1× unit-cell, non-overlap circuitry, and comparator (bottom).

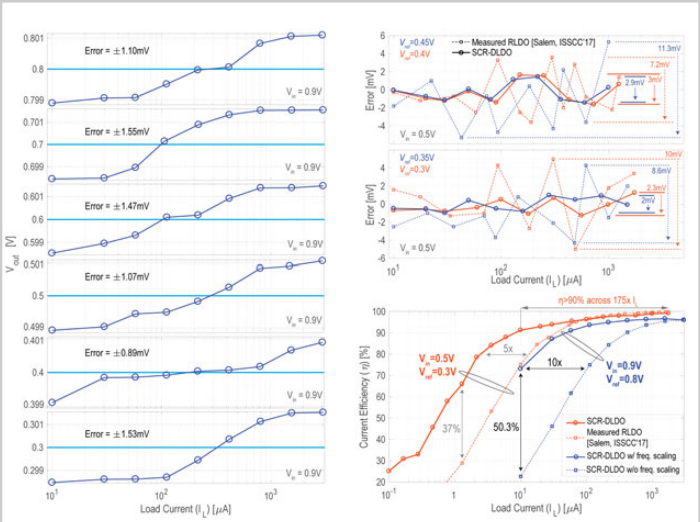


Figure 18.7.4: SCR-DLDO output voltage and current efficiency at the corners V_{in} = 0.9V and V_{in} = 0.5V.

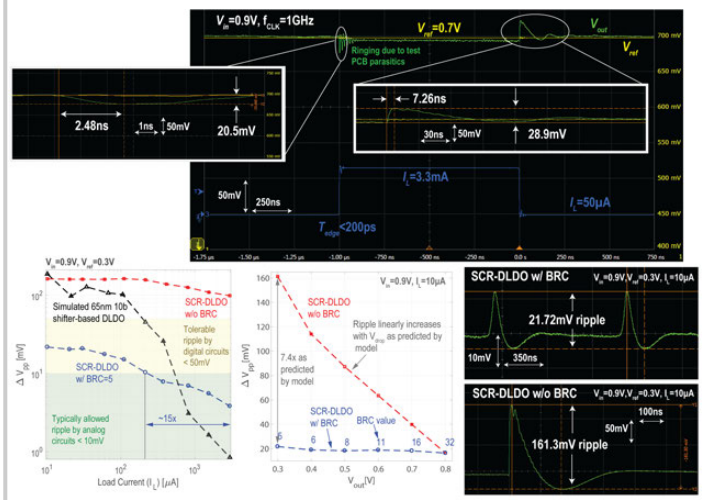
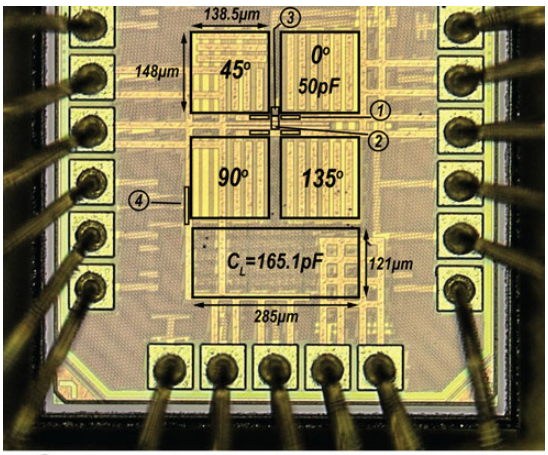


Figure 18.7.5: Measured transient response of the SCR-DLDO to an on-chip periodic load-step (top). Illustration of the efficacy of the proposed binary ripple control scheme in mitigating the SCR ripple (bottom).

Design	Huang, ISSCC'17	Tsou, ISSCC'17	Kim, ISSCC'17	Salem, ISSCC'17	This Work
Process	65nm	40nm	65nm	65nm	65nm
Active area [mm²]	0.03	0.193	0.03	0.0023	0.00137
V_{in} [V]	0.6	0.6 – 1.1	0.45-1	0.5 – 1	0.5-0.9
V_{out} [V]	0.5	0.5 – 1.0	0.4-0.95	0.3 – 0. 45	0.3-0.8
Loop Actuator	9b PMOS array	7b PMOS array	4x 7b PMOS array	7b PMOS array	SC Resistance
Control Loop	Barrel Shifter+Analog Assisted	PID	Event-Driven multi-bit ADC	SAR/PD/PWM	Hysteretic Relaxation Oscillator
Limit-Cycle Capability for Accuracy Improvement	No limit-cycle	No limit-cycle	No limit-cycle	No limit-cycle	Not Applicable
Load range	2m– 12mA (6x)	15mA – 210mA (14x)	14μA – 3.36mA (~100x *)	100nA – 2mA (20,000x)	10μA-3mA (300x)
Load range with η>90%	N.R.	N.R.	~ 10x	33.6μA – 2mA (60x)	10μA-1.75mA (175x)
C_L [nF]/ Total C	0/0.1	20 / 20	0.1 / 0.1	0.4/0.4	0.165/ 0.365
Quiescent I_Q [μA] during load transient test	N.R.	N.R.	258	14	48.4
V_{droop} @ load step size for load transient test	105mV @ 10mA	38mV @ 200mA	34mV @ 1.44mA	40mV @ 1.06mA	20.5mV @ 3.25mA
Response time† T_R [ns] from relation: $C_{out}V_{droop}/\Delta I$	1.05‡	1000††	2.36	15.1	2.3 (2.48 measured)
FOM² for load transient test [ns]	-	0.493††	0.423	0.199	0.0343 (0.0369)
Load step rise/fall time for load transient test²	1n†	1000	N.R.	< 1ns	<200ps
Peak current efficiency η [%]	N.R.	N.R.	99.2	99.8	99.3
Sampling clock range	10MHz	N.R.	Not Applicable	1MHz – 240MHz (240x)	100K-1.55GHz (15,500x)
Steady-state error (mV)	N.R.	<150†††	<15	<5.2	<1.55
DC Accuracy: peak-error/V_{in}	N.R.	±13.6%†††	±1.5%	1.04%	±0.17%
Load regulation: worst peak-peak error/ΔI [mV/mA]	N.R.	0.8	<15	< 11.3 across range	< 1.03 across range

N.R. = Not Reported
† load rise/fall time should be < T_R/10 for a valid *unit-step* FOM measurement
‡ FOM = $C_L V_{droop}/(I_{max}-I_{min}) \times I_Q/(I_{max}-I_{min})$, P. Hazucha et al., JSSC'05
§ Measured ranges are only depicted
† I_Q consumption during transient test was not reported. Also, the reported T_R is approximately the load rise/fall time, and hence the reported FOM is for the unit-ramp response and not the unit-step response.
†† Observed from transient measurement
††† Best-case theoretical value across the reported 15-150mA (10x) range
* For the same V_{ref}

Figure 18.7.6: Comparison of the proposed SCR-DLDO with state-of-the-art switch-array DLDOs illustrating the smallest area, best FOM, and highest accuracy.



① PMOS switches & non-overlap circuit of cell 0° [275μm²]
② Four D flip-flops [123μm²]
③ Comparator [147μm²]
④ On-chip 5-bit programmable load

Figure 18.7.7: Micrograph of the fabricated SCR-DLDO chip.

978-1-5090-4940-0/18/$31.00

18.8 A High-Efficiency and Fast-Transient Digital-Low-Dropout Regulator with the Burst Mode Corresponding to the Power-Saving Modes of DC-DC Switching Converters

Jian-He Lin[1], Yu-Sheng Ma[1], Chia-Ming Huang[1], Li-Chi Lin[1], Chiao-Hung Cheng[1], Ke-Horng Chen[1], Ying-Hsi Lin[2], Shian-Ru Lin[2], Tsung-Yen Tsai[2]

[1]National Chiao Tung University, Hsinchu, Taiwan
[2]Realtek Semiconductor, Hsinchu, Taiwan

Integrated power management (PM) in a system-on-a-chip (SoC) includes a high-efficiency DC-DC switching regulator (SWR) for a high conversion ratio and multiple cascaded digital-low-dropout (DLDO) regulators for post regulation to different functional blocks. Efficient power-saving modes in the SWRs improve the light-load efficiency effectively, e.g. burst mode, skip mode, pulse frequency mode (PFM), and diode emulation mode (DEM) in the constant on-time (COT). Unfortunately, a cascaded DLDO consumes significant power to suppress a large voltage ripple DV_{SWR} from the SWR, even using recent DLDO techniques [1-6], illustrated on the left of Fig. 18.8.1. Overall, the light-load efficiency of the PM seriously decreases. A DLDO with barrel-shifter-based control [1] induces large voltage ripple due to the limiting cycle oscillation (LCO), and the DLDO with freeze mode [2] for power reduction is exposed to large voltage ripples from the SWRs (in power saving modes). The large voltage ripples result in the DLDO frequently switching between the normal and freeze modes and consuming power. The recursive all-digital LDO (RLDO) [3] abruptly changes its control code Q[6:0] due to oscillations between hybrid proportional-derivative successive-approximation recursive (PD-SAR) and pulse width modulation (PWM) duty control, while ΔV_{SWR} is larger than the pre-defined hysteretic window. In power saving modes, the obvious disadvantage is that state-of-the-art DLDO designs cause extra switching loss and induce large output voltage ripple ΔV_{OUT} due to large ΔV_{SWR}. Thus, this paper proposes a DLDO employing a burst mode technique (BMT) to reduce the DV_{OUT}, thereby enhancing the overall light-load efficiency corresponding to the power saving modes in SWRs. The proposed non-linear switch control (NLSC) technique reduces both the number of on/off power switches and varies the switching frequency corresponding to the ΔV_{SWR}. Moreover, the proposed transient enhance (TE) technique improves transient performance when the DLDO leaves burst mode.

The DLDO needs to differentiate transient response and power saving modes since the ΔV_{SWR} is large in both conditions, but with different slew rates. The slew-rate detection (SRD) technique in Fig. 18.8.2 checks signals SR[6:0], representing the difference between current and previous state Q_N[6:0]. If SR[6:0] is larger than 10, a transient response occurs and the DLDO needs to react to the transient conditions. Conversely, if the value is less than or equal to 10, power-saving mode commences and the DLDO enters burst mode by setting the signal 'Burst' to 1. In burst mode, the DLDO can reduce the number of on/off switched by the linear switch control (LSC), which increases (or decreases) Q_{BST}[6:0] linearly to add (or subtract) the burst current I_{BST} ($=Q_{BST}\times I_{UNIT}$) to (or from) the Integral (I) part current I_I ($=I_N+I_{BST}$), where I_N ($=Q_N\times I_{UNIT}$) is the normal part current. The power switches of the DLDO are turned on and off corresponding to the falling and rising of V_{SWR}, respectively, in the upper right of Fig. 18.8.2. The LSC turns on one PMOSFET every $T_F/(Q_{N(MAX)}-Q_{N(MIN)})$ time periods when V_{SWR} Is falling. Likewise, the LSC turns off one switch every $T_R/(Q_{N(MAX)}-Q_{N(MIN)})$ time periods when V_{SWR} Is rising. I_{step} of the I_{BST} gradually increases and I_N increases accordingly, since the source-to-drain (V_{SD}) of the power switch array varies from 50mV to 100mV as V_{SWR} is rising. The overall response implies $\Delta I_I > 0$ in the beginning, but $\Delta I_I < 0$ at the end of the rise of V_{SWR}. Consequently, it induces a large ΔV_{OUT}. The ΔV_{OUT} of the LSC technique is smaller than that without the burst mode control, but its switching frequency is the same. Thus, the proposed NLSC technique replaces the LSC to vary the switching frequency to further suppress the ΔDV_{OUT} and reduce power loss.

Figure 18.8.3 illustrates the complete DLDO regulator, comprising the BMT with the NLSC technique to further reduce ΔV_{OUT} by varying the switching frequency corresponding to ΔV_{SWR}. In the bottom of Fig. 18.8.3, as V_{SWR} rises, Rising[10:0] are divided into three parts to turn off P-type power switches rather than the equal division in the LSC technique. The NLSC adjusts the turn-off frequency by changing the switching-frequency select signal F_{SEL}[1:0]. Initially, the NLSC speeds up the switching frequency, where Q_{BST}[6:0] decreases by 1 every $2/3\times T_1$ ($=T_R/(Q_{N(MAX)}-Q_{N(MIN)})$ time periods. At the end, it slows down the clock frequency, where Q_{BST}[6:0] decreases by 1 every $4/3\times T_1$ time periods to suppress DV_{OUT}. Analogously, the switching frequency varies from slow to fast as V_{SWR} falls. The P-type power switch array contains three parts: Proportional (P) part, Derivative (D) part (D1-3[2:0]), and Integral (I) part (I[126:0]). The I part is controlled by the BMT technique to improve load regulation and reduce ΔV_{OUT} in burst mode. Moreover, the transient enhancement (TE) controls the P, I and D parts to improve transient response.

Figure 18.8.4 illustrates the real-time load tracking (RTLT) circuit and the state control in the TE circuit. Its operation consists of three parts: (1) searching the value of $Q_{N(MAX)}/Q_{N(MIN)}$, (2) coarse setting, and (3) fine setting. In the coarse setting of [4], it needs ($Q_{N(MAX)}$-$Q_{N(MIN)}$) cycles to set the current upper limit (UL) and lower limit (LL) signals until the value of (UL-LL) is equal to 3 (Fig. 18.8.4 bottom). The disadvantage is the coarse setting time increases greatly if $Q_{N(MAX)}$ is much larger than $Q_{N(MIN)}$. The proposed TE technique adds the searching mode to capture the peak and valley values of I_{OUT} in $Q_{N(MAX)}$ and $Q_{N(MIN)}$, respectively, in the beginning of transient response. In the following coarse setting period, the advantage is that it spends one cycle to store Avg., Avg.+2 and Avg.-2 in Q_N, UL and LL, respectively, where the average value 'Avg.' is equal to $(Q_{N(MAX)}+Q_{N(MIN)})/2$. After the coarse setting, CQE reduces to 4 and the difference between UL and LL is within 4. Finally, in the fine setting period, UL counter (ULC) and LL counter (LLC) record how many times the Q_N reaches UL or LL, respectively. If the value of ULC is larger than that of LLC, LL is increased by 1. Conversely, if the value of ULT is less than that of LLT, UL is decreased by 1. Consequently, the value of (UL-LL) reduces rapidly to 1. Thus, the CQE is equal to 1 and the ΔV_{OUT} is minimized.

Measurement results in Fig. 18.8.5 show the steady-state and transient performance. In steady state, the proposed DLDO ensures ΔV_{OUT}<6mV if ΔV_{SWR} varies from 0 to 50mV. The ΔV_{OUT} of [1] may be increased to 50mV under the same ΔV_{SWR}. The proposed DLDO achieves high current efficiency of 80% in light load under 50mV ΔV_{SWR}, and the peak current efficiency is 99.8% at heavy loads. [1] suffers from ΔV_{SWR} in power-saving mode and the ΔV_{OUT} is larger than 50mV. When the proposed LSC technique is activated, the ΔV_{OUT} is reduced from 50mV to 15mV when the SWR operates in the PFM mode. Furthermore, the NLSC technique can suppress the ΔV_{OUT} to 5mV and 6mV when the SWR operates in the PFM mode and the burst mode, respectively. Without the TE, the load current changes from 1 to 20mA, and undershoot and overshoot voltages are 40mV and 20mV, respectively, with a settling time $T_{SETTLING}$ of 1.7μs. However, in the proposed DLDO, undershoot and overshoot voltages plus $T_{SETTLING}$ are reduced to 40mV, 0mV and 1.3μs, respectively.

The top left of Fig. 18.8.6 shows the power loss between the freeze mode design [2], and the proposed DLDO, which reduces the total power consumption to 9.2μA due to variable switching frequency and disabling the RTLT in steady state. The top right of Fig. 18.8.6 endorses that the proposed technique affected by the PVT variations is below 0.6% at the V_{OUT} of 0.5V and I_{LOAD} is 20mA when the V_{SWR} ranges from 0.6 to 1.1V. The comparison table shows the DLDO has the lowest output voltage ripple, and low quiescent current due to the burst mode operation when the SWR operates in the power-saving modes. A test chip was fabricated in 40nm and Fig. 18.8.7 shows the chip micrograph with an area of 0.18mm^2.

References:

[1] Y. Okuma, et al., "0.5-V input digital LDO with 98.7% current efficiency and 2.7-μA quiescent current in 65nm CMOS," *CICC*, 2010.
[2] Y.-J. Lee, et al., "A 200mA Digital Low-Drop-Out Regulator with Coarse-Fine Dual Loop in Mobile Application Processors," *ISSCC*, pp. 150-151, 2016.
[3] L. G. Salem, et al., " A 100nA-to-2mA Successive-Approximation Digital LDO with PD Compensation and Sub-LSB Duty Control Achieving a 15.1ns Response Time at 0.5V," *ISSCC*, pp. 340-341, 2017.
[4] W.-J. Tsou, et al., "Digital Low-Dropout Regulator with Anti PVT-Variation Technique for Dynamic Voltage Scaling and Adaptive Voltage Scaling Multicore Processor," *ISSCC*, pp. 338-339, 2017.
[5] D. Kim, et al., "A 0.5V-VIN 1.44mA-Class Event-Driven Digital LDO with a Fully Integrated 100pF Output Capacitor," *ISSCC*, pp. 346-347, 2017.
[6] Y.-H. Lee, et al., "A Low Quiescent Current Asynchronous Digital-LDO with PLL-Modulated Fast-DVS Power Management in 40 nm SoC for MIPS Performance Improvement," *IEEE JSSC*, vol. 48, no. 4, pp. 1018-1030, 2013.

 978-1-5090-4940-0/18/$31.00

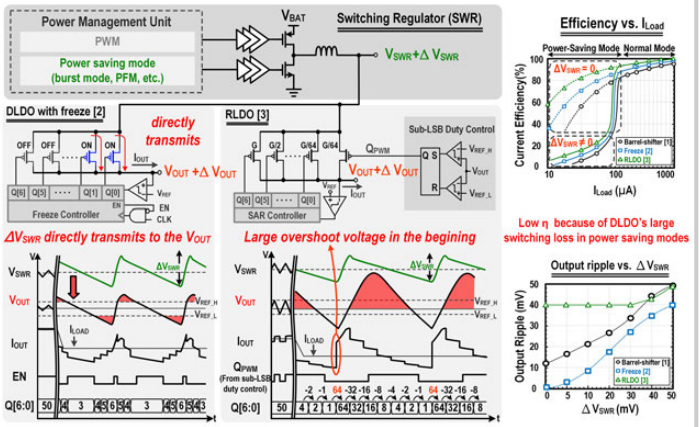


Figure 18.8.1: Light-load efficiency decreases and output voltage ripple increases due to the increasing supplying voltage ripple DV_{SWR} from the SWR in power-saving mode at light loads.

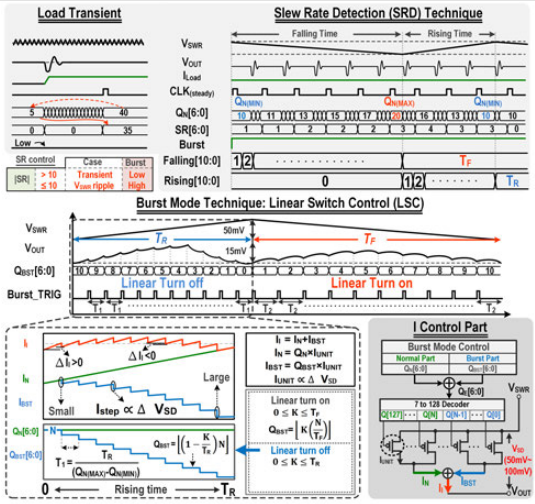


Figure 18.8.2: Proposed slew-rate detector (SRD) decides when the DLDO burst mode starts; the switch linear control (LSC) shows the disadvantages caused by the non-ideal effect of the V_{SD} variations in the power switch array.

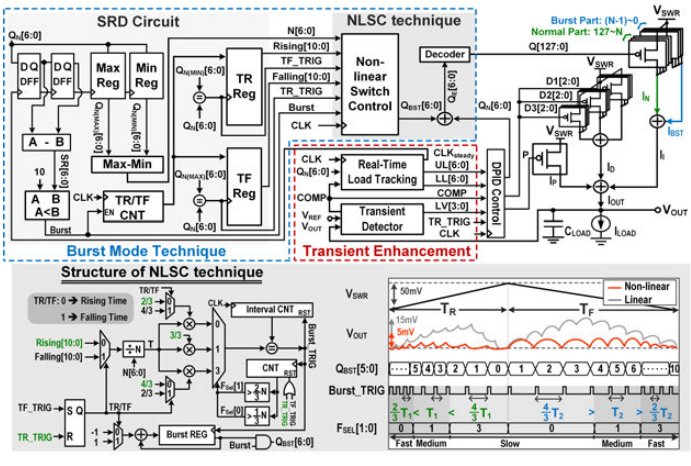


Figure 18.8.3: Overall architecture of the DLDO with burst mode and the proposed non-linear switch control (NLSC) technique.

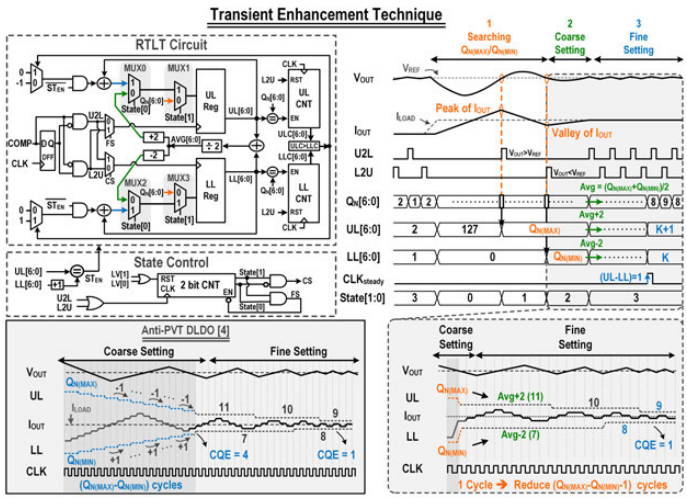


Figure 18.8.4: The proposed transient-enhancement (TE) technique and operation waveforms.

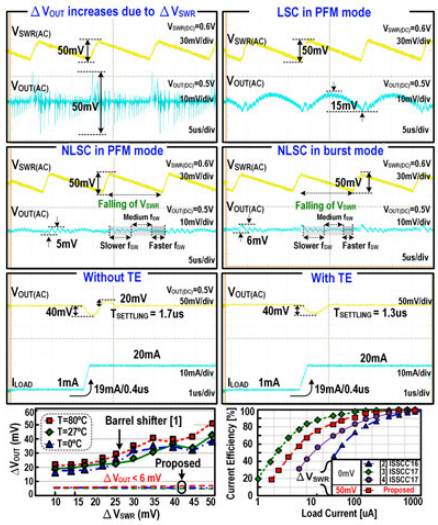
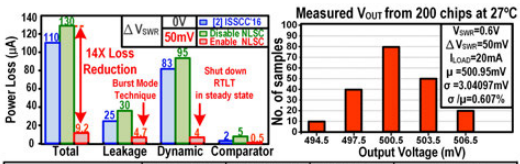


Figure 18.8.5: Measurement results in steady state in power saving mode with LSC and NLSC, the transient performance with TE.



	[2] ISSCC'16	[3] ISSCC'17	[4] ISSCC'17	This Work
Type	Digital	Digital	Digital	**Digital**
Process	28nm	65nm	40nm	**40nm**
V_{in} [V]	1.1	0.5-1	0.6-1.1	**0.6-1.1**
V_{out} [V]	0.9	0.3-0.45	0.5-1	**0.5-1**
Max. I_{LOAD} [mA]	200	2	210	**20**
Max. F_{SAMPLE} [MHz]	N.A.	240	N.A.	**100**
Load capacitor [nF]	23.5	0.4	20	**4.7**
Load regulation [mV/mA]	<0.01	<5.6	<0.075	**<0.1**
Transient droop @ ΔI_{LOAD}	120 mV @180mA	40 mV @1.06mA	36 mV @200mA	**40 mV @19mA**
Settling time @ ΔI_{LOAD}	<40μs @180mA	15ns @1.06mA	<1.3μs @200mA	**1.3μs @19mA**
Output voltage ripple [mV]	20-40	20-40	2.3-3	ΔV_{SWR}=0: 3 ΔV_{SWR}=50: **<6**
I_Q [μA]	110	14	22.6-98.5	ΔV_{SWR}=0: 19.6 ΔV_{SWR}=50: **<10**
FOM	1.91pF	199pF	0.16pF	0.19pF

$FOM = (I_Q/\Delta I_{LOAD,max})\cdot(V_{DROOP,max}/V_{OUT})\cdot C_{LOAD}$

Figure 18.8.6: Performance summary and comparison table.

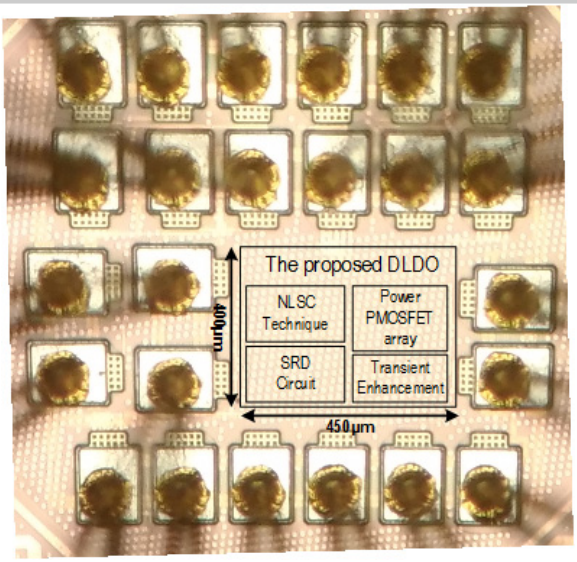


Figure 18.8.7: Chip micrograph.

978-1-5090-4940-0/18/$31.00

Session 19 Overview: *Sensors and Interfaces*

ANALOG SUBCOMMITTEE

Session Chair:
Man-Kay Law
University of Macau, Macau, China

Associate Chair:
Taeik Kim
Samsung Electronics, Hwaseong, Korea

Subcommittee Chair: ***Kofi Makinwa,*** Delft University of Technology, Delft, The Netherlands

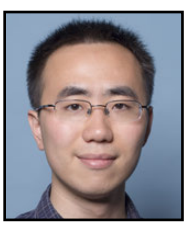
This session highlights the advances in state-of-the-art temperature, current, physical and chemical sensors. Three energy-efficient CMOS temperature sensors with the best figures of merit down to 34fJ·K^2 are reported. Two current sensors (Papers 19.4 and 19.5) are presented, one demonstrating a high room-temperature gain accuracy with a ±4A input range, while the other shows a 160dB DR biosensor readout with 7ppm INL. An energy-efficient pressure-sensing system (Paper 19.6) and a high-resolution readout IC (Paper 19.7) are also reported. An energy-efficient CO_2 sensor achieving 2× better resolution (94ppm) and >10× lower energy consumption (12mJ/meas) than the prior art is also described (Paper 19.8).

8:30 AM

19.1 An 8b Subthreshold Hybrid Thermal Sensor with ±1.07°C Inaccuracy and Single-Element Remote-Sensing Technique in 22nm FinFET

C-Y. Lu, Intel, Hillsboro, OR

In Paper 19.1, Intel Advanced Design describes a hybrid temperature sensor using subthreshold NMOS devices and parasitic PNPs. It enables single-element remote sensing with a small size (0.00021mm^2) and achieves 3× better Resolution-FOM of 0.54nJ·K^2 using a 22nm FinFET process.

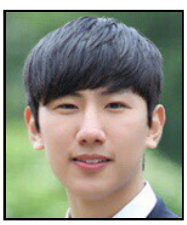


9:00 AM

19.2 A 0.25mm^2 Resistor-Based Temperature Sensor with an Inaccuracy of 0.12°C (3σ) from -55°C to 125°C and a Resolution FOM of 32fJ·K^2

S. Pan, Delft University of Technology, Delft, The Netherlands

In Paper 19.2, Delft University of Technology presents the first resistor-based CMOS temperature sensor capable of operating over the full military range (-55°C to 125°C). It achieves state-of-the-art energy efficiency of 32fJ·K^2 with 3× smaller area.



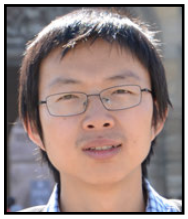
9:30 AM

19.3 A 0.53pJ·K^2 7000μm^2 Resistor-Based Temperature Sensor with an Inaccuracy of ±0.35°C (3σ) in 65nm CMOS

W. Choi, Yonsei University, Seoul, Korea

In Paper 19.3, Yonsei University describes a highly compact resistor-based CMOS temperature sensor for dense thermal monitoring in nanometer semiconductor processes. The proposed PPF-based FLL shows 13× smaller area and 15× higher resolution-FOM when compared to similar prior art.

 978-1-5090-4940-0/18/$31.00

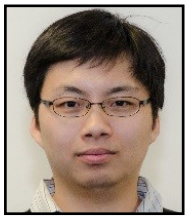


10:15 AM

19.4 A ±4A High-Side Current Sensor with 25V Input CM Range and 0.9% Gain Error from -40°C to 85°C Using an Analog Temperature Compensation Technique

L. Xu, Delft University of Technology, Delft, The Netherlands

In Paper 19.4, Delft University of Technology presents a current sensor that supports a ±4A range with 0.05% gain error at room temperature. It achieves a resolution of 150μA_{rms} in a conversion time of 2ms while consuming 10.9μA, which is 10× more energy efficient than the state-of-the-art.

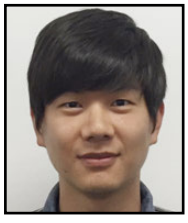


10:45 AM

19.5 A Current-Measurement Front-End with 160dB Dynamic Range and 7ppm INL

C-L. Hsu, University of California, San Diego, La Jolla, CA

In Paper 19.5, the University of California, San Diego presents an accurate current measurement with wide dynamic range and high linearity for biosensor readout. With an asynchronous Hourglass ADC, the system achieves 160dB DR and 7ppm INL.



11:15 AM

19.6 A 2.5nJ Duty-Cycled Bridge-to-Digital Converter Integrated in a 13mm³ Pressure-Sensing System

S. Oh, University of Michigan, Ann Arbor, MI

In Paper 19.6, the University of Michigan describes a duty-cycled bridge-to-digital converter for a small battery-operated pressure-sensing system. By heavily duty-cycling the bridge sensor's excitation voltage, the system requires 6000× less excitation power. It consumes 2.5nJ/conversion and achieves 49.2dB SNR and 10.6pJ/conv-step FOM.

11:45 AM

19.7 A 21.8b Sub-100μHz 1/f Corner 2.4μV-Offset Programmable-Gain Read-Out IC for Bridge Measurement Systems

J. Jun, Seoul National University, Seoul, Korea

In Paper 19.7, Seoul National University presents a fully integrated gain-programmable read-out IC with a maximum resolution of >21b. With the proposed system chopping technique utilizing an on-chip reconfigurable digital filter, the system achieves a low 1/f corner <100μHz and a maximum FOM that is 5.3dB higher than previous work.

12:00 PM

19.8 A Phase-Domain Readout Circuit for a CMOS-Compatible Thermal-Conductivity-Based Carbon Dioxide Sensor

Z. Cai, Delft University of Technology, Delft, The Netherlands and NXP Semiconductors, Eindhoven, The Netherlands

In Paper 19.8, Delft University of Technology presents a CMOS-compatible thermal-conductivity-based CO_2 sensor using a high-resolution phase-domain delta-sigma modulator to sense the thermal time constant of a hot tungsten wire. The sensor achieves a CO_2 sensor resolution of 94ppm while dissipating only 12mJ per measurement.

19.1 An 8b Subthreshold Hybrid Thermal Sensor with ±1.07°C Inaccuracy and Single-Element Remote-Sensing Technique in 22nm FinFET

Cho-Ying Lu[1], Surej Ravikumar[1], Amruta D. Sali[1], Matthias Eberlein[2], Hyung-Jin Lee[1]

[1]Intel, Hillsboro, OR, [2]Intel, Neubiberg, Germany

Thermal sensors are commonly used in modern highly dense systems-on-chip (SoC) to provide information about die temperature for thermal protection or performance optimization. To enable the deployment of multiple sensors in an SoC, the power and size of such sensors has been steadily reduced. Although most solutions are PNP-based [1-4], recently a low-power NPN-based current-mode thermal sensor [5] was implemented to meet the power and form-factor requirement with a robust architecture. However, since NPN devices are only available in a triple-well process, its use in low-cost dual-well processes is limited. This paper demonstrates a current-mode hybrid thermal sensor in an advanced 22nm FinFET process [6] based on subthreshold NMOS transistors and a parasitic PNP [7]. A simple voltage-based single-point soft-trimming was implemented to mitigate the sensitivity of the sensor to the subthreshold factor variability during manufacturing. Instead of placing the whole sensor system in multiple locations in a system, the hybrid architecture also supports single-element remote sensing. Only a 0.00021mm^2 PNP device placed in the area of interest and a signal connection to the main sensor are then needed for temperature sensing. The sensor system achieves +/-1.07°C (±3σ) precision with 0.0043mm^2 silicon area and 50uA current consumption from a 1V supply.

Figure 19.1.1 shows the architecture of the sensor along with the operational concept. Similar to the aforementioned current-mode sensor [5], the circuit senses temperature by comparing a PTAT ("proportional-to-absolute-temperature") current to a CTAT ("complementary-to-absolute-temperature") current. The PTAT current is generated by the pseudo-differential NMOS pair, biased in the subthreshold region, and the PMOS current mirror. By establishing a 1:N (N=4 in this implementation) current density ratio between M1 and M2, a PTAT voltage ΔVgs is generated across R1, and hence a PTAT current (Iptat). With the current mirror providing the DC current bias (Id), the PNP device at node X absorbs the current difference between Id and Iptat and generates a CTAT voltage, i.e. V_{EB}. The voltage at node Y is equal to V_{EB}-ΔVgs and so a CTAT current (Ictat) is created by the R-DAC R2, controlled by digital code D_{out}. The negative-feedback loop formed by the differential amplifier and the push-pull circuit stabilizes the system, detects the voltage difference between the drains of the NMOS pair, and provides a compensation current (Idiff) to balance the difference between Ictat and Iptat. Successive-approximation (SAR) logic then equalizes Ictat and Iptat by sweeping D_{out} and detecting the polarity of Idiff, with the input of the push-pull pair acting as a current comparator. A unique digital code for each temperature T (in Kelvins) is provided as shown in Fig. 19.1.2, where Rr is the ratio between R1 (2.6kΩ), and LSB resistance of R2, (125Ω), and n is the subthreshold factor. Vgo (~1.2V) is related to the silicon bandgap and tc is the slope of VEB with temperature (~ -2mV/°C). The transfer function shows that the system is robust to both resistor and PNP variation (only constant offset from the nominal tc) but is sensitive to n-factor and transistor random mismatch. To address this, the transistors are sized properly for matching and a subthreshold factor trimming methodology is implemented.

Simulations show that the subthreshold factor variation causes at least +/-3°C (3σ) temperature error. In order to reduce this, a voltage-based single-point soft-trimming, inspired by [8], is implemented. As shown in Fig. 19.1.3, by disconnecting the PNP device with a switch (SW1) and providing the known voltages V_1 and V_2 to the node X through SW3, at a fixed temperature (here 298K but not limited to), output codes D_{out1} and D_{out2} can be acquired based on the transfer function in Fig. 19.1.2. By dividing the ΔD_{out}=Dout1-Dout2 from the simulation ($\Delta D_{out,sim}$) with that from the silicon ($\Delta D_{out,DUT}$), the DUT subthreshold factor (n_{DUT}) can be determined with respect to the subthreshold factor in the model (n_{model}) from the equation $n_{DUT}=n_{model}\times(\Delta D_{out,sim}/\Delta D_{out,DUT})$. With the knowledge of the DUT subthreshold factor, The output code D_{out} from silicon can then be easily trimmed by firmware or by lookup table.

Figure 19.1.3 also shows the sensor architecture in single-element remote sensing mode. This is done by using switches SW1 and SW2 to connect the remote PNP and local PNP to node X. The upper plot in Fig. 19.1.4 shows the procedure needed to sense a remote temperature (T_{remote}) in three steps: 1: local sensing (D_{local}) with the local PNP at local temperature T_{local}, 2: remote sensing (D_{remote}) with the remote PNP under T_{remote} and sensor core under T_{local}, and 3: T_{remote} extraction in the digital domain based on the equation in Fig. 19.1.4 with the assumption of good matching between local and remote PNPs. Any parasitic resistance from the wire connections between the remote PNP and the sensor core can be derived and trimmed during the voltage-based single-point soft-trimming. In the bottom plot of Fig. 19.1.4, the simulated temperature error from -10°C to 105°C from remote temperature readings at two different local temperatures (30°C and 80°C) is shown and is observed to be less than <±1°C. This implementation requires only 0.00021mm^2 area, a 20× silicon reduction in comparison to a full sensor, and a wire connection to the area of interest without power dependency.

Two plots in the left side of Fig. 19.1.5 depict the measured inaccuracy of the 8b hybrid thermal sensor at 1V supply from 38 samples fabricated in an advanced 22nm FinFET process, including trimming at 25°C. The temperature error without trimming is +/2.81°C (±3σ) due to the sensitivity of the sensor to systematic subthreshold factor variation. The variation across samples is compensated by trimming, reducing the error to +/-1.07°C (±3σ), which is partially limited by the quantization error (0.4 to 0.9°C). Since a significant portion of the sensing inaccuracy results from quantization error, this implementation shows the potential of the architecture to achieve better accuracy by increasing the R-DAC resolution to reduce quantization error. To demonstrate its strong intrinsic PSRR performance, the supply voltage was swept at 25°C on all 38 samples. The maximum code change is 1LSB in the range from 0.97V to 1.3V, as shown in the right plot of Fig. 19.1.5. Figure 19.1.6 summarizes the sensor's performance compared to the state of the art. This hybrid sensor demonstrates its power and silicon efficiency with 0.54nJ·K^2 Resolution-FOM and 0.0043mm^2 silicon area in a low-cost dual-well FinFET process.

Acknowledgements:
The authors would like to acknowledge S. Rami and J. Bondie for the support of the silicon measurement and D. Duarte and A. Kornfeld for the valuable design discussions.

References:

[1] Y.-C. Hsu, et al., "An 18.75μW Dynamic-Distributing-Bias Temperature Sensor with 0.87°C(3σ) Untrimmed Inaccuracy and 0.00946mm^2 Area," *ISSCC,* pp. 102-103, Feb. 2017.

[2] B. Yousefzadeh and Kofi A. A. Makinwa, "A BJT-Based Temperature Sensor with a Packaging Robust Inaccuracy of ±0.3°C (3σ) from -55°C to +125°C After Heater-Assisted Voltage Calibration," *ISSCC,* pp. 162-163, Feb. 2017.

[3] M.-C. Chuang, et al., "A Temperature Sensor with a 3 Sigma Inaccuracy of ±2°C Without Trimming from -50°C to 150°C in a 16nm FinFET Process," *ESSCIRC*, pp. 271-274, Sept. 2015.

[4] T. Oshita, et al., "Compact BJT-Based Thermal Sensor for Processor Applications in a 14 nm tri-Gate CMOS Process," *IEEE JSSC*, vol. 50, no. 3, pp. 799-807, Feb. 2015.

[5] M. Eberlein and Idan Yahav, "A 28nm CMOS Ultra-Compact Thermal Sensor in Current-Mode Technique," *IEEE Symp. VLSI Circuits,* pp.1-2, June 2016.

[6] B. Sell, et al., "22FFL: A High Performance and Ultra Low Power FinFET Technology for Mobile and RF Applications," *IEEE IEDM*, 2017.

[7] M. Eberlein, "Current-mode digital temperature sensor apparatus", Patent # US9557226 B2, issued Jan. 2017.

[8] M. A. P. Pertijs, et al., "Low-Cost Calibration Techniques for Smart Temperature Sensors," *IEEE Sensors J.*, vol. 10, pp. 1098–1105, June 2010.

 978-1-5090-4940-0/18/$31.00

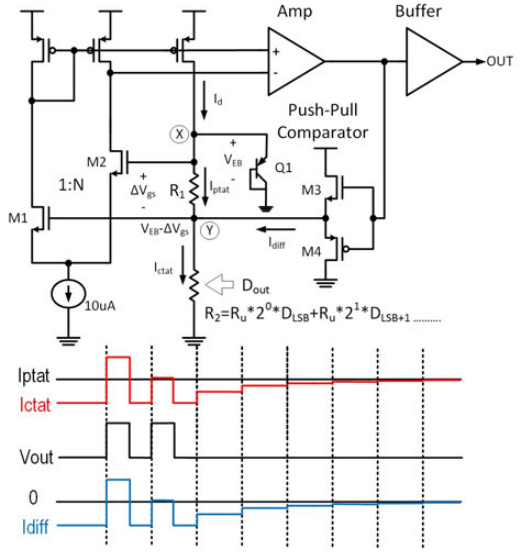


Figure 19.1.1: The hybrid thermal sensor architecture and operational concept.

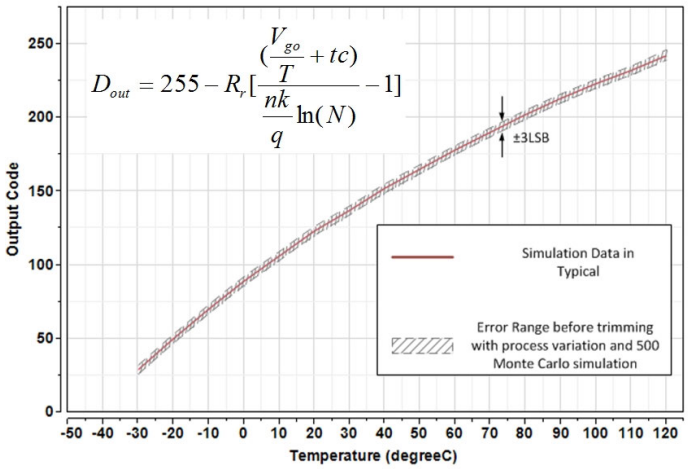
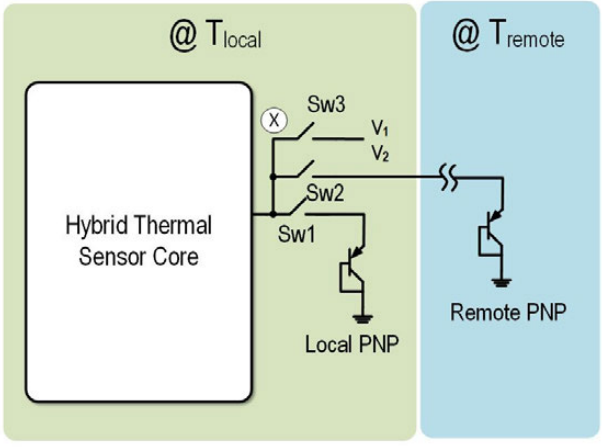


Figure 19.1.2: Temperature-to-digital-code transfer characteristics.



Operational Mode	SW1	SW2	SW3
Normal Operation/ Local Sensing	On	Off	Off
Voltage-based Single Point Trimming	Off	Off	On
Remote Sensing	Off	On	Off

Figure 19.1.3: Sensor architecture with voltage-based single-point soft-trimming and single-element remote sensing.

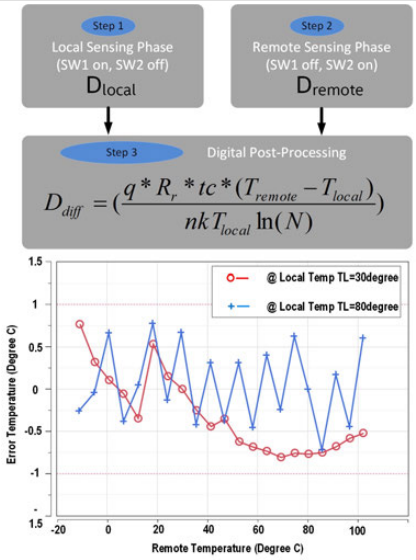


Figure 19.1.4: Remote sensing procedure and inaccuracy.

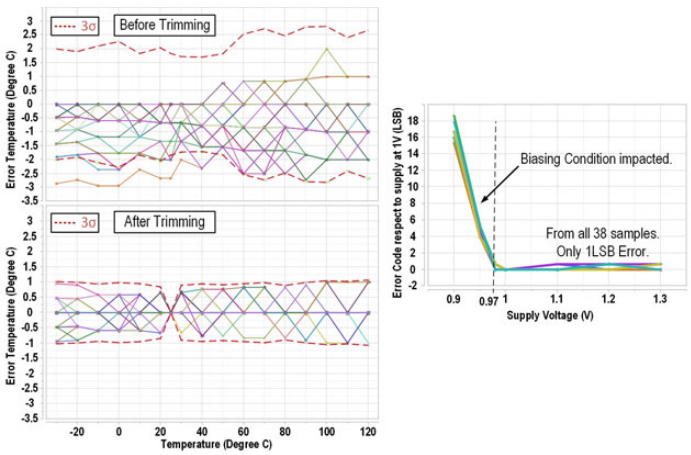


Figure 19.1.5: Error temperature across measured 38 samples before and after trimming and voltage sensitivity.

	This Work		[1]	[2]	[3]	[4]	[5]
Process	22nm		28nm	160nm	16nm	14nm	28nm
Supply Voltage (V)	1		1.8	1.5	1.8	1.35	1.8
Type	PNP & MOS		PNP	PNP	PNP	PNP	NPN
Power Consumption (uW)	50		18.75	6.9	1210	1111.2	28.8
Conversion Time (ms/conv)	0.032		8.192	5	0.27	0.022	0.032
Energy/Conversion (nJ)	1.6		153.6	34.5	326.7	24.45	0.92
Temperature Range (°C)	-30~120		-25~125	-55~125	-50~150	0~100	-20~130
Resolution (°C)	0.58		0.15	0.015	0.4	0.5	0.5
Trimming Type	1-point	No	No	1-point	No	2-point	No
3σ Inaccuracy (°C)	±1.07	±2.81	±1.85	±0.3	±2	±3.3	±1.8
Silicon Area (mm²)	0.0043/0.00021*		0.0095	0.17	0.0128	0.0087	0.0038
Resolution FOM (nJ*K²)	0.54		3.46	0.0078	52.3	6.11	0.46

*Silicon requirement of sensing area when using remote sensing technique.

Figure 19.1.6: Summary of measurement result and comparison with state-of-the-art BJT-based thermal sensors.

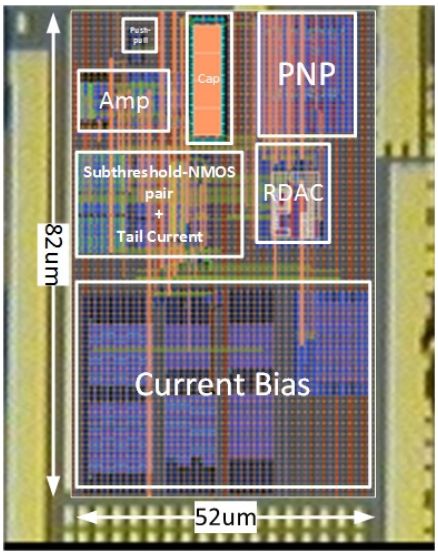


Figure 19.1.7: Hybrid thermal sensor silicon micrograph.

978-1-5090-4940-0/18/$31.00

19.2 A 0.25mm² Resistor-Based Temperature Sensor with an Inaccuracy of 0.12°C (3σ) from -55°C to 125°C and a Resolution FOM of 32fJ·K²

Sining Pan, Kofi A. A. Makinwa

Delft University of Technology, Delft, The Netherlands

Temperature sensors based on Wheatstone bridges, e.g. [1,2], have recently achieved higher resolution and greater energy efficiency than conventional BJT-based sensors [3]. However, this comes at the expense of area, making them less attractive in industrial applications. This paper presents a Wheatstone-bridge sensor that uses a zoom-ADC architecture to reduce area (by 3× over [2]) and achieve state-of-the-art energy-efficiency for an integrated temperature sensor. After a 1st-order fit and a systematic non-linearity correction [2,4], it also achieves state-of-the-art inaccuracy: 0.12°C (3σ) over the full military temperature range (–55°C to 125°C).

An energy-efficient way of reading out a Wheatstone bridge (WhB) is by directly connecting it to the virtual ground established by the 1st integrator of a single-bit continuous-time delta-sigma modulator (CTΔΣM) [1,2]. A single-ended model of this system is shown in Fig. 19.2.1 (left). For maximum sensitivity, the chosen bridge resistors R_p and R_n have positive and negative temperature coefficients, respectively. The modulator's single-bit DAC consists of a resistor R_{DAC} that can be connected either to V_{DD} or GND depending on the bitstream state. The resulting current I_{DAC} compensates for the bridge output current I_{sig} such that the average error current (I_{err}) flowing into the modulator's 1st integrator is zero. As the operating range of the sensor increases, however, the magnitudes of I_{sig} and I_{err} both increase. A large integration capacitor C_{int} (180pF in [2]) is then required to constrain the swing of the 1st integrator.

As shown in Fig. 19.2.1 (right), using a multibit DAC decreases the magnitude of I_{err}, and thus the required size of C_{int}. As a further advantage, the current consumption of the active integrator can also be reduced, thus improving the energy efficiency of the sensor. Due to their relatively large temperature sensitivity and low voltage dependency, a silicided p-poly (s-p-poly) resistor R_p and a non-silicided n-poly resistor R_n are used in the WhB. Their values, R_p = 100kΩ and R_n = 80kΩ, are chosen to restrict $|I_{sig}|$ to < 7μA over PVT. In a compromise between DAC area and that of C_{int}, a 3b DAC made from non-silicided n-poly elements (R_{DAC}= 960kΩ) was chosen.

To further reduce area, the CTΔΣM was realized as a 2nd-order zoom-ADC, since this only requires a single comparator to drive a multibit DAC, as shown in Fig. 19.2.2. The ADC digitizes the ratio $X = I_{sig}/(2I_{DAC})$ (full range from –4 to 4) in two steps. First, a coarse 3b SAR conversion determines the integer part n of X. During this phase, the loop filter is used as a preamplifier by resetting its capacitors before each bit conversion. The fractional part μ is then determined by a single-bit 2nd-order incremental conversion, using reference currents of $(n-1)\cdot 2I_{DAC}$ and $(n+1)\cdot 2I_{DAC}$ to achieve over-ranging and absorb coarse-conversion errors [5].

The 2nd-order CTΔΣM employs a feedforward architecture, with a compensating zero realized by including R_{ff} in the feedback path of the 2nd integrator. Since I_{err} is now quite small, the 1st integrator is based on an energy-efficient current-reuse OTA, rather than the two-stage opamps used in [1,2]. It achieves 80dB gain over PVT, and uses high-V_T transistors to achieve good output swing (~0.9Vpp) [6], as shown in Fig. 19.2.3 (top left). Pole-zero compensation is implemented by inserting $R_{com} \approx 1/g_m$ in series with C_{int}, which improves the stability of the modulator. As in [2], the OTA is chopped at the sampling frequency (fs = 500kHz) to suppress its offset and 1/f noise of the 1st integrator. From simulations, the OTA dissipates 22μW at room temperature (RT), which is ~60% of the sensor's total power. In order to accommodate the swing of the 1st integrator, the 2nd integrator is based on a source-degenerated telescopic OTA, which consumes only 3μW.

Non-linearity is a key challenge for multibit ADCs. For the proposed zoom-ADC, the two major contributors are R_{DAC} mismatch and the non-linearity of the1st-integrator. Although the former can be mitigated by dynamic element matching (DEM), e.g. data-weighted averaging (DWA), dealing with the latter is more challenging. It is caused by the nonlinear variation of the input impedance ($\sim 1/g_m$) of the 1st integrator with I_{DAC}, and leads to step errors of ~0.1°C at coarse code transitions (Fig. 19.2.3, top right). However, the use of over-ranging means that, in principle, the same ADC result can be obtained from either of two adjacent coarse codes, e.g. n and $n+1$. Noting that the resulting errors then have opposite polarity, improved linearity can be achieved by segment averaging, i.e. averaging the result of two conversions centered on adjacent coarse codes (Fig. 19.2.3, bottom right). Although a 2nd-order modulator will remain stable over its full DAC range for DC signals, its quantization noise will increase rapidly at the extremes. To avoid this, the second conversion is only based on an adjacent coarse code if the expected result lies within 95% of the DAC's range. From simulations, the residual non-linearity is then less than 0.02°C, which is well below the measured inaccuracy of the sensor.

The prototype sensor is realized in a 0.18μm CMOS process (Fig. 19.2.7). It consumes 52μA from a 1.8V supply, and occupies 0.25mm². About 45% of the area is occupied by the WhB and the DAC, and about 30% by the capacitors (2×45pF) of the 1st integrator. For flexibility, the DWA and SAR logic were implemented off-chip. Simulations show that an on-chip realization would dissipate only 0.7μW and occupy 0.005mm².

19 samples from one wafer were mounted in ceramic DIL packages and characterized in a temperature-controlled oven. To minimize temperature drift during the measurements, the sensors were mounted in good thermal contact with a large aluminum block. After segment averaging, the sensor output vs. temperature is shown in Fig. 19.2.4 (left). After a 1st-order fit and a systematic non-linearity correction [2,4], the sensor achieves an inaccuracy of 0.12°C (3σ), from –55°C to 125°C, as shown in Fig. 19.2.4 (bottom right). Without segment averaging, the step errors due to OTA non-linearity can be clearly seen (Fig. 19.2.4, top right).

FFTs of the sensor's bit-stream outputs are shown in Fig. 19.2.5 (top) for different types of DEM. Although barrel-shifting DEM is simpler than data-weighted averaging (DWA), it elevates the sensor's noise floor. With DWA, the sensor achieves a thermal-noise-limited resolution of 260μK_{rms} in a 5ms conversion time (Fig. 19.2.5, bottom). This is not affected by segment averaging (2.5ms/segment). Furthermore, the use of a multibit DAC makes the sensor fairly robust to clock jitter: simulation shows that 40ps (rms) jitter corresponds to only a 5% increase in thermal noise power. The sensor's 1/f noise (10Hz corner frequency) is mainly due to the n-poly resistors used in the WhB and DAC [4]. At RT, the sensor achieves a supply sensitivity of 0.02°C/V from 1.6 to 2V.

In Fig. 19.2.6, the performance of the proposed temperature sensor is summarized and compared with other energy-efficient sensors. It is 3× smaller than [2], and achieves higher energy efficiency. In fact, its resolution FOM (32fJ·K²) is even lower than that of a recent MEMS-based sensor [7]. Notably, for a resistor-based sensor, it is also capable of operating over the full military temperature range (–55°C to 125°C). Over this range it achieves a similar relative inaccuracy as [2], which, however, only operates over the industrial temperature range (–40°C to 85°C).

Acknowledgment:
We thank Hui Jiang for his helpful comments and valuable support.

References:

[1] C. Weng, et al., "A CMOS Thermistor-Embedded Continuous-Time Delta-Sigma Temperature Sensor With a Resolution FoM of 0.65 pJ°C²," *IEEE JSSC*, vol. 50, no. 11, pp. 2491-2500, Nov. 2015.
[2] S. Pan, et al., "A CMOS Temperature Sensor with a 49fJ·K² Resolution FoM," *IEEE Symp. VLSI Circuits*, pp. C82-C83, 2017.
[3] K.A.A. Makinwa, "Smart Temperature Sensor Survey," [Online]. Available: http://ei.ewi.tudelft.nl/docs/TSensor_survey.xls
[4] S. Pan, et al., "A Resistor-Based Temperature Sensor with a 0.13pJ·K² Resolution FOM," *IEEE ISSCC*, pp. 158-159, Feb. 2017.
[5] Y. Chae, et al., "A 6.3 μW 20 bit Incremental Zoom-ADC with 6 ppm INL and 1 μV Offset," *IEEE JSSC*, vol. 48, no. 12, pp. 3019-3027, Dec. 2013.
[6] B. Yousefzadeh, et al., "A BJT-Based Temperature-to-Digital Converter With ±60 mK (3σ) Inaccuracy From -55°C to +125 °C in 0.16 μm CMOS," *IEEE JSSC*, vol. 52, no. 4, pp. 1044-1052, April 2017.
[7] M. H. Roshan, et al., "A MEMS-Assisted Temperature Sensor With 20-μK Resolution, Conversion Rate of 200 S/s, and FOM of 0.04 pJK²," *IEEE JSSC*, vol. 52, no. 1, pp. 185-197, Jan. 2017.

 978-1-5090-4940-0/18/$31.00

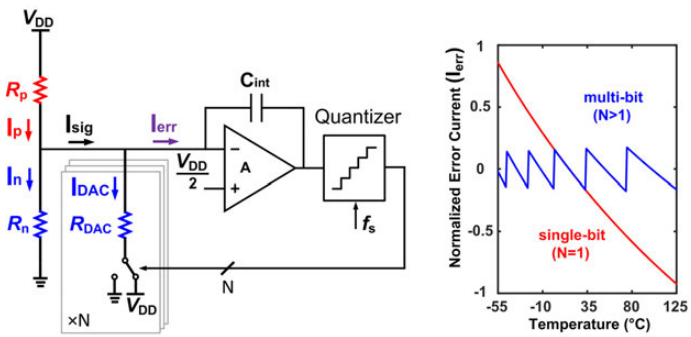


Figure 19.2.1: ΔΣ readout of a Wheatstone-bridge temperature sensor (left), error current I_{err} over temperature (right).

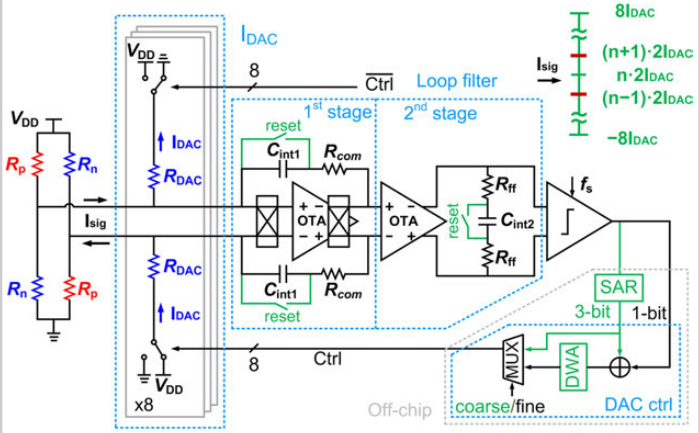


Figure 19.2.2: System block diagram of the proposed sensor.

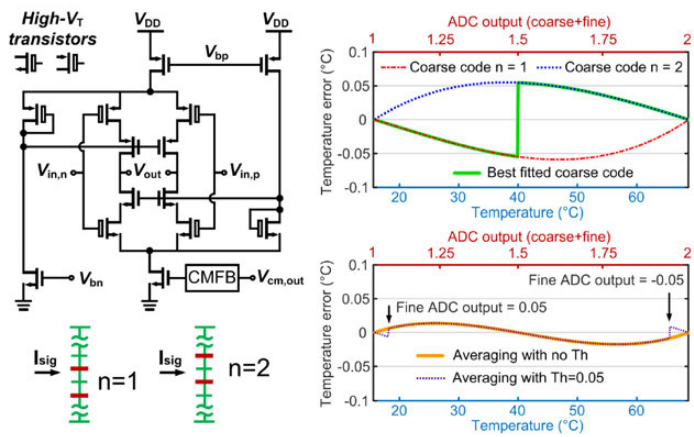


Figure 19.2.3: Current-reuse OTA (top left), zoom-ADC levels (bottom left), simulated error due to OTA non-linearity (top right), and after segment averaging (bottom right).

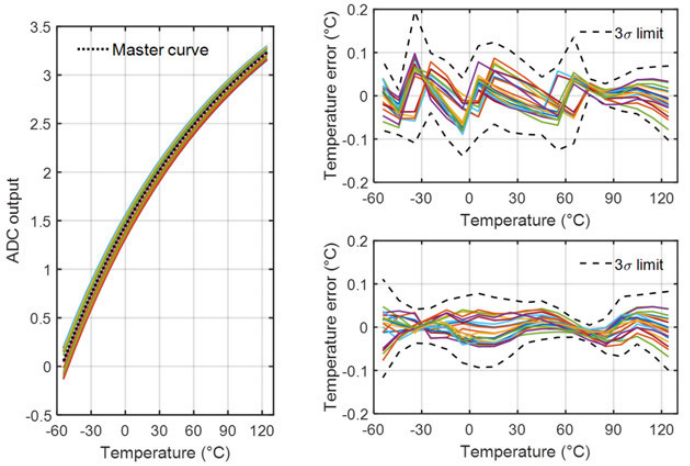


Figure 19.2.4: Sensor characteristic (left), temperature error after a 1st-order fit and fixed non-linearity correction, without and with segment averaging (top & bottom right).

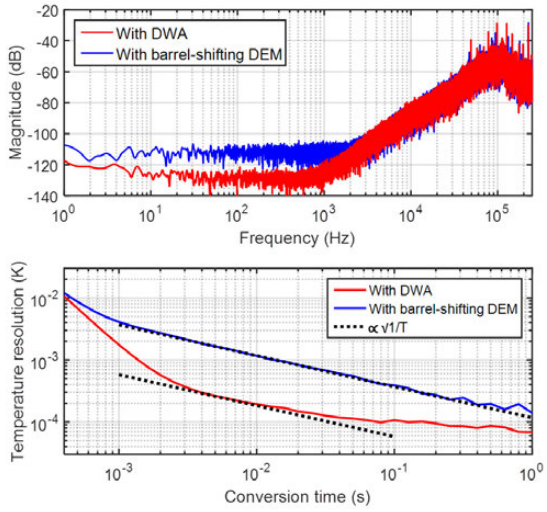


Figure 19.2.5: PSD of the sensor output bitstream (top), resolution vs. time plot (bottom).

	This work	[7]	[2]	[4]	[1]	[6]
Sensor type	**Resistor WhB**	Dual-MEMS Resonator	Resistor WB	Resistor WhB	Resistor WhB	BJT
CMOS Technology	**0.18µm**	0.18µm	0.18µm	0.18µm	0.18µm	0.16µm
Area [mm2]	**0.25**	0.54	0.72	0.72	0.43	0.16
Temperature range	**-55°C to 125°C**	-40°C to 85°C	-40°C to 85°C	-40°C to 85°C	-40°C to 125°C	-70°C to 125°C
3σ inaccuracy [°C] (Trimming points)	**0.12 (2[a])**	--	0.07 (2[a])	0.1 (2[a])	0.4 (2[b])	0.06 (1)
Power consumption [µW]	**94**	13000	160	180	65	7
Conversion time [ms]	**5**	5	5	10	0.1	5
Resolution [mK]	**0.26**	0.02	0.41	0.16	10	15
Resolution FOM [fJ·K²][c]	**32**	40[d]	130	49	650	7300

[a] 1st-order fit. [b] 1-point trim with 1st-order fit, min-max.
[c] Energy / Conversion × (Resolution)². [d] MEMS die + CMOS readout IC.

Figure 19.2.6: Performance summary and comparison with previous work.

19

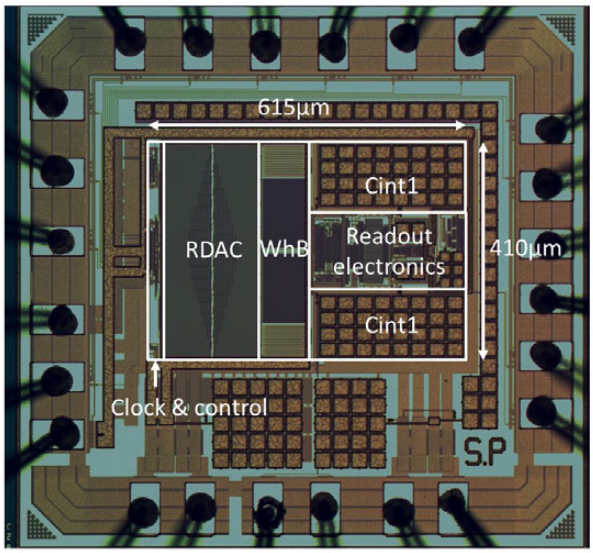


Figure 19.2.7: Die micrograph of the fabricated temperature sensor.

978-1-5090-4940-0/18/$31.00

19.3 A 0.53pJ·K² 7000µm² Resistor-Based Temperature Sensor with an Inaccuracy of ±0.35°C (3σ) in 65nm CMOS

Woojun Choi[1], Yong-Tae Lee[1], Seonhong Kim[2], Sanghoon Lee[2], Jieun Jang[2], Junhyun Chun[2], Kofi A. A. Makinwa[3], Youngcheol Chae[1]

[1]Yonsei University, Seoul, Korea; [2]SK hynix, Icheon, Korea
[3]Delft University of Technology, Delft, The Netherlands

In microprocessors and DRAMs, on-chip temperature sensors are essential components, ensuring reliability by monitoring thermal gradients and hot spots. Such sensors must be as small as possible, since multiple sensors are required for dense thermal monitoring. However, conventional BJT-based temperature sensors are not compatible with the sub-1V supply of advanced processes. Subthreshold MOSFETs can operate from lower supplies, but at high temperatures their performance is limited by leakage [1,2]. Thermal diffusivity (TD) sensors achieve sub-1V operation and small area with moderate accuracy, but require milliwatts of power [3]. Recently, resistor-based sensors based on RC Wien-Bridge (WB) filters have realized high resolution and energy efficiency [4,5]. Fundamentally, they are robust to process and supply-voltage scaling. However, their readout circuitry has been based on continuous-time (CT) ΔΣ ADCs or frequency-locked loops (FLLs), which require precision analog circuits and occupy considerable area (>0.7mm²).

This paper presents a highly digital resistor-based temperature sensor in 65nm CMOS, which achieves a 3σ inaccuracy of ±0.35°C from -40 to 85°C and 2.8mK resolution at a 1kS/s sampling rate. The sensor can operate from 0.85V supplies, while consuming only 68µW. This corresponds to a resolution FOM of 0.53pJ·K², which is more than 15× less than a previous WB-based FLL sensor [5]. Furthermore, the sensor occupies only 7000µm², which is 13× less than [5] and comparable to state-of-the-art BJT-, MOS-, and TD-based sensors [1-3]. These advances are achieved by the use of an RC polyphase filter (PPF) as a sensing element, which is then read out by a highly digital frequency-locked loop (FLL).

The PPF consists of two silicided N-poly resistors (R=35kΩ) and two MIM capacitors (C=0.5pF). Silicided poly resistors were chosen because of their large temperature coefficient, low voltage dependency, and low 1/*f* noise [4]. Compared to a WB, a PPF requires less passive components. Furthermore, when driven by anti-phase clocks (P, $\overline{P}$), its output voltage V_{PPF} has a 5× larger swing than that of a similarly driven WB, which proportionally reduces the error contribution of the succeeding readout circuitry [6].

As shown in Fig. 19.3.1, the phase shift of an RC filter can be read out by embedding it in an FLL [4,5]. The loop settles when the input of the integrator is zero, which corresponds to a fixed phase-shift in the RC filter and hence to a fixed VCO frequency. In this work, the analog-intensive phase-demodulation scheme of [4,5] is replaced by a simple comparator that digitizes the zero-crossing of V_{PPF}, whose timing is also a function of the filter phase-shift. A phase/frequency detector (PFD) then compares the phase of the comparator output V_O with that of a quadrature feedback signal (Q) and then drives an oscillator, via a digital PI controller, such that its output frequency (F_{PPF}) corresponds to a 90° phase-shift. As a result, F_{PPF} will be proportional to $1/(R_{PPF}C)$.

Figure 19.3.2 shows the block diagram and operation principle of the proposed PPF-based FLL. By comparing the comparator output V_O with the rising edge of the Q signal, the PFD provides an up/down signal to increase/decrease the loop filter output current, which drives a current-controlled oscillator (CCO). The gain of the loop filter locks the CCO frequency F_{PPF} to the point when the phase error (φ_{DIFF}) is zero. At steady state, the current from the integral path I_{INT} is stabilized (Fig. 19.3.2, bottom). Since the VCO phase noise is shaped with the loop gain of the FLL, an analog front-end often limits the phase noise performance. In the analog-intensive approach [5], the current buffer (CB) continuously provides the output phase-demodulated current into the integration capacitor, so a considerable amount of power is required to mitigate the noise. However, since the FLL comparison instants can be predicted within a half duty cycle, the proposed front-end noise can be mitigated by increasing the power of the comparator only at the comparison moment. In the loop filter, a proportional path is added to reduce the lock-time of the loop from power-on-reset condition (<100 cycles), which is important for dense thermal monitoring applications. It also works as an added g_m, designed with a push-type charge pump (CP), and increases the loop bandwidth to 20kHz, thus reducing VCO phase noise more effectively. The proportional path is implemented with current sources that are directly driven by the PFD and the current in the proportional path I_{PROP} is defined through the voltage charged on C_{INT} (4pF) of the integral path.

Figure 19.3.3 shows the circuit-level implementation of the proposed readout circuit. To facilitate scaling in advanced processes, an inverter-based comparator is used. Two inverters serve as a preamplifier, which then drives a cross-coupled latch. Since the output of the PPF is only sampled on the rising edge of Q, the comparator can be disabled half the time to reduce the power consumption. Its power consumption is only 11µW and an input-referred noise of 18.4µV is estimated. The proportional path (CP_{PROP}) is implemented with PMOS current sources to be operated in a push-type, which also avoids a possible source of mismatch. The CP of the integral path (CP_{INT}) generates the bias voltage V_B, which is converted to I_{PROP} and I_{INT} (5 and 10µA at steady state) through PMOS transistors (M_1-M_3). I_{PROP} and I_{INT} are summed at the supply node of the CCO. Depending on the PFD state, i.e. up, reset, and down, the I_{PROP} is weighted by 2, 1 and 0 respectively, which is implemented by two switches controlled by the PFD outputs: $\overline{UP}$ and DN. The CCO is implemented using a 9-stage current-starved ring oscillator, whose gain is 1MHz/µA. Its delay cell consists of two inverters coupled with each other using transmission gates, attenuating common-mode signals by ensuring pseudo-differential operation. The CCO achieves the target output frequency range of 38 to 48MHz. The output buffer includes a level-shifter for rail-to-rail operation and an inverter-based latch for 50% duty cycle.

The prototype temperature sensor was fabricated in TSMC 65nm CMOS. The sensor occupies an area of 7000µm² (Fig. 19.3.7). It consumes 68µA from a 1V supply. The FLL output frequency changes from 38 to 48MHz over the temperature range from -40 to 85°C (Fig. 19.3.4, left). It exhibits an average temperature coefficient of 0.19%/°C. The temperature error is measured for 16 samples in ceramic DIL packages. As shown in Fig. 19.3.4 (right), after one- and two-point trimming with the removal of the systematic non-linearity following a 1st-order fit, a 3σ inaccuracy of ±3.65°C and ±0.35°C is achieved from -40 to 85°C, respectively. At room temperature, the sensor supply sensitivity of 0.5°C/V is measured from 0.85 to 1.05V. The measured phase noise at 100kHz offset is -124dBc/Hz, and the RMS jitter integrated from 1Hz to 100kHz is 12ps (Fig. 19.3.5). The accumulated jitter increases up to 10^4 cycles, showing √N behavior due to the thermal noise. Also, an accumulated jitter of 5.2ns (rms) is achieved in a 1ms time window, corresponding to 2.8mK temperature resolution. As a result, the sensor can track millisecond thermal transients with high energy efficiency (FOM of 0.53pJ·K²).

Figure 19.3.6 shows the performance summary and comparison with state-of-the-art. The proposed sensor is 13× smaller than a conventional FLL readout [5] and 100× smaller than a ΔΣ ADC readout [4]. Due to the supply voltage scaling and improved jitter performance, the resolution FOM is also highly improved by 15× compared to [5]. Even compared to a compact MOS-based sensor [2], this work is 6× more energy-efficient with similar size. It can be seen that, compared to BJT- and TD-based designs [1,3], this sensor consumes significantly less energy. These results demonstrate that the proposed PPF-based FLL can be used to realize reliable on-chip temperature sensors for dense thermal monitoring, in nanometer CMOS processes.

Acknowledgements:
This paper was the result of a research project supported by SK Hynix Inc. and it was also supported by NRF (National Research Foundation of Korea) Grant funded by the Korean Government (NRF-2016-Global Ph.D. Fellowship Program).

References:

[1] T. Oshita, et al., "Compact BJT-Based Thermal Sensor for Processor Applications in a 14 nm tri-Gate CMOS Process," *IEEE JSSC*, vol. 50, no. 3, pp. 799-807, Mar. 2015.

[2] K. Yang, et al., "A 0.6nJ –0.22/+0.19°C Inaccuracy Temperature Sensor Using Exponential Subthreshold Oscillation Dependence," *ISSCC*, pp.160-161, Feb. 2017.

[3] U. Sönmez, et al., "1650µm² Thermal-Diffusivity Sensors with Inaccuracies Down to ±0.75°C in 40nm CMOS," *ISSCC*, pp.206-207, Feb. 2016.

[4] S. Pan, et al., "A Resistor-Based Temperature Sensor with a 0.13pJ·K² Resolution FOM," *ISSCC*, pp.158-159, Feb. 2017.

[5] P. Park, et al., "A Thermistor-Based Temperature Sensor for a Real-Time Clock with ±2 ppm Frequency Stability," *IEEE JSSC*, vol. 50, no. 7, pp. 1571-1580, July 2015.

[6] J. Lee, et al., "A 1.4V 10.5MHz Swing-Boosted Differential Relaxation Oscillator with 162.1dBc/Hz FOM and 9.86ps_{rms} Period Jitter in 0.18µm CMOS," *ISSCC*, pp.106-108, Feb. 2016.

 978-1-5090-4940-0/18/$31.00

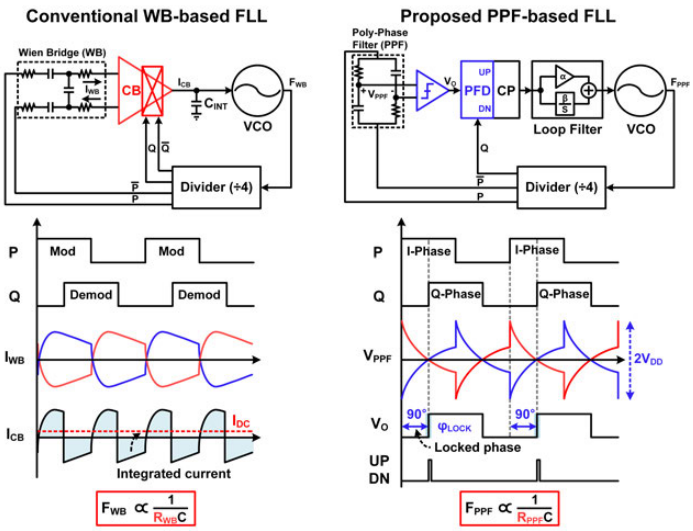


Figure 19.3.1: Architecture of the WB-based FLL and the proposed PPF-based FLL.

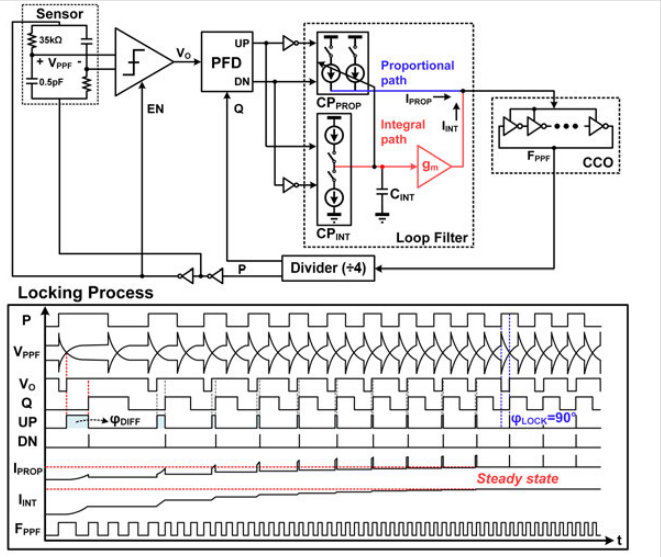


Figure 19.3.2: Block diagram and operation principle of the PPF-based FLL.

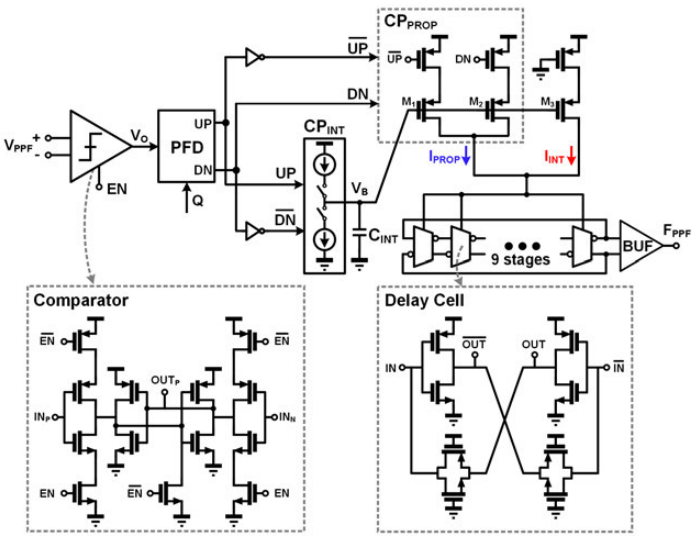


Figure 19.3.3: Circuit implementation of the readout circuit.

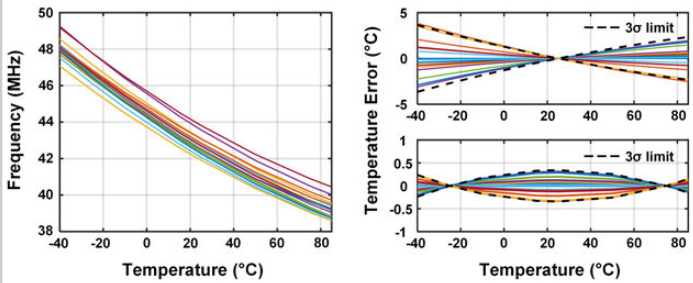


Figure 19.3.4: Measured sensor output frequency (left) and Temperature error after 1-point (top right) and 2-point trimming (bottom right).

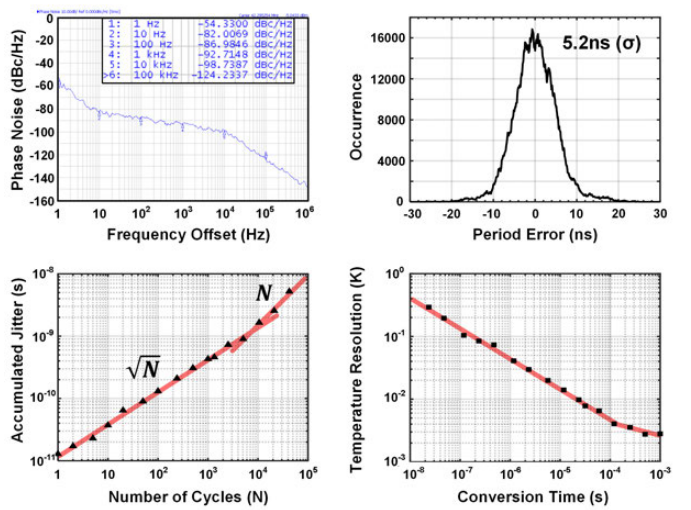


Figure 19.3.5: Measurement Results: Phase noise, Accumulated jitter on a 1ms time window, Accumulated jitter vs. number of cycles, and Temperature resolution vs. conversion time.

Publication	This Work	JSSC15 Park [5]	ISSCC17 Pan [4]	ISSCC17 Yang [2]	JSSC15 Oshita [1]	ISSCC16 Sönmez [3]
Sensor Type (Configuration)	**Resistor (PPF)**	Resistor (WB)	Resistor (WB)	MOS	BJT	TD
Readout Type	**FLL**	FLL	ΔΣ	OSC	ΔΣ	ΔΣ
Technology	**65nm**	180nm	180nm	180nm	14nm	40nm
Area [μm²]	**7000**	90000	720000	8865	8700	1650
Power [μW]	**68**	31	160	0.075	1100	2500
Supply voltage [V]	**0.85-1.05**	3.3	1.6-2	0.8-1.8	1.35	0.9-1.2
Supply sensitivity [°C/V]	**0.5**	0.4	0.17	0.13	-	2.8
Temperature range [°C]	**-40 to 85**	-40 to 85	-40 to 85	-20 to 100	0 to 100	-40 to 125
Inaccuracy [°C]	**±0.35**** (3σ)	±0.12*** (p-p)	±0.2* (3σ)	-0.22/0.19** (3σ)	±0.7** (3σ)	±0.75* (3σ)
Conversion time [ms]	**1**	32	5	8	0.02	1
Energy/Conversion [nJ]	**68**	992	800	0.6	22	2500
Resolution [°C]	**0.0028**	0.0028	0.00041	0.073	0.5	0.36
Resolution FOM [pJ·K²]	**0.53**	8	0.13	3.2	5500	324000

* 1-point trimming, ** 2-point trimming, *** 3-point trimming

Resolution FOM = Energy/conversion × (Resolution)²

Figure 19.3.6: Performance summary and comparison with the state of the art.

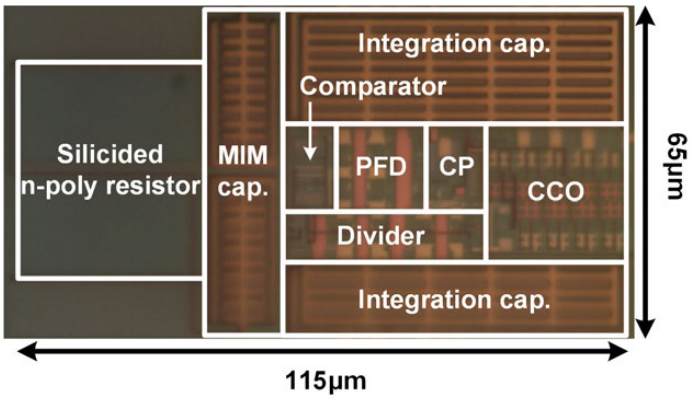


Figure 19.3.7: Die micrograph.

978-1-5090-4940-0/18/$31.00

19.4 A ±4A High-Side Current Sensor with 25V Input CM Range and 0.9% Gain Error from -40°C to 85°C Using an Analog Temperature Compensation Technique

Long Xu, Johan H. Huijsing, Kofi A. A. Makinwa

Delft University of Technology, Delft, The Netherlands

This paper presents a fully integrated ±4A current sensor that supports a 25V input common-mode voltage range (CMVR) while operating from a single 1.5V supply. It consists of an on-chip metal shunt, a beyond-the-rails ADC [1] and a temperature-dependent voltage reference. The beyond-the-rails ADC facilitates high-side current sensing without the need for external resistive dividers or level shifters, thus reducing power consumption and system complexity. To compensate for the shunt's temperature dependence, the ADC employs a proportional-to-absolute-temperature (PTAT) reference voltage. Compared to digital temperature compensation schemes [2,3], this analog scheme eliminates the need for a temperature sensor, a band-gap voltage reference and calibration logic. As a result, the current sensor draws only 10.9µA and is 10× more energy efficient than [2]. Over a ±4A range, and after a one-point trim, the sensor exhibits a 0.9% (max) gain error from -40°C to 85°C and a 0.05% gain error at room temperature. The former is comparable with that of other fully-integrated current sensors [2-4], while the latter represents the state-of-the-art.

Figure 19.4.1 shows a simplified block diagram of the current sensor. The load current is measured by digitizing the voltage drop V_S across a metal shunt resistor R_S inserted between the battery and the load. To safely handle ±4A currents, the 10mΩ shunt consists of four metal layers (M2-M5) and occupies 450µm×880µm (Fig. 19.4.1). Being made of aluminum, its resistance has a large temperature dependence: $R_S = R_0 \times (1+\alpha_{shunt} \times (T-T_0))$, $T_0 = 25°C$, where $\alpha_{shunt} \approx 0.34\%/°C$, which means that Joule heating or ambient temperature variations will cause significant gain error.

In previous work, the shunt's temperature dependence has been calibrated in the digital domain, by sensing the shunt temperature *T* and then using this information to correct the ADC's output with the help of a calibration polynomial [2,3]. Noting that the metal shunt's temperature dependence is almost perfectly PTAT ($R_S \approx k_R \cdot T_A$, T_A is absolute temperature) over the industrial temperature range, we propose an analog compensation scheme in which the ADC is driven by a PTAT voltage reference $V_{Ref} = k_V \cdot T_A$ (Fig. 19.4.1). Consequently, the shunt's 1st-order temperature dependency is corrected in a ratiometric manner without calibration. Although this approach does not correct for the non-linear components of the shunt's temperature dependence, simulations show that the resulting gain error will be less than ±1% from -40°C to 85°C. Furthermore, the spread of the nominal values of the shunt resistance R_0 (±10%) and the magnitude of V_{Ref} can both be corrected by a single trim at room temperature.

The schematic of the V_{Ref} generator is shown in Fig. 19.4.1. It consists of a bias circuit and a bipolar core, both based on pairs of NPN transistors with an emitter-area ratio *p = 7*. In the bias circuit, one pair is biased by two identical current sources, thus generating a base-emitter voltage difference $\Delta V_{BE} = (k/q) \times \ln(p) \times T_A$. This is then forced across a poly-resistor R_b, resulting in a PTAT current $\Delta V_{BE}/R_b$. Leveraging the vertical NPNs available in the chosen process means that this can be done without the extra low-offset amplifier required by PNP-based bias circuits, e.g. as used in [2]. The PTAT biasing current is then mirrored (1:4) to the bipolar core and used to bias the second pair of NPNs, thus generating the accurate ΔV_{BE} that is used as V_{Ref}. The two current sources in the bipolar core are chopped to suppress their *1/f* noise. To avoid intermodulation issues, the chopping frequency is the same as the ADC's sampling frequency. The NPN transistors are located underneath the shunt to ensure good thermal coupling between the metal shunt and the voltage reference. This is further improved by using thermal vias to connect the shunt to a sheet of M1 around the NPNs [2]. Compared to the analog compensation scheme described in [5], which uses a bandgap voltage reference followed by a reference buffer with a temperature-dependent gain, the proposed solution is much simpler and more power efficient.

Figure 19.4.2 shows the schematic of the beyond-the-rails ADC [1]. It is based on a 2nd-order feedforward SC ΔΣ ADC built around two current-reuse OTAs. During $\phi 1$, the input signal, V_S, and the OTA offset are sampled on the 2.5pF input capacitors, C_{S1}. During $\phi 2$, the HV chopper CH_{HV} reverses the polarity of V_S and thus transfers a charge packet proportional to $2 \cdot C_{S1} \cdot V_S$ to the integration capacitors, C_{INT}. In a similar manner, the PTAT voltage reference ΔV_{BE} is sampled onto feedback capacitors C_{S2} (2.5pF) via an LV chopper CH_{LV} with the polarity determined by the modulator's bitstream. This cross-coupled sampling scheme ensures that the only components exposed to the input CM voltage are the input capacitors and the HV chopper. Together with switches S_{1-2}, it also realizes a correlated-double-sampling (CDS) scheme that suppresses the offset and 1/*f* noise of the 1st OTA. The switch timing is designed to ensure that the residual offset is mainly due to the charge-injection mismatch of switches S_{1-2}, and so can be further reduced by low-frequency chopping (CHL). In [1], this was implemented by an additional capacitively-coupled HV input chopper, which then had to be periodically toggled to keep its coupling capacitors charged. In this design, the same functionality is achieved by swapping the clock signals $\phi 1$, $\phi 2$ applied to the input chopper CH_{HV} (Fig. 19.4.2), thus allowing CHL to be completely disabled if necessary. For good matching, both C_{S1} and C_{S2} are implemented as fringe capacitors with a 70V breakdown voltage.

The schematic of CH_{HV} is shown in Fig. 19.4.2. Its clock signals ϕ_1, ϕ_2 are capacitively-coupled to the gates of four sampling switches M_{1-4} via two HV capacitors $C_{1,2}$. A minimum selector $M_{S1,2}$ connected between the input terminals V_{ip} and V_{in} selects the lowest input voltage. Its output is tied to the reference of the clock level shifter comprising coupling capacitors $C_{1,2}$ and a latch $M_{5,6}$. As a result, the coupled clocks are always superimposed on V_{min} (the lower of V_{ip} and V_{in}), which minimizes the leakage of M_{1-4} in the presence of bidirectional input voltages [1].

The current sensor was implemented in a 0.18µm HV BCD CMOS technology and occupies 1.4mm² (Fig. 19.4.7). It draws 10.9µA from a 1.5V supply at room temperature. The reference generator, the ADC and the digital clock generator consume 4µA, 5.2µA and 1.7µA respectively. Figure 19.4.3 shows the output spectrum of the ADC for different input currents. At a sampling frequency of 250kHz, the ADC achieves a resolution of $1.5\mu V_{rms}$ in a conversion time of 2ms, which translates into a current-sensing resolution of $150\mu A_{rms}$.

10 sensors were characterized in a current range of ±4A from -40°C to 85°C (Fig. 19.4.4). After trimming its digital output (at +3A and ~25°C), the sensor gain error is only 0.05% at room temperature, increasing to 0.9% over the full temperature range. Over a 25V input CMVR, the ADC maximum offset is 6.4µV (640 µA), dropping below 400nV (40µA) when CHL is enabled (Fig. 19.4.5). This varies by less than 700nV over the full CMVR, corresponding to a CMRR of 151dB, which is improved to 158dB after CHL. The sensor input CMVR is limited by the ESD diodes at the input terminals to -0.7V to 25V.

The performance of the sensor is summarized in Fig. 19.4.6. Its energy efficiency, like that of a temperature sensor, can be expressed in terms of a resolution FOM [6]. Compared to other fully integrated current sensors in the table, this design achieves 10× better energy efficiency, the best accuracy at room temperature, and comparable accuracy over the industrial temperature range.

Acknowledgement:
The authors would like to thank Zuyao Chang for chip bonding and Saleh Heidary Shalmany for layout review.

References:
[1] L. Xu, et al., "A 110dB SNR ADC with ±30V Input Common-Mode Range and 8µV Offset for Current Sensing Applications," *ISSCC*, pp. 374-375, Feb. 2015.
[2] S. H. Shalmany, et al., "A ±36A Integrated Current-sensing System with 0.3% Gain Error and 400µA Offset from –55°C to +85°C," *IEEE JSSC*, vol. 52, no. 4, pp. 1034-1043, Apr. 2017.
[3] Linear Technology, LTC2947 Data Sheet. Accessed on Aug. 21, 2017. Available: http://cds.linear.com/docs/en/datasheet/2947fa.pdf
[4] Texas Instruments, INA260 Data Sheet. Accessed on Aug. 21, 2017. Available: http://www.ti.com/lit/ds/symlink/ina260.pdf
[5] A. Nagari, et al., "An 8Ω 2.5W 1%-THD 104dB(A)-Dynamic-Range Class-D Audio Amplifier With Ultra-Low EMI System and Current Sensing for Speaker Protection", *IEEE JSSC*, vol. 47, no. 12, pp. 3068-3080, Dec. 2012.
[6] K. A. A. Makinwa, "Smart Temperature Sensors in Standard CMOS," Procedia Engineering, vol. 5, pp. 930–939, Sept. 2010.

 978-1-5090-4940-0/18/$31.00

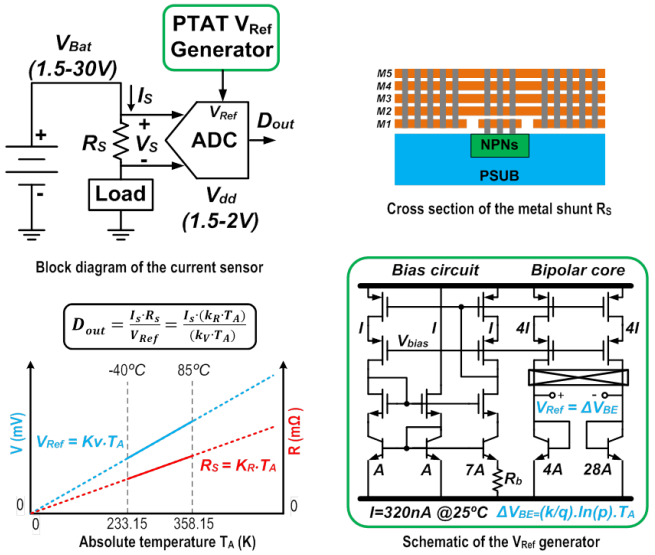


Figure 19.4.1: Block diagram of the current sensor.

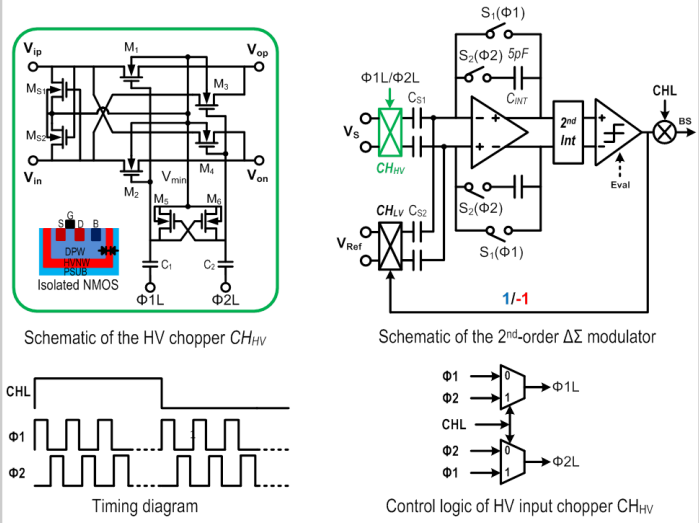


Figure 19.4.2: Block diagram of the beyond-the-rails ADC.

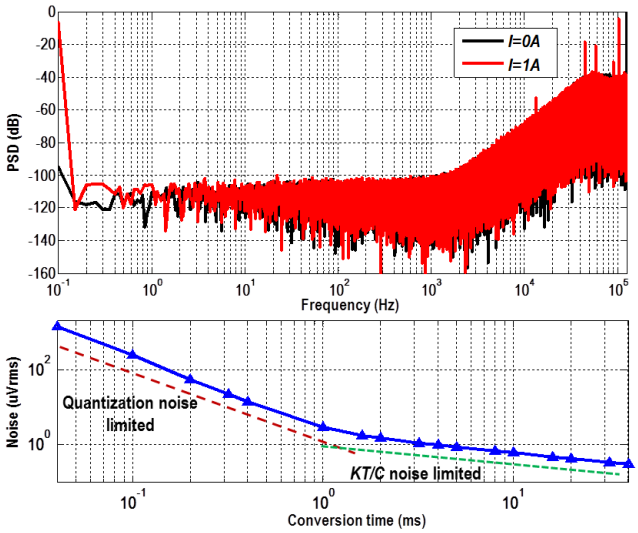


Figure 19.4.3: Measured output spectrum of the beyond-the-rails ADC (CHL is off).

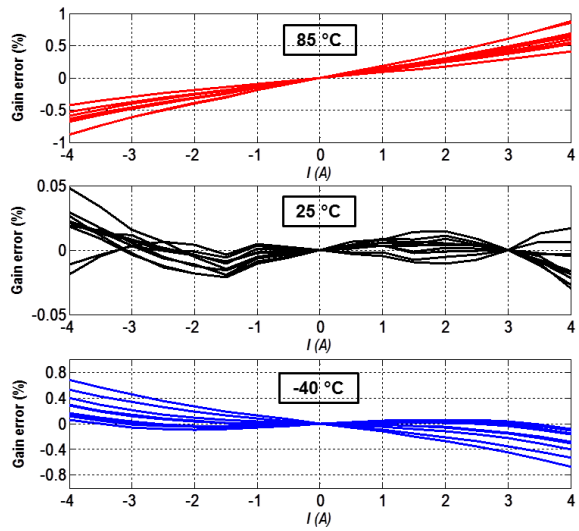


Figure 19.4.4: Measured current-sensing gain error at different ambient temperatures (10 samples).

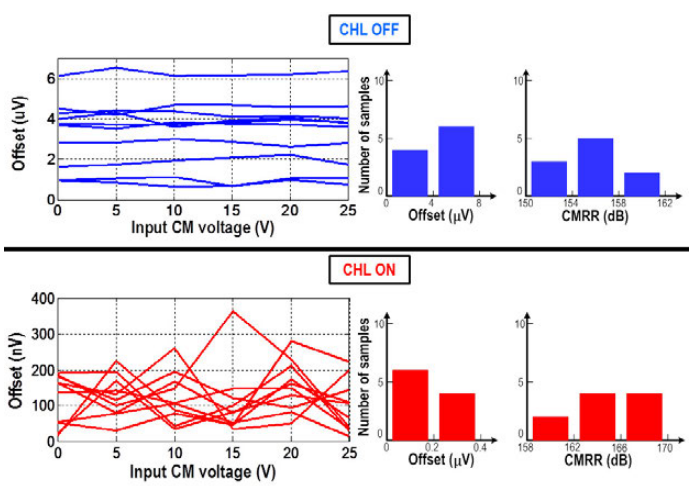


Figure 19.4.5: Measured CMRR and offset (10 samples) over input CMVR with CHL off (top) and on (bottom).

	This work	JSSC 17 [2]	LT2947 [3]	INA260 [4]
I-range	**±4A**	±5A	±30A	±10A
Temperature range	**-40-85°C**	-55-85°C	-40-85°C	-40-125°C
Shunt	**10mΩ**	10mΩ	300µΩ	2mΩ***
Input CM range	**0-25V**	0-0.75V	0-15V	0-36V
Gain error (25°C)	**0.05%**	0.1%	0.75%	0.15%
Gain error (-40-85°C)	**0.9%**	0.3%	1%	0.5%
Offset	**40µA**	4µA	9mA	5mA
Resolution	**150µA**	200µA	3mA	1.25mA
ENOB	**13bit**	13bit	11.5bit	11bit
Conversion time	**2ms**	10ms	100ms	8.2ms
Supply voltage	**1.5-2V**	1.3-1.7V	4.5-15V	2.7-5.5V
Supply current	**10.9µA**	13µA	9mA**	310µA
Polynomial Calibration	**No**	Yes	Yes	No
FOM*	**0.74fJ·A²**	**7.8fJ·A²**	--	--

* FOM = (Energy / Conversion) × Resolution²
** Includes the power of current-sense ADC, voltage-sense ADC, temp. sensor and digital circuitry
*** Uses a custom low-TC shunt

Figure 19.4.6: Performance summary and comparison table.

19

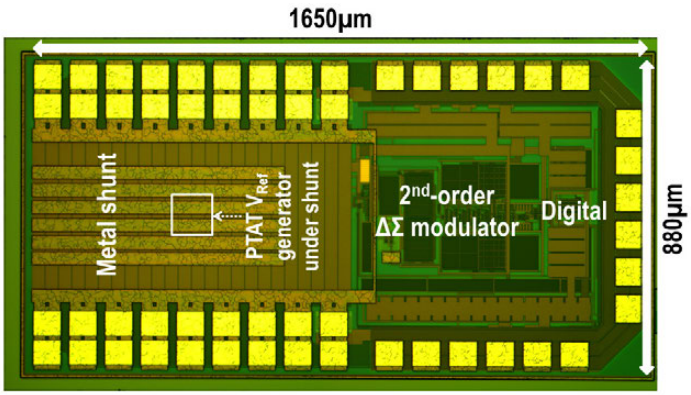


Figure 19.4.7: Die micrograph.

978-1-5090-4940-0/18/$31.00

19.5 A Current-Measurement Front-End with 160dB Dynamic Range and 7ppm INL

Chung-Lun Hsu, Drew A. Hall

University of California, San Diego, La Jolla, CA

Accurate current measurement is crucial in many biosensing applications, such as the detection of neurotransmitters [1] and the monitoring of intercellular molecular dynamics. This need has become even more critical recently with single-molecule biosensors where sub-pA signal currents are superimposed on a slowly varying nA-to-μA background current, as is the case with nanopores [2]. As such, the readout circuitry requires wide dynamic range (>120dB) and high linearity (>14b) albeit often with low bandwidth (a few Hz to kHz). This paper presents a current measurement front-end using a modified asynchronous ΔΣ modulator architecture that achieves 7ppm INL and 160dB dynamic range (100fA to 10μA) for a state-of-the-art 197dB FoM due to: 1) a continuous-time, oscillator-based Hourglass ADC that asynchronously folds the input signal within the supply, 2) noise shaping to suppress the quantization noise, and 3) a digital linearity correction technique that relaxes the amplifier bandwidth requirement thus reducing power.

Figure 19.5.1 shows a block diagram of the wide-dynamic-range (DR) current-mode analog front-end (AFE) that consists of two main blocks: 1) a 9b predictive, current-steering DAC and 2) an 8b oversampling, asynchronous "Hourglass" ADC. Unlike conventional ΔΣ modulators, the Hourglass ADC can tolerate the entire full-scale input current (10μA), but it does so with reduced linearity as described later. To constrain the input range ($i_{fine} \leq FullScale/2^8$), a digital predictor [3], a first-order digital differentiator with one oversampling cycle delay controls the DAC to generate an approximation of the input signal, i_{coarse}. This approximation is subtracted at the input thus closing the loop. The DAC is implemented using a binary-weighted, tri-state topology to minimize the noise, area, and capacitance at the input node [4]. The DAC mismatch is randomized using a tree-structure, segmented dynamic element matching (DEM) technique [5]. The residual current, i_{fine}, is quantized by the Hourglass ADC that is designed to handle 2× the DAC unit current to tolerate prediction errors and remaining mismatch. The linearity of the Hourglass ADC is further improved from <4b to >8b by a one-time offline calibration routine. The 17b digital code, D_{out}, is obtained by combining the digital outputs of the predictor and the Hourglass ADC.

The core of the Hourglass ADC is an open-loop asynchronous ΔΣ consisting of a capacitive-feedback transimpedance amplifier (C-TIA) in conjunction with an Hourglass switch driven by the outputs of two continuous-time comparators (Fig. 19.5.2). The C-TIA continuously integrates the input current and folds the output voltage within a predefined window, $\pm V_R$, by flipping the polarity of the input signal, i_{fine}, using the Hourglass switch, resulting in a current-to-frequency conversion (*I*-to-*F*). In contrast to a conventional periodically reset C-TIA, the asynchronous folding prevents the C-TIA from saturating by alternating between charging and discharging the feedback capacitors, C_F. Using the input current to charge and discharge C_F removes the need for an explicit DAC. Because the quantization error is retained by not resetting C_F, this structure provides first-order noise shaping. Unlike an asynchronous ΔΣ, which has an asymmetric triangular waveform with a frequency inversely proportional to input amplitude, the C-TIA output is a symmetric triangular waveform with a fundamental frequency ($f_{dir} = i_s/4V_RC_F$) that is linearly proportional to the input amplitude. Due to the high OSR and DAC, the harmonic tones (equivalent to idle tones in a conventional ΔΣ) are guaranteed to be out-of-band and are removed by the decimation filter. A counter accumulates the number of comparator pulses, *cp* and *cn*. Like most oscillator-based quantizers, a digital representation of the signal is obtained by sampling the output of the counter and digitally differentiating at the oversampling frequency, f_{OSR}. This Hourglass structure enables wide dynamic range while simultaneously providing the necessary low input impedance for current measurements.

The linearity of the Hourglass ADC can be understood by examining the output of the C-TIA (Fig. 19.5.3). An ideal triangle wave has an infinite number of odd harmonics, but due to the filtering from the finite bandwidth of the amplifier in the C-TIA, the output waveform is distorted. As the input current is increased, f_{dir} linearly increases resulting in poorer linearity for a fixed-bandwidth amplifier. By bounding the input current with the DAC, the number of harmonics, and thus the linearity of the Hourglass ADC, can be ensured. For 8b linearity, the bandwidth of the amplifier must be at least 52× larger than the maximum f_{dir}. Rather than implementing such a wide-bandwidth (>75MHz), power-hungry amplifier, the linearity is corrected digitally using an amplifier with a bandwidth only 3.2× larger than the maximum f_{dir}. Since the distortion can be precisely expressed once the finite loop gain and bandwidth of the amplifier are known, the calibration routine consists of using the DAC to sweep a subset of the *I*-to-*F* transfer function and fitting with a 5th-order polynomial. This approach results in 16× lower power compared to simply implementing a faster amplifier while ensuring >8b linearity.

A two-stage fully differential amplifier (Fig. 19.5.3) was designed using a dual cascode compensation technique to increase the unity-gain bandwidth with 2× smaller compensation capacitance than the equivalent Miller capacitor and reduce gain peaking beyond the unity-gain frequency. In simulations, the amplifier has >71° phase margin with C_f=100fF and up to 5pF of sensor capacitance. The high DC gain (99dB) in conjunction with autozeroing minimizes the input offset voltage that modulates the sensor current during switching. A low-leakage reset switch was designed using three transmission gates to obtain off-leakage less than 100fA. The Hourglass switch was implemented with transmission gates to minimize charge injection. The comparators consist of a single-stage preamplifier and a latch that is autozeroed during the start-up phase to remove offset. The propagation delay of the comparator is less than 5ns to minimize deadzone time and harmonic distortion caused by excess loop delay.

This AFE was implemented in a 0.18μm CMOS process with a 1.8V supply and 0.5V and 1.3V reference voltages. It was characterized with one of the differential inputs connected to a test source while the other was connected to a matched impedance network. Figure 19.5.4 shows the measured *I*-to-*F* conversion of the Hourglass ADC. The Hourglass ADC INL was improved from >±50ppm to ±7ppm after enabling the calibration where the fitted parameters ($A_{DC,closed\text{-}loop}$=64dB and $f_{closed\text{-}loop}$=1.5MHz) closely match the simulation results. Figure 19.5.4 also shows a spectrum of the Hourglass ADC with f_{0SR}=100kHz illustrating the first-order noise shaping. For a conversion time of 400ms (1.8Hz BW), an input-referred noise of 79fA$_{rms}$ was measured. Figure 19.5.5 shows the full DR of the AFE as the current is swept from 100fA to 10μA (160dB) with a measured linearity of ±7ppm.

This AFE consumes 295μW with the amplifier consuming most of the power (Fig. 19.5.6). For flexibility, the digital logic including the predictor, DEM, and linearity correction were implemented off-chip in an FPGA. Simulation of the synthesized digital logic consumed 8μW. Figure 19.5.6 summarizes the AFE performance in comparison to the state-of-the-art current-input ADCs with similar DR and conversion time. A micrograph of the 1.5×2.0mm^2 chip is shown in Fig. 19.5.7 where the AFE occupies an active area of only 0.2mm^2. In summary, this work achieves state-of-the-art performance in terms of normalized conversion time for a 1nA current (0.04ms) and Schreier FoM (197dB) demonstrating an energy efficient, wide dynamic range, high-linearity design for current input biosensors.

References:

[1] M. Stanacevic, et al., "VLSI Potentiostat Array with Oversampling Gain Modulation for Wide-Range Neurotransmitter Sensing," *IEEE TBioCAS*, vol. 1, pp. 63-72, 2007.

[2] S. Dai, et al., "A 155-dB Dynamic Range Current Measurement Front End for Electrochemical Biosensing," *IEEE TBioCAS,* vol. 10, pp. 935-944, 2016.

[3] N. Wood, et al., "Predicting ADC: A New Approach for Low Power ADC Design," *IEEE DCAS*, pp. 1-4, 2014.

[4] K. Nguyen, et al., "A 108dB SNR 1.1mW Oversampling Audio DAC with a Three-Level DEM Technique," *IEEE ISSCC*, pp. 488-489, 2008.

[5] K. L. Chan, et al., "Segmented Dynamic Element Matching for High-Resolution Digital-to-Analog Conversion," *IEEE TCAS-I*, pp. 3383-3392, 2008.

 978-1-5090-4940-0/18/$31.00

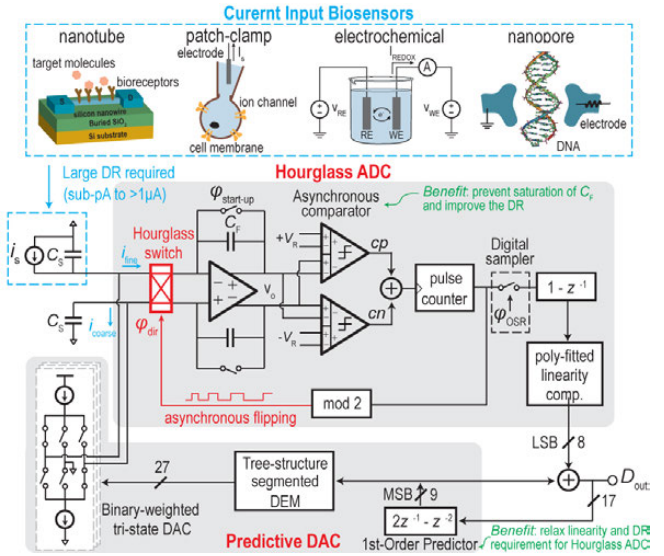


Figure 19.5.1: System architecture of the current measurement front-end.

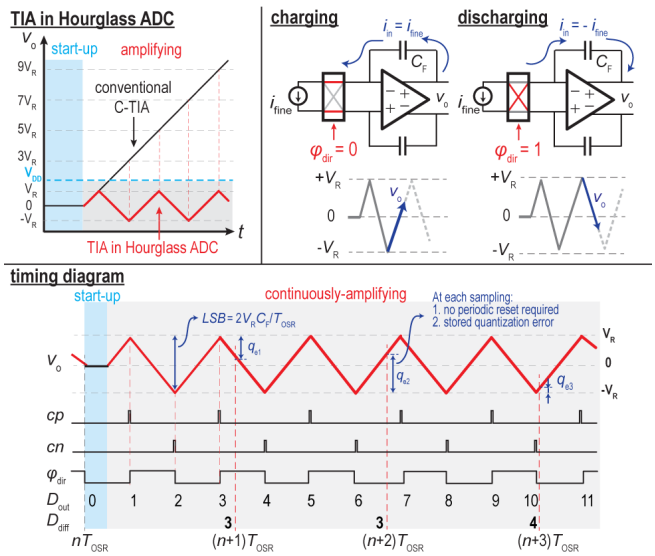


Figure 19.5.2: Operation of the Hourglass ADC.

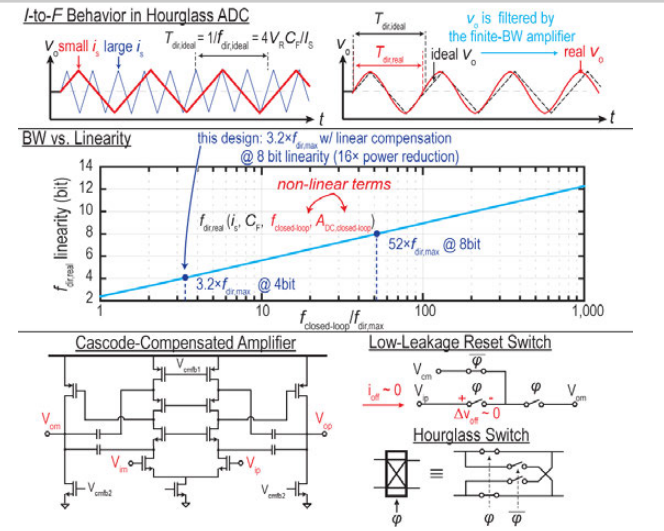


Figure 19.5.3: The *I*-to-*F* behavior and linearity compensation in the Hourglass ADC.

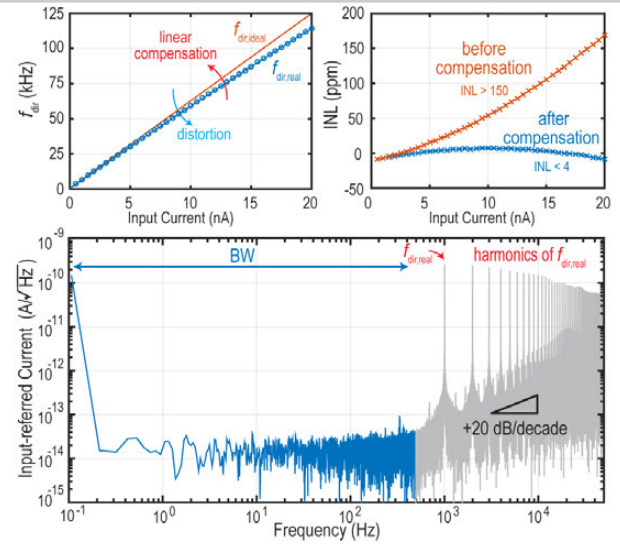


Figure 19.5.4: Measured data showing f_{dir} and INL as a function of the input current and representative spectrum showing noise shaping.

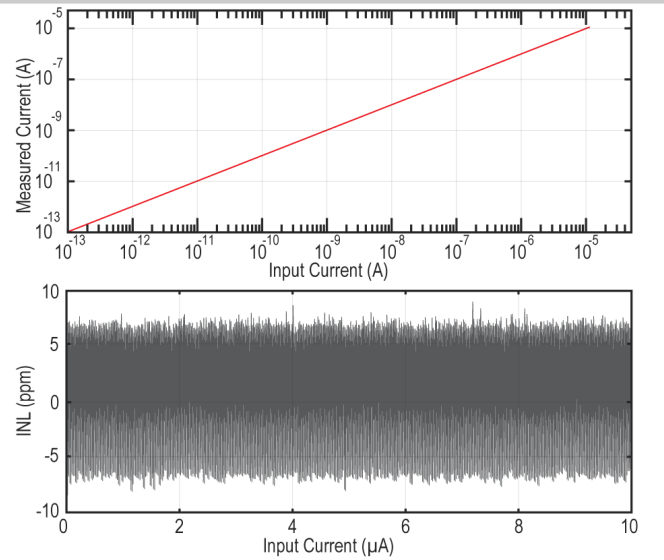
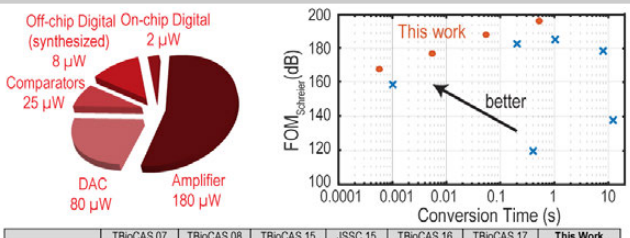


Figure 19.5.5: Measured linearity vs. input amplitude.



	TBioCAS 07	TBioCAS 08	TBioCAS 15	JSSC 15	TBioCAS 16	TBioCAS 17	This Work
Architecture	SDM	SDM	I-to-F	SDM	SDM	SDM	**A-SDM**
Process [μm]	0.5	0.5	0.18	0.18	0.5	0.35	**0.18**
Power [μW]	80	60	5,220	80	240	16.8	**295**
Min Input	100 fA	12 pA	204 fA	1 nA	100 fA	100 pA	**100 fA**
Max Input	1 μA	430 nA	11.6 μA	4 μA	16 μA	3 μA	**10 μA**
Conversion time for Min Input [ms]	8,330	10,000	250	400	1,000	4	**400**
Normalized conversion time for 1nA [ms]	838	120	0.1	400	1	0.4	**0.04**
Max Linearity Error	3.9 nA	3.0 nA	1.16 μA	0.9 nA	1.6 nA	0.1 nA	**70 pA**
Dynamic Range @ BW [dB]	140 @ <0.1Hz	91 @ <0.1Hz	155 @ 1.4Hz	72 @ 2.5Hz	164 @ 1.0Hz	88.9 @ 1 kHz	**100 @ 1.8kHz 120 @ 180Hz 140 @ 18Hz 160 @ 1.8Hz**
FOM$_{Schreier}$ [dB]	172	132	181	103	183	158	**167 @ 1.8kHz 177 @ 180Hz 187 @ 18Hz 197 @ 1.8Hz**

FOM$_{Schreier}$ (dB) = DR(dB) +10log(1/2/T_{conv}/P)

Figure 19.5.6: Performance summary and comparison with previous work.

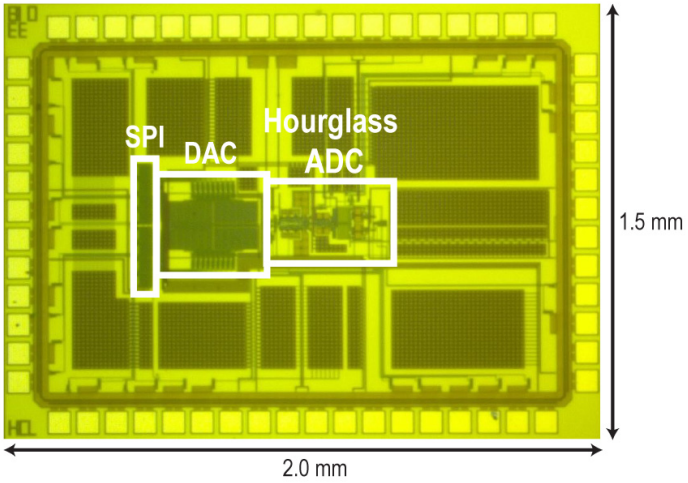


Figure 19.5.7: Die micrograph.

978-1-5090-4940-0/18/$31.00

19.6 A 2.5nJ Duty-Cycled Bridge-to-Digital Converter Integrated in a 13mm³ Pressure-Sensing System

Sechang Oh[1], Yao Shi[1], Gyouho Kim[1,2], Yejoong Kim[1,2], Taewook Kang[1], Seokhyeon Jeong[1,2], Dennis Sylvester[1], David Blaauw[1]

[1]University of Michigan, Ann Arbor, MI; [2]CubeWorks, Ann Arbor, MI

Small form-factor piezoresistive MEMS sensors, often configured in a Wheatstone bridge, are widely used to measure physical signals such as pressure [1-3], temperature [4], force [1], and gas concentration. A common method to realize a digital output from the bridge involves biasing the bridge with a DC voltage source and using a low-noise amplifier followed by an ADC. While a bridge measurement can achieve high resolution and linearity, it is very power hungry [3] because the bridge resistance is low (typically 1-10kΩ). Both the high power and high instantaneous current make it unsuitable as a sensing interface in miniaturized microsystems with battery capacities of <10μAh and ~15kΩ internal resistance [5]. Duty cycled excitation was proposed in [1] to reduce power in moderate dynamic range (DR) applications, lowering bridge excitation energy by up to 125× compared to static biasing. However, the excitation energy consumption (~250nJ) is still much larger than the interface circuit conversion energy, and therefore limits overall sensor energy efficiency. To address this challenge, we propose an energy-efficient highly duty-cycled excitation bridge-sensor readout circuit for small battery-operated systems. Due to high battery resistances, the excitation voltage (V_{EX}) is sourced from an on-chip decoupling capacitance that drops ~100mV during excitation and then slowly recharges from the battery. To avoid accuracy degradation from this voltage fluctuation, the design samples not only the inputs ($V_{IN+/-}$) but also V_{EX}, from which it generates a DAC reference voltage (V_{DAC}). We also propose an offset calibration and input-range matching method. We demonstrate operation of the bridge-to-digital converter (BDC) integrated with a complete and fully functional pressure-sensing system, including a processor, battery, power management unit, RF transmitter, and optical receiver.

Figure 19.6.1 shows the structure of the bridge-sensor interface circuit. The BDC provides V_{EX} to the bridge and senses $V_{IN+/-}$ and V_{EX} with a sampling circuit, followed by V_{DAC} generation, SAR ADC, and an FSM. An RC-relaxation oscillator generates an internal 17.2kHz clock and the FSM is controlled by a bus controller, which connects to other chips in the microsystem. A sampling pulse generator applies a short pulse (SPL) to sampling switches for V_{EX}, the V_{EX} sampling capacitor (C_{EXS}), and input sampling capacitors (C_S). The value of C_S and SPL pulse width are determined by the input resolution requirement. We target 200μV $V_{IN+/-}$ resolution at 3.6V V_{EX}. C_S is set to 4pF so that kT/C noise is <50μV and SPL width is set to 170ns to satisfy >10b accuracy with the RC settling of $V_{IN+/-}$. The bridge is exposed to the supply voltage for only 170ns within the 1ms total conversion time, enabling bridge power consumption to be 6000× less than conventional DC biasing.

Figure 19.6.2 shows the detailed implementation of V_{EX} and V_{DAC} generation circuits. Since V_{EX} is at ground for most of the conversion time and its large V_{SD} and V_{GD} voltage incurs significant GIDL current, these circuits use GIDL reduction devices G1 and G2. SPL is generated by an inverter delay chain, which is 4b-programmable from 60 to 240ns. Once propagation reaches a selected stage, the remainder of the delay chain is gated to reduce energy consumption. Since battery internal resistance is high, it cannot directly supply the bridge excitation current. Instead, sampling current is provided from a 0.48mm² 1.2nF on-chip decoupling capacitor (C_{DECAP}) made up of M1-M4 MOM and MIM. V_H drop is ~100mV during sampling and it is then slowly restored by the battery during subsequent conversion phases. To avoid negative impact of this supply voltage fluctuation effect during ADC conversion, it is necessary to dynamically adjust the conversion to the reduced V_{EX} at the end of excitation. To achieve this, a DAC reference voltage (V_{DAC_REF}) is internally generated by sampling V_{EX} with C_{EXS} (4pF) when SPL=1. Then, V_{DAC_REF} is multiplied by 10/11 to provide >200mV V_{DS} to ensure transistors are in saturation within the amplifier that generates the final regulated output V_{DAC}. Simulated amplifier PSRR is –56dB. The amplifier is designed for 30kHz bandwidth and 60μV integrated noise and draws 160nA. By sampling the excitation voltage in this way, the BDC is also insensitive to supply variation, which is important for sensor nodes operating on small batteries and hence often unstable supplies. The BDC timing diagram is shown in the bottom left of Fig. 19.6.2. After ST_SPL pulses, PREP_VDAC is on and acts to multiply V_{DAC_REF} by 10/11. V_{DAC} settles during the on period of PREP_VDAC, after which the bit-cycle phase is entered.

Figure 19.6.3 shows the proposed SAR ADC with input range matching and offset calibration features. In conventional SAR ADCs, the input voltage is sampled to a binary DAC array. However, in this implementation such an approach would require $V_{IN+/-}$ to drive >12pF for linearity in the target process technology, increasing the sampling time constant and V_{EX} energy by 3×. Targeting a 4pF sampling capacitor instead (as determined by kT/C constraints), C_S is separated from the DAC [6] as shown. Bridge sensor resistance changes at most a few % at full-scale input. To match the input range of the bridge an additional MSB DAC is used. To accommodate an input range from ±50 to 100mV the MSB DAC is implemented with 31b unary capacitors with selection switches. The MSB DAC uses a split-DAC structure to reduce both total DAC capacitance and V_{DAC} amplifier current, which is proportional to the total DAC capacitance load for the same bandwidth constraint as the DAC settling constraint.

During the sampling phase, the DAC purges all its charges while the input is sampled on C_S. To avoid the power consumption of a common-mode reference voltage generation, the DAC uses only V_{DAC} and ground. At the beginning of the bit-cycling phase, the DAC top plate (V_X) connects to the left plate of C_S (V_Y). C_S charge is conserved during the bit-cycling phases, and V_Y change is directly coupled to the comparator input. The comparator operates at 1.2V (V_{1P2}) to reduce power while the DAC operates at 3.3V (V_{DAC}). The comparator is a conventional two-stage clocked comparator with 400fF internal loading capacitor to enhance noise performance. During the transition the common-mode voltage is shifted to the correct range ($0V < V_Z < 1.2V$) by adjusting the M code of the MSB DAC. The BDC can optionally run offset calibration. It operates with shorted inputs (SHRT=1) and B_{OS}=512 during the calibration, and its output is set as B_{OS} in normal operation. The remainder of the conversion process is identical to a conventional SAR ADC. In order to accommodate multiple applications that require different resolution, the BDC conversion can be oversampled with an oversampling rate (OSR) of 1 to 256. This approach repeats the entire conversion process OSR times and accumulates the output codes.

The proposed BDC was integrated in a stacked sensor node system (Fig. 19.6.5, left) composed of a MEMS pressure sensor, battery, and 6 IC layers: radio, decap, processor, energy harvester, photovoltaic cells, and power management unit. The system is powered by two 8μAh thin-film batteries with 3.6-to-4.1V output voltage, which is downconverted to 1.2V and 0.6V by the switched-capacitor power management unit. The system includes 8kB SRAM and an ARM Cortex-M0 processor, which controls BDC operation. The MEMS pressure sensor is on top of the entire stack with a pressure-sensitive top diaphragm. The four electrodes are directly wirebonded to the proposed BDC chip.

The BDC was implemented in 0.18μm CMOS technology with an area of 1.7mm². The BDC output was measured with different MSEL (# of C_M selected) and input voltages (Fig. 19.6.4, top left). MSEL changes the slope and input range from 45 to 110mV. Measured SNR is 46 to 51dB across MSEL from 12 to 31 (Fig. 19.6.4, bottom left). BDC line sensitivity is measured at 3.6, 3.8, and 4.0V V_H (Fig. 19.6.4, top right). The codes shift 0.07code/mV. Total BDC conversion energy is 2.5nJ and Fig. 19.6.4 (bottom right) provides its breakdown from measurement. Testing the BDC with the pressure sensor is shown in the top right of Fig. 19.6.5. With OSR of 4 it achieves 1.1mmHg resolution at 4ms conversion time. The complete sensor system was tested and is fully functional, as shown by system operation in Fig. 19.6.5 (bottom right). The system periodically wakes up from a sleep mode and enters an active mode by releasing power gates and isolation gates, turning on its RC clock, and executing the BDC. Measured data is saved to SRAM and can be transmitted out by radio when needed. Figure 19.5.6 summarizes the BDC and overall system performance from measurement, and also compares it with previous related BDC work.

References:

[1] R. Grezaud, et al., "A Robust and Versatile, –40°C to +180°C, 8Sps to 1kSps, Multi Power Source Wireless Sensor System for Aeronautic Applications," *IEEE Symp. VLSI Circuits*, pp. C310-C311, June 2017.

[2] T. T. Nguyen, et al., "An Energy-Efficient Implantable Transponder for Biomedical Piezo-Resistance Pressure Sensors," *IEEE Sens. J.*, vol. 14, no. 6, pp. 1836-1843, June 2014.

[3] H. Jiang, et al., "An Energy-Efficient 3.7nV/√Hz Bridge-Readout IC with a Stable Bridge Offset Compensation Scheme," *ISSCC*, pp. 172-173, Feb. 2017.

[4] S. Pan, et al., "A CMOS Temperature Sensor with a 49fJK2 Resolution FoM," *IEEE Symp. VLSI Circuits*, pp. C82-C83, June 2017.

[5] S. Oh, et al., "A Dual-Slope Capacitance-to-Digital Converter Integrated in an Implantable Pressure-Sensing System," *IEEE JSSC*, vol. 50, no. 7, pp. 1581-1591, July 2015.

[6] S. Jeong, et al., "A 12nW Always-on Acoustic Sensing and Object Recognition Microsystem Using Frequency-Domain Feature Extraction and SVM Classification," *ISSCC*, pp. 362-363, Feb. 2017.

 978-1-5090-4940-0/18/$31.00

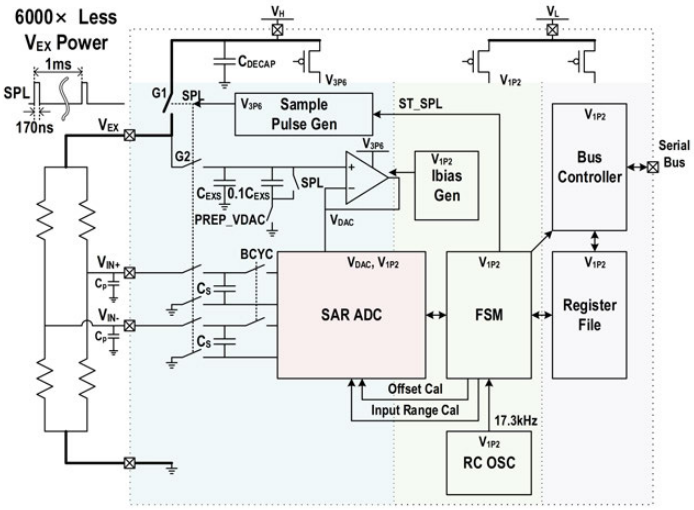


Figure 19.6.1: Structure of the bridge-sensor interface circuit.

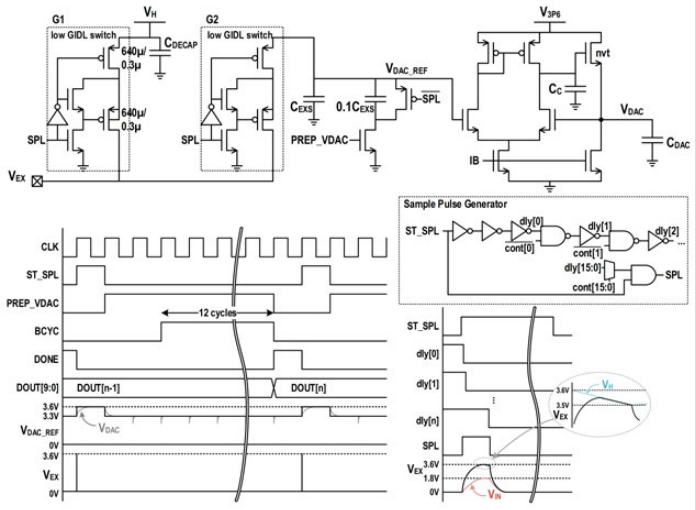


Figure 19.6.2: Detailed implementation of V_{EX} and V_{DAC} generation and BDC timing diagram.

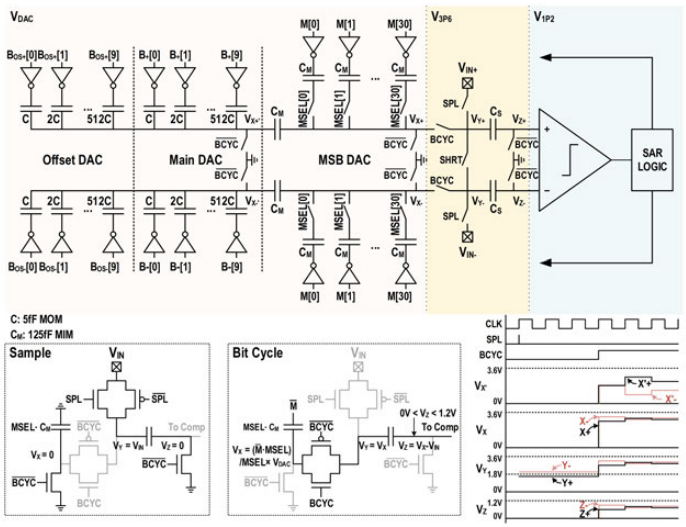


Figure 19.6.3: Implementation of 10b ADC with range matching and offset calibration.

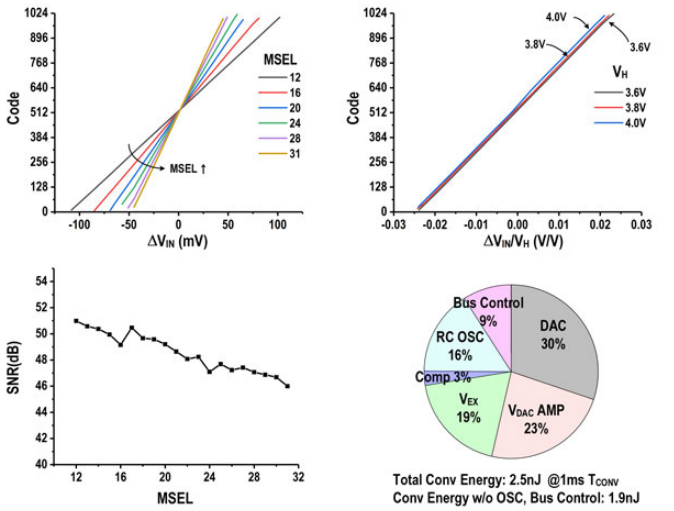


Figure 19.6.4: BDC Measurement results of Code vs. MSEL (# of C_M selected) and V_{IN} (top left), Code vs. V_H and V_{IN} (top right), SNR vs. MSEL (bottom left), and conversion energy breakdown (bottom right).

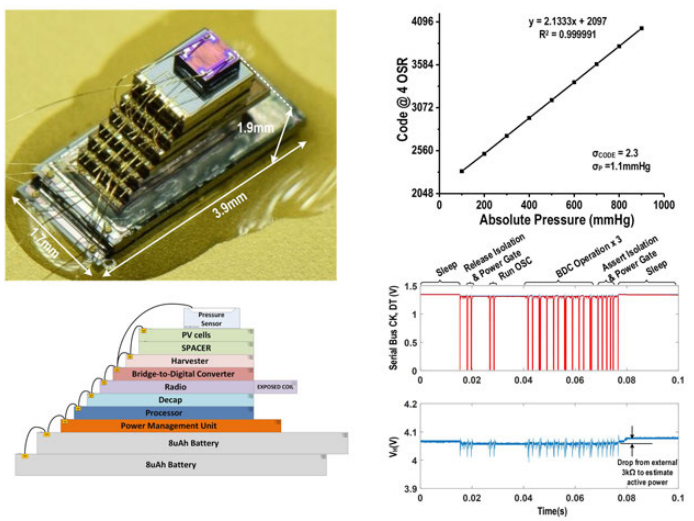


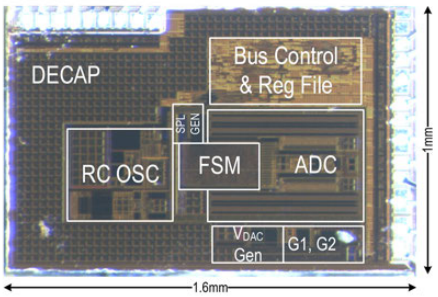
Figure 19.6.5: The proposed 3D stacked pressure-sensing system and the measurement results.

	This Work	This Work OSR=4	VLSI 2017 [1]	Sens J 2014[2]	ISSCC 2017[3]
Technology (nm)	180		180	90	180
Supply Voltage (V)	1.2, 3.6		1.8	1	1.8
Supply Current (µA)	0.65 @1.2V, 0.52 @3.6V		140	52	1200 @1.8V, 1500 @5V
Bridge Voltage (V)	3.6		1.8	1	5
Bridge Resistance (kΩ)	6		1	12	3.3
Conv. Time(µs)	1000	4000	1000	96	500
Energy/Conv exclude Bridge (nJ/Conv)	2	8	61	1.88	1080
Energy/Conv include Bridge (nJ/Conv)	2.5	10	246	5	4870
+/-Input Range(mV)	68		16	12.8	10
SNR (dB)	49.2	54.2	59.0	44.1	95.5
FOMW[1] (pJ/c.s.)	10.6	23.9	337.9	38.3	100.0
FOMS[2] (dB)	132.2	131.2	122.1	124.1	145.6

[1] $FOMW = \frac{E_{CONV\_INC\_BRIDGE}}{2^{(SNR-1.76)/6.02}}$

[2] $FOMS = SNR(dB) + 10\log\left(\frac{1}{2E_{CONV\_INC\_BRIDGE}}\right)$

Figure 19.6.6: Performance summary and comparison.



System Dimension	1.7×3.9×1.9mm³
Battery Capacity	16μAh
Processor	ARM Cortex M0
System Storage	8KB SRAM
Active Power	27μW
Sleep Power	24nW
Processor Clock	80kHz
Internal Supply	0.6, 1.2, 3.6V
RF TX Energy	250nJ/bit

Figure 19.6.7: BDC die micrograph and system specification.

978-1-5090-4940-0/18/$31.00

19.7 A 21.8b Sub-100μHz 1/f Corner 2.4μV-Offset Programmable-Gain Read-Out IC for Bridge Measurement Systems

Jaehoon Jun, Cyuyeol Rhee, Minsung Kim, Junho Kang, Suhwan Kim

Seoul National University, Seoul, Korea

High-resolution read-out integrated circuits (ROICs) are often used in DC measurement systems such as bridge transducers. Since these typically output small-amplitude signals with a bandwidth of a few hertz, a highly linear gain ROIC is required with a resolution above 20b [1,2]. In previous work, high-precision instrumentation amplifiers (IAs) followed by analog-to-digital converters (ADCs) were reported [1-4]. However, IA topologies based on switched-capacitor (SC) or multiple operational amplifiers (opamps) suffer from a poor power-noise tradeoff due to noise folding [2], or the number of amplifiers [3], respectively. Current-feedback IAs (CFIAs) are more power-efficient, but require highly linear feedback resistors [4,5]. This drawback can be addressed by capacitively-coupled IAs (CCIAs), which are even more power efficient [1].

This paper describes a fully integrated 21.8b programmable-gain ROIC for DC measurement applications, which consists of a CCIA, an incremental 2nd-order SC ΔΣ ADC, and a reconfigurable digital filter including a serial interface, as shown in Fig. 19.7.1. The ROIC is designed to have a low 1/f noise corner frequency, which is less than 100μHz. To mitigate flicker noise and any offset error in the main signal path, a chopping and a correlated-double-sampling (CDS) technique are used in the CCIA and the programmable-gain ADC (PG-ADC), respectively. A 2nd-order system-level chopping is also applied to the ROIC to further reduce residual offset and low-frequency noise resulting from circuit mismatch and 1/f noise. At the start of each conversion, the system is reset and the clock signal f_{sys_chop} flips the state of the system-level choppers. After decimation by a sinc^3 filter, the results of four conversions are then combined by a moving-average Finite Impulse Response (FIR) filter, increasing the effective resolution of the ROIC by 1b (Fig 19.7.1). Compared to conventional system-level chopping, in which the results of successive groups of two conversions are averaged, this approach increases the ROIC's throughput and implements a highpass filter that suppresses 1/f noise and reduces residual offset [4]. The input system-level chopper is incorporated into an input multiplexer (MUX) stage to reduce circuit complexity and input-referred thermal noise, which directly results from the on-resistances of the switches. A single channel of the input MUX consists of 3 switches, a low on-resistance main switch and two dummy switches, to minimize the side effects of clock feedthrough and charge injection. An output system-level chopper is implemented in the digital domain, which is relatively tolerant to the circuit nonlinearities.

The CCIA, which is shown in simplified form in Fig. 19.7.1, has an input noise voltage density of 16nV/√Hz, and consists of an input MUX, a ripple-reduction loop (RRL), an impedance-boosting loop (IBL), and a main amplifier. The gain of the CCIA with a rail-to-rail input common-mode voltage range is determined by the ratio between two on-chip capacitors, C_{IN}/C_{FB}. C_{IN} = 16pF and C_{FB} = 1pF or 0.5pF, resulting in a CCIA gain of 16 or 32. The negative-feedback RRL, implemented as an SC topology, suppresses an analog ripple induced by the inner choppers of the CCIA, and the positive-feedback IBL through C_{IB} boosts input impedance by introducing a compensating current, drawn from the input of the ROIC.

Figure 19.7.2 shows a detailed schematic diagram of the main amplifier of the CCIA, which has no resistive loading, together with a sampler stage of the incremental ADC. A differential difference amplifier (DDA) is used to efficiently operate the main signal path and the RRL path. The Miller-compensated chopper-stabilized 2-stage amplifier, which has continuous-time common-mode feedback (CMFB), is designed with a DC gain of 140 dB. To reduce current consumption during slewing, two clamp transistors are added in the 1st-stage of the amplifier. The current consumption of the CCIA, including a Class-AB output stage of the DDA and auxiliary paths, is 292μA, assuring full-driving capability of the ADC.

Figure 19.7.3 shows a schematic diagram of the incremental PG-ADC. It consists of a 2nd-order, 1b ΔΣ modulator, implemented as a cascade of integrators with feedforward form (CIFF) followed by a reconfigurable on-chip decimation filter. The use of a fully differential discrete-time ΔΣ modulator avoids applying a resistive loading to the amplifier of the CCIA. Separate sampling capacitors C_{S1} and C_{DAC} are used to sample the input signal and to feed back the digital-to-analog converter (DAC) signal, respectively [6]. The separated capacitor structure is employed to realize a variable gain of the ADC as 1 or 4, which is defined by C_{S1}/C_{DAC}. The CDS technique is applied to the 1st integrator to reduce flicker noise of the amplifier, and the noise of the 2nd integrator is shaped out of band by the 1st integrator. The amplifiers with SC-CMFB in the ADC are implemented in a similar way to that of the CCIA, except for the 2nd stage, which is omitted to reduce power consumption. The input sampling capacitor and sampling clock frequency of the unity-gain-mode ADC are 6pF and 61.44kHz, respectively, while only drawing 34μA from a 3V analog supply. The digital filter has a controllable output data-rate (ODR), and consists of a sinc^3 filter followed by an FIR filter, and a serial interface. The programmable ODR (5/10/20/80 SPS) digital filter has notches, which effectively suppress out-of-band noise and 50/60Hz interference.

The programmable-gain ROIC was fabricated in a 0.13μm standard CMOS process with an active area of 0.88mm^2, which includes both the analog and digital circuits. The gain of the ROIC is determined by the combined gain of the CCIA and the ADC, which can be set to 1, 4, 16, 32, 64, or 128. Figure 19.7.4 shows the measured effective resolution (ER) of the system versus the final digital ODR for each gain of the ROIC. The circuit draws a current of 326μA from a 3V analog supply and 20μA from a 1.5V digital supply. Each measurement point in the plot is a value averaged over 2^{12} output data values. The noise histogram shows that output rms noise is 4.29 LSB, resulting in an equivalent ER of 21.8b in a conversion time of 200ms. The maximum noise-free resolution (NFR) is 19.1b, calculated from 6.6σ noise distribution satisfying the industry-standard reliability of 99.9%. As illustrated in Fig. 19.7.5, the measured output code is highly linear with a coefficient of determination of 0.9999.

Figure 19.7.6 summarizes the performance of our programmable-gain ROIC and compares it with previously published ROICs designed for high-resolution DC measurement. The ROIC achieves a resolution of over 21b with a full-on-chip reconfigurable digital filter. The combination of the 2nd-order system-level chopping, the inner chopping, and the CDS technique, results in a 1/f corner frequency of less than 40μHz. The ROIC exhibits an average offset of 2.4μV, a CMRR of 124dB, and a PSRR of 118dB, based on measurements across 8 tested sample chips. An annotated die micrograph is shown in Fig. 19.7.7.

Acknowledgement:
The authors acknowledge very useful comments from Prof. Kofi Makinwa.

References:

[1] H. Jiang, et al., "An Energy-Efficient 3.7nV/√Hz Bridge-Readout IC with a Stable Bridge Offset Compensation Scheme", *ISSCC*, pp. 172-173, Feb. 2017.
[2] C. D. Ezekwe, et al., "A 6.7nV/√Hz Sub-mHz-1/f-Corner 14b Analog-to-Digital Interface for Rail-to-Rail Precision Voltage Sensing", *ISSCC*, pp. 246-247, Feb. 2011.
[3] M. Maruyama, et al., "An Analog Front-End for a Multifunction Sensor Employing a Weak-Inversion Biasing Technique with 26 nVrms, 25 aCrms, and 19 fArms Input-Referred Noise," *IEEE JSSC*, vol. 51, no. 10, pp. 2252-2261, Oct. 2016.
[4] R. Wu, et al., "A 20-b 40-mV Range Read-Out IC With 50-nV Offset and 0.04% Grain Error for Bridge Transducers", *IEEE JSSC*, vol. 47, no. 9, pp.2152-2163, Sept. 2012.
[5] F. Michel and M. Steyaert, "On-Chip Gain Reconfigurable 1.2V 24μW Chopping Instrumentation Amplifier with Automatic Resistor Matching in 0.13μm CMOS", *ISSCC*, pp. 372-373, Feb. 2012.
[6] J. Jun, et al., "An SC Interface With Programmable-Gain Embedded ΔΣ ADC for Monolithic Three-Axis 3-D Stacked Capacitive MEMS Accelerometer," *IEEE Sensors J.*, vol. 17, no. 17, pp. 5558-5559, Sept. 2017.

 978-1-5090-4940-0/18/$31.00

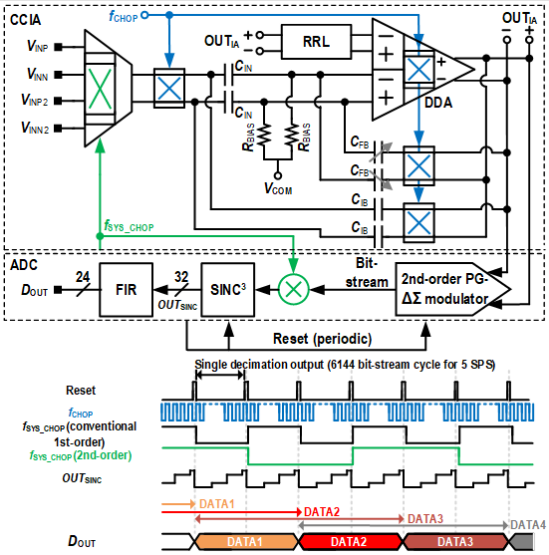


Figure 19.7.1: Simplified block diagram of the ROIC.

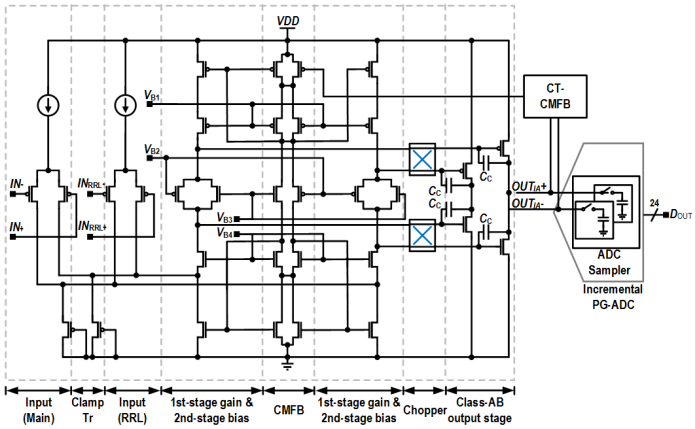


Figure 19.7.2: Circuit diagram of the DDA.

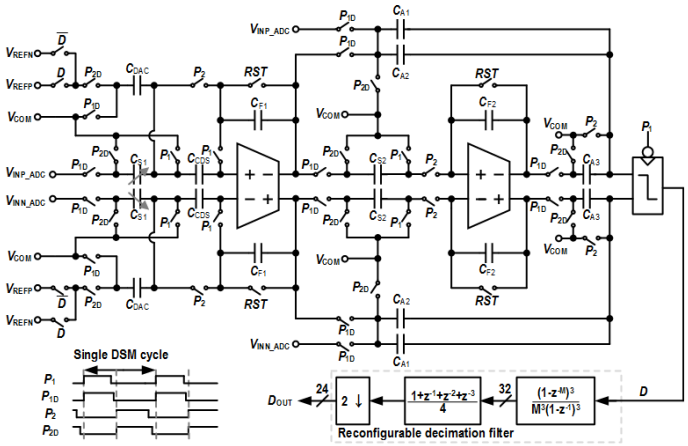


Figure 19.7.3: Simplified schematic diagram of the programmable-gain incremental ADC.

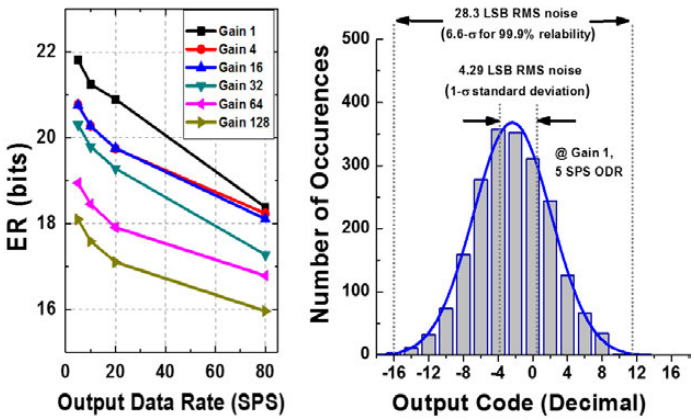


Figure 19.7.4: Measured ER versus ODR and noise histogram.

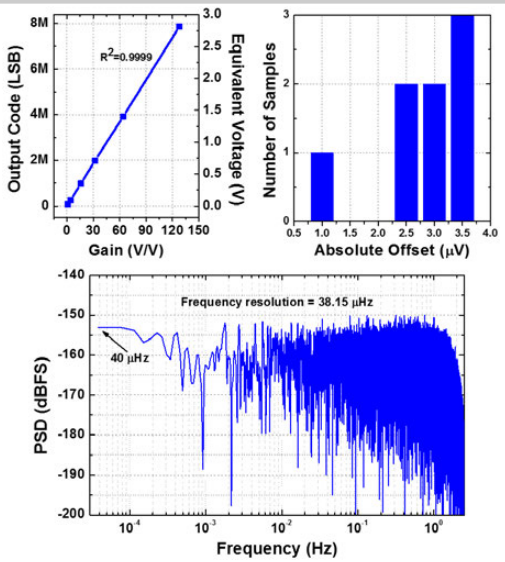


Figure 19.7.5: Measured ROIC output versus gain, offset histogram and PSD.

Parameter	This work	[1] ISSCC' 17	[3] JSSC' 16	[4] JSSC' 12
Architecture	CCIA+ DT-ΔΣ ADC	CCIA+ CT-ΔΣ DSM	3-AMP IA+ CT-ΔΣ ADC	CFIA+ DT-ΔΣ DSM
Technology (μm)	0.13	0.18	0.18	0.7
Die area (mm²)	0.88	0.73	4.5	5
Supply voltage (V)	3.0 (analog) 1.5 (digital)	1.8	1.55	5
Supply current (μA)	326 (analog) 20 (digital)	1200	1560	270
Conversion time (ms)	1.56 ~ 200	0.5	50	170
Gran range	1 ~ 128	100 (fixed)	5 ~ 160	Stable for gain>20 (off-chip resistor)
IRN measured noise$_{RMS}$ (μV)	1.53 (@ gain 1)	—	—	—
Effective resolution$_{MAX}$ (bit)	21.8	15.4	20.3 (only ADC)	20
DC offset (μV)	2.4	7	100	0.048
DC CMRR/PSRR (dB)	124/118 (typical)	134/— (typical)	—/—	140/— (typical)
±Input range (V)	2.8	0.01	0.31	0.04
NEF / PEF* of ROIC	11.1/372	5.0/44	12.5/242	10.4/541
1/f corner (mHz)	<0.04	4	1000	1
Digital filter	On-chip	Off-chip	On-chip	Off-chip
FOM$_{MAX}$ (dB) **	166.9	151.1	160.1	155.5

* PEF = NEF²×VDD
** FOM$_{MAX}$ (dB) = SNR$_{MAX}$ + 10·log(1/2·Power·Conversion time))

Figure 19.7.6: Performance summary and comparison.

19

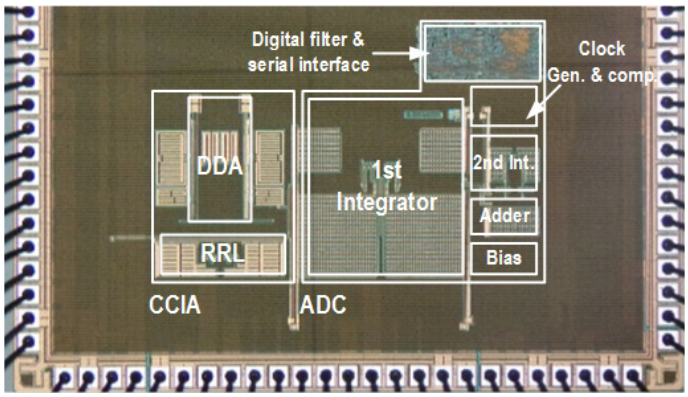


Figure 19.7.7: Die micrograph.

978-1-5090-4940-0/18/$31.00

19.8 A Phase-Domain Readout Circuit for a CMOS-Compatible Thermal-Conductivity-Based Carbon Dioxide Sensor

Zeyu Cai[1,2], Robert van Veldhoven[2], Hilco Suy[3], Ger de Graaf[1], Kofi A. A. Makinwa[1], Michiel Pertijs[1]

[1]Delft University of Technology, Delft, The Netherlands
[2]NXP Semiconductors, Eindhoven, The Netherlands
[3]ams AG, Eindhoven, The Netherlands

The measurement of carbon-dioxide (CO_2) concentration is very important in home and building automation, e.g. to control ventilation in energy-efficient buildings. This application requires compact, low-cost sensors that can measure CO_2 concentration with a resolution of <200 ppm over a 2500ppm range. Conventional optical (NDIR-based) CO_2 sensors require components that are CMOS-incompatible, difficult to miniaturize and power-hungry [1]. Due to their CMOS compatibility, thermal-conductivity-based sensors are an attractive alternative [2,3]. They exploit the fact that the thermal conductivity (TC) of CO_2 is lower than that of the other constituents of air, so that CO_2 concentration can be indirectly measured via the heat loss of a hot wire to ambient. However, this approach requires the detection of very small changes in TC (0.25 ppm per ppm CO_2 [3]).

This paper presents a TC-based CO_2 sensor that achieves a resolution of 94ppm (rms) while dissipating only 12mJ per measurement, >10× less than prior CMOS-compatible CO_2 sensors [3]. This is achieved by using a high-resolution phase-domain $\Delta\Sigma$ modulator (PD$\Delta\Sigma$M) to sense the thermal time constant τ_{th} of a hot wire. The time constant can be approximated by the product of the wire's thermal capacitance (C_{th}) and its thermal resistance to ambient (R_{th}). Since the wire loses part of its heat to the surrounding air, R_{th} depends on the TC of the air, and thus on CO_2 concentration (Fig. 19.8.1), while C_{th} can be considered constant. Driving the wire with periodic heat pulses then results in phase-shifted temperature variations $\Delta T(t)$, which are digitized by the PD$\Delta\Sigma$M, and from which τ_{th} can be derived [4, 5]. To maximize the sensitivity of the detected phase shift to τ_{th}, the wire is driven at $f_{drive} = 1/2\pi\tau_{th}$ = 9.26kHz, i.e. at the pole of the thermal filter. Compared to measuring the steady-state temperature of a hot wire [3], this approach has the important advantage that the absolute temperature and power levels of the transducer do not need to be accurately stabilized or measured. In contrast with earlier TC sensors based on transient measurements [2,4], which use separate heaters and temperature sensors, we combine these two functions in a single resistive transducer. This greatly simplifies fabrication, because only one extra etch step is required to realize a tungsten hot-wire transducer in the via layer of a standard CMOS process [3].

To produce heat pulses at a frequency f_{drive}, the hot wire is driven by a pulsed current I_d (2.5mA). Since its resistance $R(t)$ is temperature dependent, the resulting temperature variations $\Delta T(t)$ can then be sensed via the corresponding resistance changes $\Delta R(t)$ in the hot wire. To sense $R(t)$ independently of the switched drive current I_d, an additional sense current I_s (0.5 mA), switched at $f_{sense} = 15f_{drive}$, produces a modulated voltage that is proportional to $R(t) = R_0(1+\alpha\Delta T)$, where α = 0.4%/°C for tungsten, and thus has a constant sensitivity to $\Delta T(t)$ (Fig. 19.8.2, left). To ease the detection of this voltage in the presence of the large voltage transients at f_{drive} (~ 300mV$_{pp}$), two identical transducers (R_{t1} and R_{t2}) are heated simultaneously and read out differentially via out-of-phase sense currents. Thus, the signal at f_{drive} is converted into a common-mode signal that can be rejected, while the differential signal is demodulated using a chopper switch (Fig. 19.8.2, middle).

Even with this arrangement, the subsequent readout circuit must still have a large dynamic range. This is because the temperature-induced resistance change ΔR (~3Ω) is small compared to the baseline resistance R_0 (~110Ω), while the change in ΔR due to changes in CO_2 concentration is even smaller (~1.5μΩ per ppm CO_2). To cancel the voltage steps associated with R_0, two poly resistors $R_{p1,2}$ (~R_0) are connected in series with the transducers, and the sense currents are routed such that the additional voltage drop I_sR_p cancels out I_sR_0 (Fig. 19.8.2, right). The remaining differential signal V_s is ideally equal to $I_s\Delta R$ (~1.5mV$_{pp}$, 200× smaller than the initial transients), and is thus proportional to $\Delta T(t)$. However, the mismatch between the transducers and the poly resistors leads to ripple. To minimize this, the drive and sense currents can be trimmed using three 6b current-trimming DACs with an LSB of 0.4%I_s (not shown), thus reducing the residual ripple to < 0.1mV.

To detect the CO_2-dependent changes in τ_{th}, the phase shift of ΔR relative to f_{drive} is digitized by a PD$\Delta\Sigma$M similar to [5] (Fig. 19.8.3). Rather than directly demodulating the sense voltage, as in Fig. 19.8.2, it is first converted into a current by a transconductor g_m and then detected by a chopper demodulator, resulting in a signal proportional to ΔR. A second chopper then multiplies this signal by a phase reference (a phase-shifted version of f_{drive}), resulting in a signal of which the DC component is proportional to their phase difference. This difference is integrated and quantized by a comparator clocked at $f_{samp} = f_{drive}$ to form a $\Delta\Sigma$ loop in which the bit-stream output *bs* switches between two phase references ϕ_0 and ϕ_1 ($\phi_0 - \phi_1 = 4°$). This loop nulls the integrator's average input, thus ensuring that the average reference phase tracks the phase of $\Delta T(t)$, which can therefore be derived from the bit-stream average. To simplify the circuit, the two choppers at the output of the transconductor have been merged into a single chopper, which is driven by the logical product of f_{sense} and the chosen phase reference.

Both the transducers and the readout circuit have been implemented in the same 0.16μm CMOS technology (Fig. 19.8.7), with active areas of 0.3mm^2 and 3.14mm^2, respectively. For flexibility, they have been realized on separate chips and connected at the PCB level, and so they can be readily co-integrated. The modulator control signals were generated using an FPGA. The readout circuit consumes 6.8mW from a 1.8V supply, 6.3mW of which is dissipated in the transducers. Fig. 19.8.4 shows the measured resolution at different oversampling ratios (OSR). A resolution equivalent to 94ppm CO_2 is reached at an OSR of 16384, which corresponds to a measurement time of 1.8s, and an energy consumption of 12mJ. The measured phase shift as a function of the drive frequency, measured using a larger full scale $\phi_0 - \phi_1 = 12°$ for clarity, shows the first-order behaviour associated with the hot wire thermal time constant.

To measure its CO_2 response, the sensor was placed in a sealed box along with an NDIR reference CO_2 sensor [1]. Like other TC-based CO_2 sensors [2,3], the readings of the sensor are affected by variations in ambient conditions, which need to be compensated in a final product. In our experiment, ambient temperature, humidity and pressure sensors were placed in the box to facilitate cross-sensitivity compensation. Figure 19.8.5 shows the good agreement between the readings of our sensor and the CO_2 concentration measured by the reference sensor.

Figure 19.8.6 summarizes the performance of the chip and compares it with the prior art. By using a low-noise phase-domain $\Delta\Sigma$ modulator and substantially reducing the required dynamic range using differential sensing and baseline compensation, this work achieves the lowest energy consumption per measurement, while using a transducer fabricated in standard CMOS technology with minimum post-processing. This results in a fully integrated CO_2 sensor in only ~3mm^2, making it a promising candidate for CO_2 sensing in cost- and energy-constrained applications.

Acknowledgement:
This work is supported by NXP Semiconductors and ams AG. The authors would like to thank Zu-yao Chang, Lukasz Pakula, and Zhuoling Liao for their support.

References:
[1] SenseAir K30 datasheet, SenseAir [Online].
Available: http://www.senseair.com/.
[2] K. Kliche, et al., "Sensor for Thermal Gas Analysis Based on Micromachined Silicon-Microwires," *IEEE Sensors J.*, vol. 13, no. 7, pp. 2626–2635, July 2013.
[3] Z. Cai, et al., "A Ratiometric Readout Circuit for Thermal-Conductivity-Based Resistive CO_2 Sensors," *IEEE JSSC*, vol. 51, no. 10, pp. 2463-2474, Oct. 2016.
[4] C. van Vroonhoven, et al., "Phase Readout of Thermal Conductivity-Based Gas Sensors," *IEEE Int. Workshop on Advances in Sensors and Interfaces (IWASI),* pp. 199-202, 2011.
[5] S. M. Kashmiri, et al., "A Scaled Thermal-Diffusivity-Based 16 MHz Frequency Reference in 0.16 μm CMOS," *IEEE JSSC*, vol. 47, no. 7, pp. 1535-1545, July 2012.
[6] T.A. Vincent and J.W. Gardner, "A Low Cost MEMS Based NDIR System for the Monitoring of Carbon Dioxide in Breath Analysis at ppm Levels," *Sens. and Actuators B, Chem., (Elsevier)*, vol. 236, pp. 954-964, Nov. 2016.

 978-1-5090-4940-0/18/$31.00

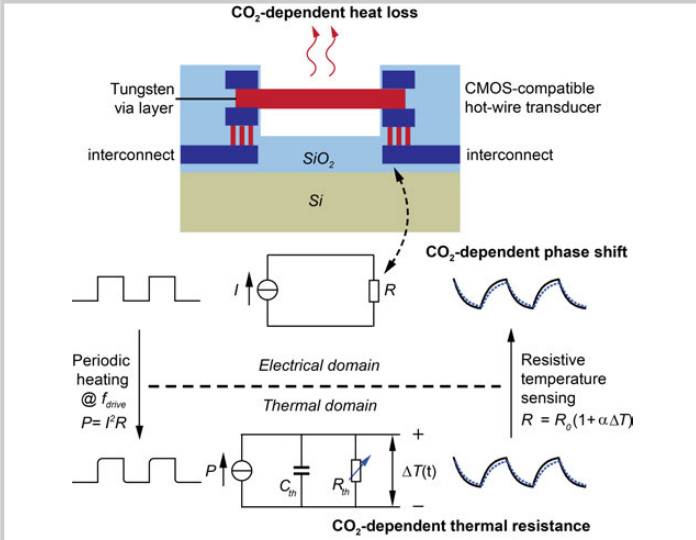


Figure 19.8.1: Transient thermal-resistance (thermal time-constant) measurement principle.

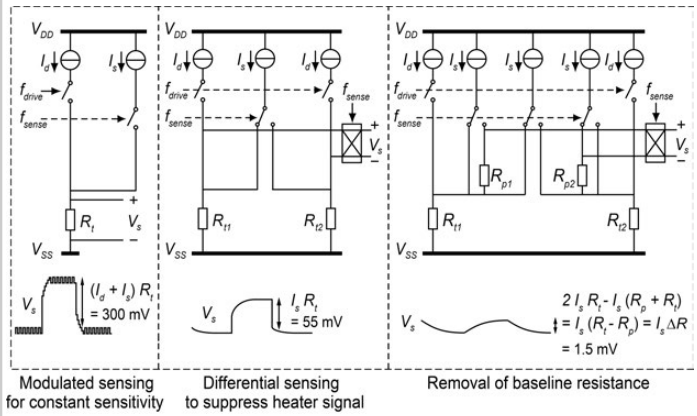


Figure 19.8.2: Sensing the temperature-induced resistance changes using modulation, differential sensing and baseline cancellation.

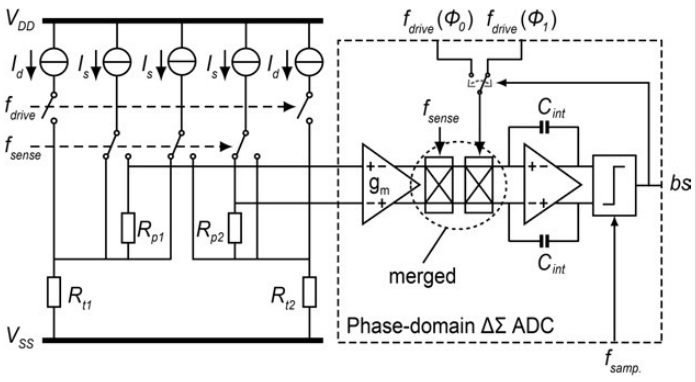


Figure 19.8.3: Circuit diagram of the proposed readout circuit.

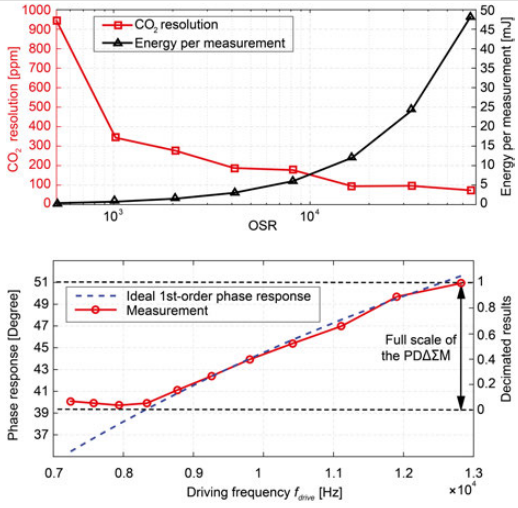


Figure 19.8.4: (top) Measured resolution (standard deviation of 20 consecutive measurements) and energy per measurement as a function of OSR; (bottom) measured phase shift as a function of the drive frequency.

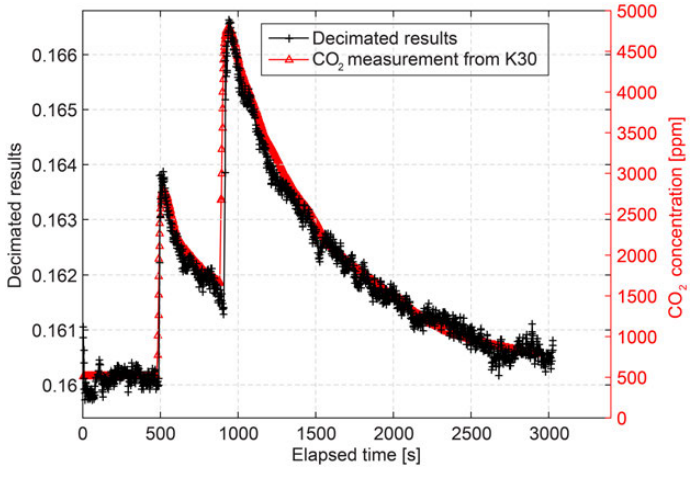


Figure 19.8.5: Transient CO_2 response of the CO_2 sensor and an NDIR-based reference sensor (K30).

Parameter	This work	[3]	[2]	[1]	[6]
Method	TC	TC	TC	NDIR	NDIR
Technology	CMOS (0.16 µm)	CMOS (0.16 µm)	SOI MEMS	Module	SOI MEMS
On-chip readout	Y	Y	N	N	N
Area (sensor)	0.3 mm²	0.6 mm²	16 mm²	-	†0.3 mm²
Area (readout)	3 mm²	3 mm²	-	-	-
Supply voltage	1.8 V	1.8 V	-	5-14 V	-
Power consumption	6.8 mW	11.2 mW	3 mW	200 mW	200 mW
Meas. time	1.8 s	30 s	60 s	2 s	2.4 s
CO_2 resolution	94 ppm	202 ppm	456 ppm	20 ppm	250 ppm
Energy / meas.	12 mJ	336 mJ	180 mJ	400 mJ	480 mJ

†Area of the IR emitter only, excluding 80-mm light tube and an IR detector

Figure 19.8.6: Performance summary and benchmarking.

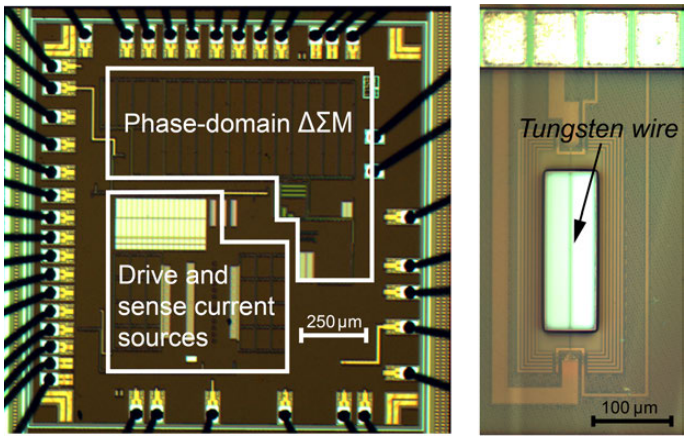


Figure 19.8.7: Die micrograph of the readout circuit and the transducer.

978-1-5090-4940-0/18/$31.00

Session 20 Overview: *Flash-Memory Solutions*

MEMORY SUBCOMMITTEE

Session Chair:
Ki-Tae Park
Samsung, Hwaseong, Korea

Associate Chair:
Yan Li
Western Digital, Milpitas, CA

Subcommittee Chair: ***Leland Chang,*** IBM, Yorktown Heights, NY

Continued proliferation of semiconductors for a smarter society drives the evolution of flash memory technologies towards higher density, lower power consumption, and lower cost. This year, a new generation of 3D NAND Flash memory with up to 96-stacked word-line layers is introduced. For the first time, a memory with over 1Tb density is demonstrated using a 4b/cell 3D NAND technology. An ultra-low latency flash controller with a new high-speed 3D NAND is proposed in order to fill a large performance gap between DRAM and Flash memories.

 978-1-5090-4940-0/18/$31.00

8:30 AM

20.1 A 512Gb 3b/Cell 3D Flash Memory on a 96-Word-Line-Layer Technology

H. Maejima, Toshiba Memory, Yokohama, Japan

In Paper 20.1, Toshiba and Western Digital present a 512Gb 3b/cell 3D NAND Flash with an advanced 96-stacked WL-layer technology. A start bias control scheme designed for 3D NAND is proposed to enhance program performance by 7% and a new smart on-chip V_{th}-tracking read that can support program suspend function is also presented.

9:00 AM

20.2 A Flash Memory Controller for 15µs Ultra-Low-Latency SSD Using High-Speed 3D NAND Flash with 3µs Read Time

W. Cheong, Samsung Electronics, Hwaseong, Korea

In Paper 20.2, Samsung presents an ultra-low latency Flash controller using a high-performance 3D NAND that supports a 15µs-latency SSD, which is over 3-5 times faster than a conventional SSD. To provide such a high-performance gain, a split DMA architecture and an advanced suspend/resume DMA operation are proposed.

9:30 AM

20.3 A 1Tb 4b/Cell 64-Stacked-WL 3D NAND Flash Memory with 12MB/s Program Throughput

S. Lee, Samsung Electronics, Hwaseong, Korea

In Paper 20.3, Samsung presents a 1Tb NAND Flash in 64 stacked layers by using a 4b/cell technology. It achieves a 5.63Gb/mm^2 areal density; the highest density ever. In order to control tight V_{th} distributions at 12MB/s, a new fast unselect precharging scheme and a slow bit bypass scheme are proposed.

20.1 A 512Gb 3b/Cell 3D Flash Memory on a 96-Word-Line-Layer Technology

Hiroshi Maejima[1], Kazushige Kanda[1], Susumu Fujimura[1], Teruo Takagiwa[1], Susumu Ozawa[1], Jumpei Sato[1], Yoshihiko Shindo[1], Manabu Sato[1], Naoaki Kanagawa[1], Junji Musha[1], Satoshi Inoue[1], Katsuaki Sakurai[1], Naohito Morozumi[1], Ryo Fukuda[1], Yuui Shimizu[1], Toshifumi Hashimoto[1], Xu Li[1], Yuuki Shimizu[1], Kenichi Abe[1], Tadashi Yasufuku[1], Takatoshi Minamoto[1], Hiroshi Yoshihara[1], Takahiro Yamashita[1], Kazuhiko Satou[2], Takahiro Sugimoto[1], Fumihiro Kono[1], Mitsuhiro Abe[1], Tomoharu Hashiguchi[1], Masatsugu Kojima[1], Yasuhiro Suematsu[2], Takahiro Shimizu[1], Akihiro Imamoto[1], Naoki Kobayashi[1], Makoto Miakashi[1], Kouichirou Yamaguchi[1], Sanad Bushnaq[1], Hicham Haibi[1], Masatsugu Ogawa[1], Yusuke Ochi[1], Kenro Kubota[2], Taichi Wakui[2], Dong He[1], Weihan Wang[1], Hiroe Minagawa[1], Tomoko Nishiuchi[1], Hao Nguyen[3], Kwang-Ho Kim[3], Ken Cheah[3], Yee Koh[3], Feng Lu[3], Venky Ramachandra[3], Srinivas Rajendra[3], Steve Choi[3], Keyur Payak[3], Namas Raghunathan[3], Spiros Georgakis[3], Hiroshi Sugawara[3], Seungpil Lee[3], Takuya Futatsuyama[1], Koji Hosono[1], Noboru Shibata[1], Toshiki Hisada[1], Tetsuya Kaneko[1], Hiroshi Nakamura[1]

[1]Toshiba Memory, Yokohama, Japan
[2]Toshiba Memory Systems, Yokohama, Japan; [3]SanDisk, Milpitas, CA

The first multi-layer stacked 3D Flash memory was proposed as BiCS FLASH in 2007 [1]. Since then, memory bit density has grown rapidly due to the increase in the number of stacked layers from continuous 3D technology innovations. On the other hand, the multi-level-cell technology, which was initially proposed for 2D Flash, has also been adopted to 3D Flash memories. The first 3b/cell 32-layer Flash was presented in 2015 [2], followed by a 48-layer one in 2016 [3], and a 64-layer one in 2017 [4,5]. This paper describes a 512Gb 3b/cell 3D Flash memory in a 96-word-line-layer BiCS FLASH technology. This work implements three key technologies to improve performance: (1) a string based start bias control scheme achieves a 7% shorter program time; (2) a smart V_t-tracking read improves read retry performance by minimizing the tracking time and supporting a program suspend read function, and; (3) a low-pre-charge sense-amplifier bus scheme reduces both the power consumption and the data-transfer time between the sense amplifier (SA) and the data cache by half. Figure 20.1.1 shows the die micrograph and the summary of the key features of the chip.

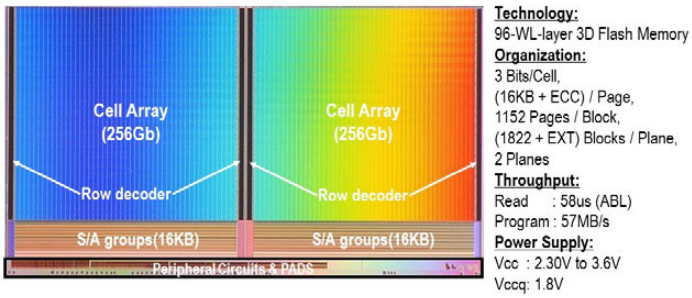
In 2D NAND Flash the cell-to-cell interference is large, that is each single cell is subjected to at least two incremental step-up program pulse (ISPP) sequences during the programming of multi-level data to the cell. A start bias control scheme was proposed to improve the programming throughput [6], where the optimal start programming voltage (V_{PGM}) is measured during the prior ISPP sequence (LSB page programming) and is applied to the subsequent ISPP sequence (MSB page programming). On the other hand, due to the small cell-to-cell interference, the full-sequence (FS) programming method is commonly used for 3D Flash memory, where all of the states are programmed directly from the erased state by one ISPP sequence [2] as shown in Fig. 20.1.2. As such, an adequate V_{PGM} for each program sequence cannot be acquired, as a result there is no room to shorten the programming time. This work introduces a string-based start-bias control (SSBC) scheme to improve program performance. As shown in Fig. 20.1.2, each block consists of four strings. The optimal V_{PGM} level is measured during the first string program sequence and is applied to the remaining cells in other strings of the same block. This results in saving a number of program loops for the second-to-last cells. The program speed for 3D Flash has a strong dependency on the WL layers as the memory-hole size gradually varies from layer to layer. The SSBC scheme is suitable for 3D Flash structures as the program speed for cells on common WL layers is almost the same. The measured V_{th} distribution and t_{PROG} for all WL layers are shown in Fig. 20.1.3, the SSBC V_{th} distribution is practically the same as that without SSBC. However, the program throughput is improved by 7% on average, because t_{PROG} is reduced by 9% for the second to fourth strings.

A key design challenge in recent years is enhancing the retry read performance of a 3D Flash memory. The valley tracking read [5] minimizes the BER by adaptively finding the optimal read voltage. We propose a smart V_{th} tracking read (SVTR) to improve the retry read performance by reducing the number of tracking cycles, and to support a program-suspend read function. Figure 20.1.4 shows the SVTR waveform for a middle-page read. The read operation consists of two parts; (1) a V_t tracking read (VTR), where the selected WL level moves from the MP1-state to the MP2-state, and then to the MP3-state. The optimal read voltage is to be tracked for each state. The number of stages related to each state is parameterizable and the WL voltage step has enough resolution to guarantee the accuracy of VTR. A bit count operation is executed in parallel with VTR, and the bit whose V_t is distributed between the read voltage of the N^{th} stage read and $(N+1)^{th}$ stage read is counted concurrently with the $(N+2)^{th}$ stage read operation. VTR is used with a shielded-bitline current-sense (SBL) scheme [4], because SBL suppresses BL coupling noise thereby realizing a shorter read time [4]. The number of S/A latches in use is reduced by half in comparison with a conventional all bitline current sense (ABL) scheme. (2) A calibrating read (CALR), in which the optimal read voltage that is determined from the bit count result of the VTR is applied. A CALR is executed with ABL to readout all of the data. These two parts are sequentially executed by one command. To improve read performance a high-state option is implemented to minimize the sequence of the 1^{st} part (VTR). This option works based on the fact that the V_t shift of each cell due to data retention is correlated with its level. Moreover, the V_t shift of lower states can be determined from the measured V_t shifts of higher states. In this option, an optimal read voltage of the lower state CALR is calculated using the high-state VTR result and the V_t shift correlation formula. The coefficient of that formula are stored in the registers of the chip in advance. VTR is carried out only on the highest state. For a middle-page read, the CALR level for the 3 states are determined from MP3-state VTR result, thus tracking time is reduced 1/3. To further improve read performance, the program suspend read function with SVTR is realized by utilizing SA latches, which are not occupied by the stored programming data as explained in Fig. 20.1.5. The wait time for an interrupting read is successfully reduced to 50µs by suspending an ongoing program; without suspend it can take up to 1.85ms to initiate SVTR. Figure 20.1.5 also shows the program-suspend read sequence and data assign events for the SA data latches. When a read command is issued during program, (1) the program sequence is suspended. (2) a part of the data stored in the even-page latches is temporarily moved to the odd-page ones to make space for VTR. (3) VTR is executed for even 8kB page data by SBL. (4) the data temporarily moved to the odd-page latches is restored back to the even-page latches, and (5) CALR is run for 16kB page data with ABL. After reading out the data, the programming sequence is resumed.

The presented memory has 16kB SA groups per plane. Each SA group consists of a sense amplifier and multiple sense-amplifier data latches (SADLs). These SADLs are connected to the cache data latches (XDLs) and are placed next to the SA groups by bus lines named DBUS. The total parasitic capacitance of the DBUS is large and so the peak current will be large if all DBUS' are pre-charged simultaneously before transferring the data. In prior work, the data transfer from XDL to SADL via DBUS is divided into 16 cycles to disperse the peak current at the cost of throughput. There is an innate tradeoff between reducing the data transfer time and suppressing the peak current. A low pre-charge SA bus scheme (LPSAB) is proposed to eliminate this tradeoff, where the pre-charge level of DBUS is suppressed without any SA area overhead. Figure 20.1.6 shows a data transfer from SADL to XDL with LPSAB. (1) DBUS is pre-charged to 1V by clamping the gate of the pre-charger (DPC), while conventionally DBUS is pre-charged to V_{DD} (2.2V). (2) V_{TG} which is 0.5V lower than V_{PC} is applied to the gate of the transistor connecting DBUS and SADL. DBUS will be discharged to V_{SS} if SADL has stored-0. For a stored-1, DBUS will stay at 1V. (3) the V_{TG} level is also applied to the transfer gate of XDL, and data is transferred to XDL. The V_{PC} generator contains a replica transistor to compensate for PVT variation. LPSAB reduces the power consumed by the bus operation by 50%. As the data transfer is not as partitioned as in the former scheme, the data throughput is doubled.

Acknowledgements:
The authors thank H. Mukai, H. Saito, H. Maekawa, M. Kajimoto, K. Ino, S. Yoshikawa and the entire Design, Layout, CAD, Device, Evaluation, Test, and Process teams.

References:
[1] H. Tanaka, et al., "Bit Cost Scalable Technology with Punch and Plug Process for Ultra High Density Flash Memory," *IEEE Symp. VLSI Tech.*, pp. 14-15, 2007.
[2] J. W. Im, et al., "A 128Gb 3b/cell V-NAND flash memory with 1Gb/s I/O rate," *ISSCC*, pp. 130-131, 2015.
[3] D. Kang, et al., "256Gb 3b/Cell V-NAND Flash Memory with 48 Stacked WL layers," *ISSCC*, pp. 130-131, 2016.
[4] R. Yamashita, et al., "A 512Gb 3b/Cell Flash Memory on 64-Word-Line-Layer BiCS Technology," *ISSCC*, pp. 196-197, 2017.
[5] C. Kim, et al., "A 512Gb 3b/cell 64-Stacked WL 3D V-NAND Flash Memory," *ISSCC*, pp. 202-203, 2017.
[6] S. Lee, et al., "A 128Gb 2b/cell NAND Flash Memory in 14nm Technology with tPROG=640us and 800MB/s I/O Rate," *ISSCC*, pp. 138-139, 2016.



Technology:
96-WL-layer 3D Flash Memory
Organization:
3 Bits/Cell,
(16KB + ECC) / Page,
1152 Pages / Block,
(1822 + EXT) Blocks / Plane,
2 Planes
Throughput:
Read : 58us (ABL)
Program : 57MB/s
Power Supply:
Vcc : 2.30V to 3.6V
Vccq: 1.8V

Figure 20.1.1: Die micrograph and key features.

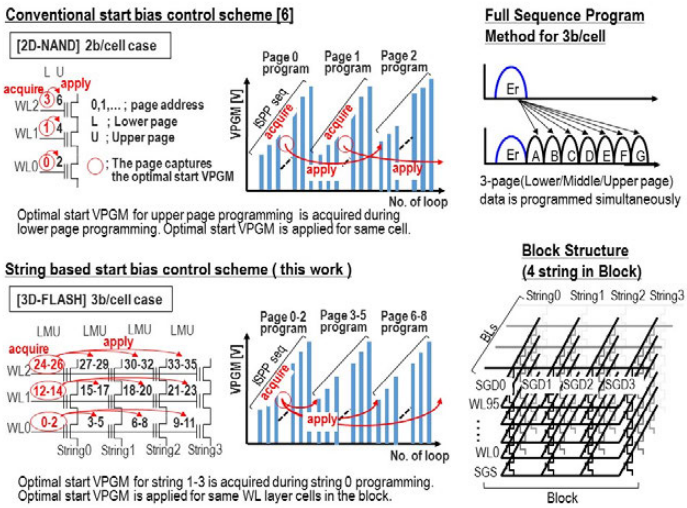


Figure 20.1.2: String-based start-bias control scheme (SSBC).

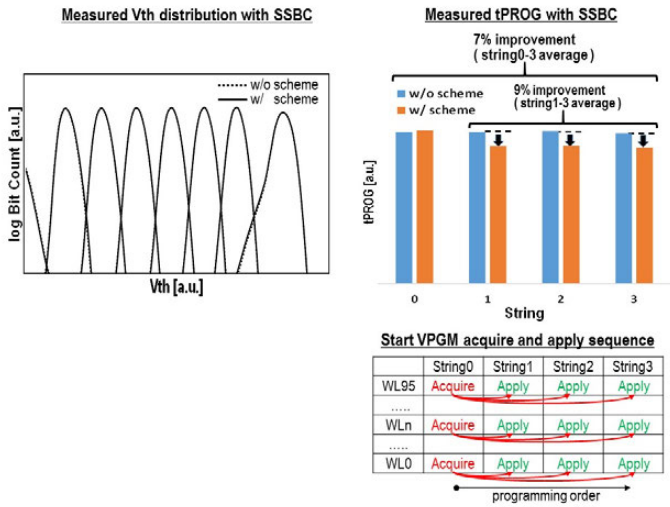


Figure 20.1.3: SSBC measurement results.

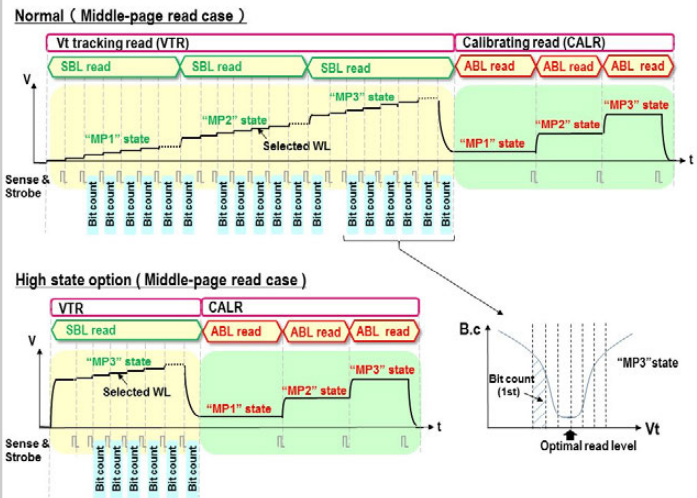


Figure 20.1.4: Smart V_t tracking read (SVTR) and high-state option.

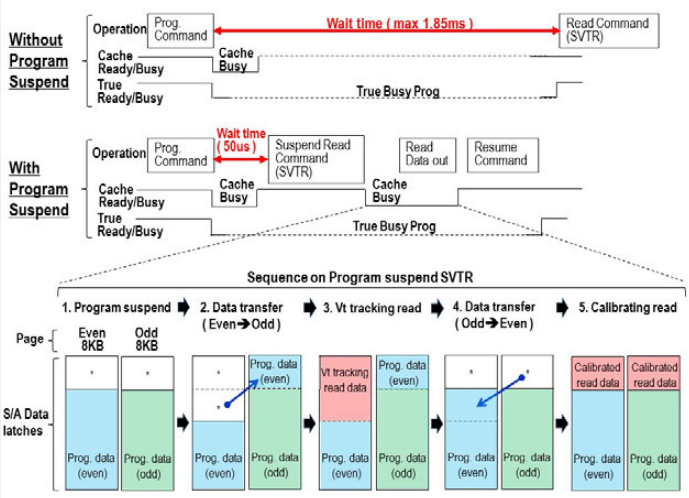


Figure 20.1.5: Program-suspend smart V_t-tracking read.

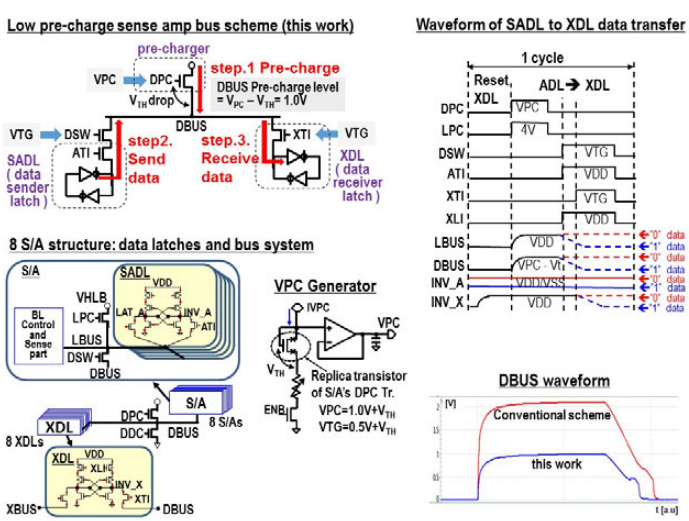


Figure 20.1.6: Low pre-charge sense amplifier bus scheme (LPSAB).

20.2 A Flash Memory Controller for 15µs Ultra-Low-Latency SSD Using High-Speed 3D NAND Flash with 3µs Read Time

Wooseong Cheong, Chanho Yoon, Seonghoon Woo, Kyuwook Han, Daehyun Kim, Chulseung Lee, Youra Choi, Shine Kim, Dongku Kang, Geunyeong Yu, Jaehong Kim, Jaechun Park, Ki-Whan Song, Ki-Tae Park, Sangyeun Cho, Hwaseok Oh, Daniel DG Lee, Jin-Hyeok Choi, Jaeheon Jeong

Samsung Electronics, Hwaseong, Korea

In a memory hierarchy, there are various classes of memory systems depending on the access latency. A typical memory hierarchy consists of a CPU cache, DRAM, and an SSD or HDD. The DRAM has an access latency of 100ns, while flash memory has a latency of about 50µs [1]. Recently, new non-volatile memories with latencies of less than 10µs, including PRAM, MRAM, and ReRAM [2], are getting attention for business-critical systems such as big-data analysis and storage caches. To meet the low latency requirements, a new type of NAND flash, Z-NAND, with a read time (t_R) of 3µs has also been introduced [3]. Figure 20.2.1 shows a feature comparison between Z-NAND and conventional 3D NAND [4,5]. The Z-NAND achieves a read time of 3µs, which is 15-20 times faster than conventional NAND. Write throughput reaches up to 160MB/s with a 100µs program time. To further minimize read latency, I/O circuit support a DDR interface for both ×8 and ×16 mode. To take full advantage of such low-latency memory devices, reduction of memory access overhead is necessary. In this paper, we introduce an NVMe SSD controller which leverages the advantages of the low-latency NAND and enables the reduction of total memory access time, thereby minimizing overall system latency.

Figure 20.2.2 defines the read latency of a typical SSD operation: from the time a host system issues a read command to the time data arrives at the host. The most common unit of data transfer in an NVMe SSD is 4kB, which is defined as a data chunk in this paper. t_{media} is a storage latency composed of t_R and t_{DMA}. t_{DMA} is the data transfer time of a NAND flash for a 4kB chunk. A typical value for t_{DMA} is about 8µs at a NAND I/O speed of 667Mb/s.

When the conventional 3D NAND has a t_R of 45-60µs, t_{media} is the most critical component of the system latency, and t_{DMA} is relatively small compared with t_R. On the contrary, when t_R is reduced to 3µs, t_{DMA} is a more dominant factor of t_{media}. We propose two techniques to reduce t_{DMA} in this paper, and achieve a total-storage read latency of less than 20µs.

A contemporary SSD employs a multi-channel and a multi-way architecture to get high throughput and capacity. We propose a split DMA scheme that accesses two pages of different NANDs from different channels and transmits 2kB of data simultaneously as shown in Fig. 20.2.3. In a typical address mapping scheme for SSDs a physical page number (PPN) directly identifies one physical page in NAND. To manage two pages as one PPN for the split DMA scheme, we combine two NAND channels into a super channel. For example, if the first and the second NAND channels in an SSD compose a super channel, and a 4kB PPN indicates a P^{th} page in the first NAND channel, only 2kB of data are read from the P^{th} page in the first channel and the other 2kB data are read from P^{th} page in the connected second channel. The super channel operation reduces t_{DMA} by half, and still maintains the same map table size compared with a conventional scheme.

However, there might be issues when a bad block occurs in one channel of a super channel, or when the data from either channel arrives at different times. The split DMA management engine, in Fig. 20.2.4, is a circuit that adjusts data flow during a split DMA operation. The remap checker in the engine receives the physical block address (PBA) and determines whether each channel in the super channel has a bad block or not. If one of the two separate pages mapped to the same PPN belongs to a bad block, the split DMA management engine remaps the corresponding bad block to a reserved good block to minimize waste of media space, rather than discarding both blocks in the super channel. To resolve timing skew between channels, the split DMA management engine absorbs timing skew at the arrival time of the late data and packs the two data into one packet. In the proposed controller, the split DMA method can reduce t_{DMA} to 4µs at a NAND I/O speed of 667Mb/s.

An SSD with a multi-channel and a multi-way structure improves performance by processing multiple operations concurrently. Programming performance gains are achieved as the data chunk size increases, but I/O transfer time also increases. It takes 24µs for a 16kB data transfer. This long DMA yields some disadvantages in mixed pattern operation. For example, when a write DMA is in progress to way 0 on a specific channel, and a read is attempted on way 1 of the same channel, the read command can be issued only after waiting 24µs as described in Fig. 20.2.5 top. Then, read data can be transferred after t_R. To reduce this overhead for a low-latency response and better quality of service, we propose a suspend/resume DMA function. When a read operation is required during a DMA of the other way, the function suspends a write data DMA in progress and sends a read command immediately. The write DMA is resumed after issuing the read command. When the DMA is suspended, the instantaneous write data transfer count is stored in the controller to calculate the resume starting point after attending to the more urgent command.

The suspend/resume DMA technique allows read commands to be issued without waiting for the completion of a DMA from the other NAND. It enables parallel processing of read operations and reduces the overall response latency by increasing the activation ratio of the multi-way NAND. This is effective in increasing the quality of service, especially for a system having a heavily-mixed read/write workload, such as data center applications. However, there is a switching overhead for the suspend/resume operations. A protection time is set in front and at the end of each DMA time, and suspending is prohibited during the protection time to optimize the switching overhead and latency gain.

We compare a 4kB read latency and storage benchmarking score of our low-latency SSD and a recent PRAM based SSD [6] in Fig. 20.2.6. PCMark 8 storage benchmark indicates the performance of the SSDs with traces recorded form Adobe Creative Suite, Microsoft Office and a selection of popular games. From a user experience perspective, shown in Fig. 20.2.6 top, the performance of SSD using Z-NAND and the techniques presented in this paper is comparable to that of the PRAM based SSD. The middle of Fig. 20.2.6 shows the 4kB random read latency distribution at queue depth of 1 (QD1). The proposed SSD is compared with an SSD having a conventional controller and Z-NAND. An average read latency improvement of 48% is achieved by the proposed SSD controller. The average random read latency of the proposed SSD for 4kB QD1 is as low as 15µs. Other techniques, such as hardware automation in host interfacing, map search, and improving the busy signal detection method, etc., are also used in the proposed SSD to reduce $t_{controller}$ and $t_{transfer}$. These host interface and mapping techniques are also important to achieve a QD1 latency of 15µs, but are beyond the scope of this paper. The random read latency distribution for 4kB at QD16 is shown in Fig. 20.2.6 bottom. The QD16 read latency of the proposed SSD is shorter than a conventional controller SSD using Z-NAND by 39%, and a PRAM based SSD by 20%, on average. Figure 20.2.7 shows the die photo of the proposed NVMe SSD controller. It has a PCIe Gen3 4-lane host interface and an 8-channel NAND interface.

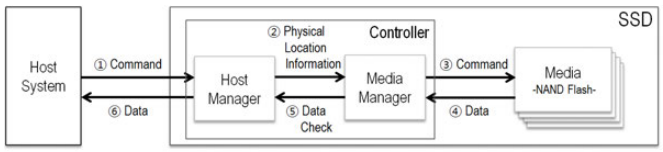
References:

[1] C. Zambelli, et al., "Phase change and magnetic memories for solid-state drive applications," *Proc. of IEEE*, vol. 105, no. 9, pp. 1790-1811, Sept. 2017.
[2] C. Matsui, et al., "Design of hybrid SSDs with storage class memory and NAND flash memory," *Proc. of IEEE*, vol. 105, no. 9, pp. 1812-1821, Sept. 2017.
[3] Y. Paik, "Developing extremely low-latency NVMe SSDs," *Flash Memory Summit*, Aug. 2017,
https://www.flashmemorysummit.com/English/Collaterals/Proceedings/2017/20170809_FA21_Paik.pdf
[4] D. Kang, et al., "256Gb 3b/Cell V-NAND Flash Memory with 48 Stacked WL Layers," *ISSCC*, pp. 130-131, 2016.
[5] C. Kim, et al., "A 512Gb 3b/cell 64-Stacked WL 3D V-NAND Flash Memory,'' *ISSCC*, pp. 202-203, 2017.
[6] "INTEL OPTANE SSD DC P4800X SERIES", accessed Nov. 2017,
https://www.intel.com/content/www/us/en/products/memory-storage/solid-state-drives/data-center-ssds/optane-dc-p4800x-series.html.

 978-1-5090-4940-0/18/$31.00

	Conventional NAND [4]	Conventional NAND [5]	Z-NAND
Technology	3D NAND 48 stacked word-line layer	3D NAND 64 stacked word-line layer	3D NAND 48 stacked word-line layer
t_R	45µs	60µs	3µs
t_{PROG}	660µs	700µs	100µs
Capacity	256Gb	512Gb	64Gb
Page Size	16kB/Page	16kB/Page	2kB/Page (×8 mode), 4kB/Page (×16 mode)

Figure 20.2.1: Comparison of Z-NAND and conventional 3D NAND memories.



Read latency = $t_{Host} + t_{Controller} + t_{Media} + t_{Transfer}$

Component	Description	Interval	Conventional SSD
t_{Host}	Time from host issuing command to controller's recognition	①	5 – 10µs
$t_{Controller}$	Time from decoding command to starting media operation	②⑤	20µs
t_{Media}	Time to read data of 4kB from media and transfer it to controller	③④	53 – 68µs
$t_{Transfer}$	Time to transfer data of 4kB from controller to host	⑥	5µs

$t_{Media} = t_R + t_{DMA}$

Component	Description	Conventional SSD	Proposed SSD
t_R	Time to transfer data of 4kB from cell array to register in media	45 – 60µs	3µs
t_{DMA}	Time to transfer read data from media to controller @ 667Mb/s NAND I/O	8µs	4µs

Figure 20.2.2: Definition of the read latency and t_{Media} for SSD.

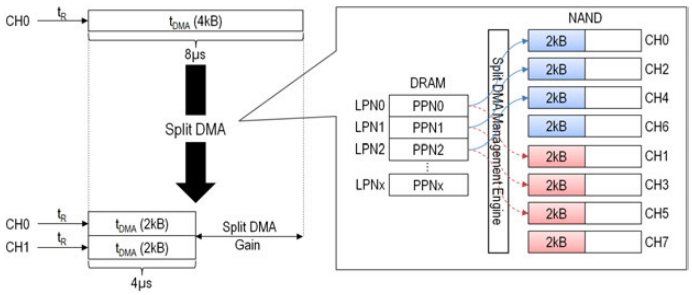


Figure 20.2.3: Split DMA scheme concept.

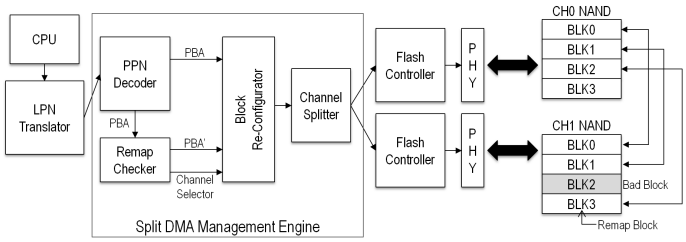


Figure 20.2.4: Bad-block management in split DMA.

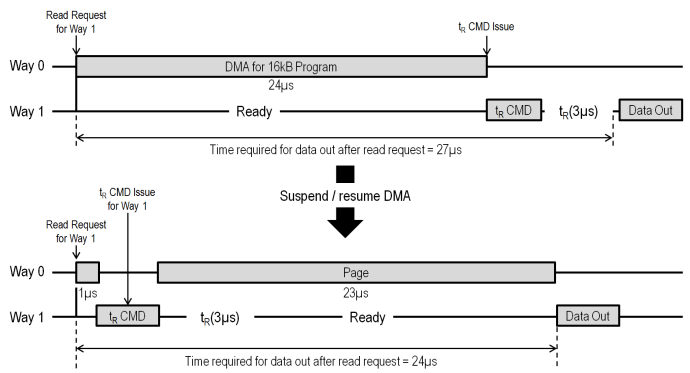


Figure 20.2.5: Suspend/resume DMA concept.

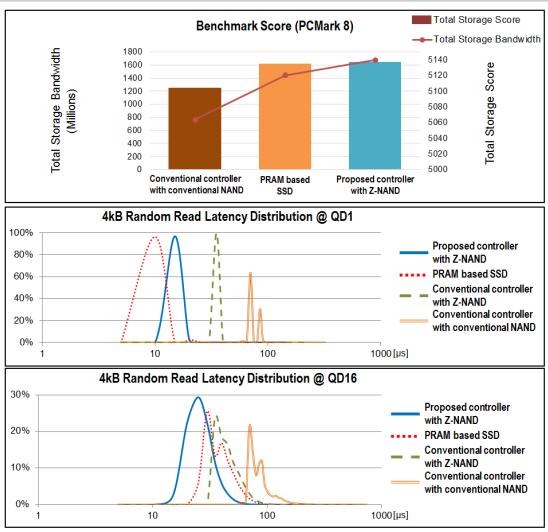


Figure 20.2.6: Evaluation results of PCMark8 overall score (top) and 4kB random read-latency distribution at QD1 (middle) and QD16 (bottom).

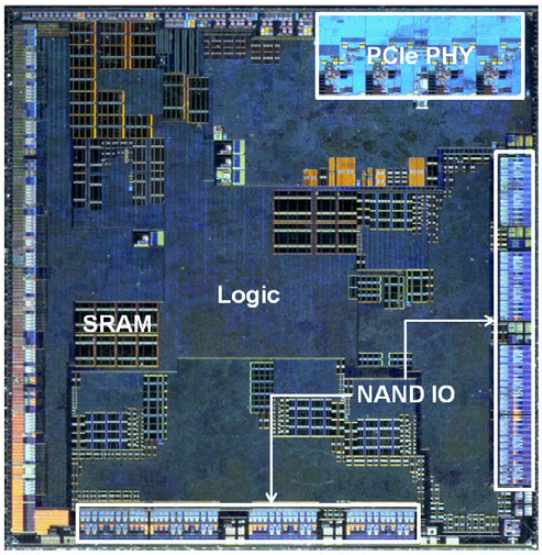


Figure 20.2.7: Photograph of proposed controller.

978-1-5090-4940-0/18/$31.00

20.3 A 1Tb 4b/Cell 64-Stacked-WL 3D NAND Flash Memory with 12MB/s Program Throughput

Seungjae Lee, Chulbum Kim, Minsu Kim, Sung-min Joe, Joonsuc Jang, Seungbum Kim, Kangbin Lee, Jisu Kim, Jiyoon Park, Han-Jun Lee, Minseok Kim, Seonyong Lee, SeonGeon Lee, Jinbae Bang, Dongjin Shin, Hwajun Jang, Deokwoo Lee, Nahyun Kim, Jonghoo Jo, Jonghoon Park, Sohyun Park, Youngsik Rho, Yongha Park, Ho-joon Kim, Cheon An Lee, Chungho Yu, Youngsun Min, Moosung Kim, Kyungmin Kim, Seunghyun Moon, Hyunjin Kim, Youngdon Choi, YoungHwan Ryu, Jinwon Choi, Minyeong Lee, Jungkwan Kim, Gyo Soo Choo, Jeong-Don Lim, Dae-Seok Byeon, Kiwhan Song, Ki-Tae Park, Kye-hyun Kyung

Samsung Electronics, Hwaseong, Korea

Since the first demonstration of a production quality three-dimensional (3D) stacked-word-line NAND Flash memory [1], the 3b/cell 3D NAND Flash memory has seen areal density increases of more than 50% per year due to the aggressive development of 3D-wordline-stacking technology. This trend has been consistent for the last three consecutive years [2-4], however the storage market still requires higher density for diverse digital applications. A 4b/cell technology is one promising solution to increase bit density [5]. In this paper, we propose a 4b/cell 3D NAND Flash memory with a 12MB/s program throughput. The chip achieves a 5.63Gb/mm^2 areal density, which is a 41.5% improvement as compared to a 3b/cell NAND Flash memory in the same 3D-NAND technology [4].

This paper presents a 1Tb 4b/cell NAND flash memory with 64 stacked WLs using 3D-NAND flash memory technology. A two-plane chip architecture is adopted and each block consists of 1024 pages with 16KB of storage. Other key parameters are summarized in Fig. 20.3.1. This paper will describe schemes for reducing the programming time and improving the read-window margin.

Figure 20.3.2(a) shows the cell array structure of the 3D NAND Flash memory. A block has four string select lines (SSL), while WLs and ground select lines (GSL) are shared within a block. The sharing of WLs allows the memory area to be reduced, by reducing the WL routing area required, but the severity of program disturbance increases. During programming a voltage (V_{pgm}) is applied to the selected strings, but the unselected strings are also disturbed since V_{pgm} is applied through the shared WLs. To decrease program disturbance an unselected string precharge operation (USP) was proposed [6]. However, a conventional USP has a negative effect on programming performance, especially for a 4b/cell NAND Flash memory, due to the large number of programming loops. To minimize write performance degradation, an adaptive USP scheme using both a fast and a conventional USP is proposed. The timing control comparison between conventional and fast USP are shown in Fig 20.3.2(b). For the fast USP, program BL is maintained at 0V while applying a positive bias to all of 4 SSLs in order to connect with BLs. This initializes the channel potentials of the unselected SSL strings without any time overhead. In early program loops, the program voltage is low, so the initialization of the channel potential by the fast USP is sufficient to overcome any disturbances. The programming performance improves by 8% when using an adaptive USP scheme.

Figure 20.3.3 shows the transition of cell V_{th} distribution of WL_n during a programming sequence. For the conventional high-speed program (HSP) scheme [2], the interference between neighboring WLs degrades the upper V_{th} distribution of WL_n cells while the WL_{n+1} pages are programmed. Furthermore, the program disturbance degrades the upper V_{th} distribution of WL_n cells in low V_{th} states while programming the pages of WL_n in other SSL strings. In addition, initial charge loss due to charge de-trapping in shallow traps broadens the under V_{th} distribution of WL_n cells. To overcome these degradation factors, the re-program scheme [5] is used in this work. First, cells in the WL_n of SSL_0 are coarsely programmed. In Fig. 20.3.3 the behaviors of cells in SSL_0 string are displayed during each stage of the program sequence. During the coarse program of WL_n in other SSLs and the coarse program of WL_{n+1} in all strings, the V_{th} distribution of WL_n becomes worse as shown in Fig. 20.3.3(b) because of program disturbance, initial charge loss, and WL interference. To eliminate these degradations, the fine program of WL_n cells in the SSL_0 string is finally performed as illustrated in Fig. 20.3.3(c). For a 3D NAND Flash memory using a charge trap cell, the re-program technique greatly suppresses initial charge loss by decreasing a proportion of shallow trap among total trapped charges, because the charges in a shallow trap are filtered during the duration between coarse program and fine program. Figure 20.3.4(a) displays the measured comparison results between using the proposed techniques and not. The relationship between an improvement to the V_{th} distribution and the bit error rate (BER) in Fig. 20.3.4(b); the BER is decreased by 83% due to using the proposed schemes.

In general, the number of programming loops is determined by the voltage required to program the slowest bits to the state with the highest Vth distribution. As increasing the number of programming loops due to the slowest bits, electric field across tunneling oxide increases during the program operation, which accelerates the degradation of tunneling oxide. And the amount of program disturbance, which is proportional to the number of programming loops and maximum value of program voltage, also increases. Furthermore, degradation of program performance is inevitable. In this work, a slow-bit bypass scheme is proposed, since usually only a few slow bits limit the number of overall programming loops. For example, the bit count of the cells to be programmed in the next loop is compared to the reference count of each state at the end of each state's verify operation. If the bit count of the processing loop is lower than the reference count, the program operation for the state ends, and the remaining the slow bits are un-programmed. For a P_{15} state with the highest V_{th} distribution, the reference count can be set larger than other states, which can reduce the number of programming loops and program disturbance. Figure 20.3.5(a) illustrates the above-mentioned operation, *A* denotes the reference cell count of normal states and *B* denotes the reference cell count of P_{15} state. Figure 20.3.5(b) shows the measured V_{th} distribution with and without the proposed scheme. Note that the program disturbance decreases without degrading the V_{th} distribution of P_{15} state.

Charge loss that is proportional to the retention time is one of the critical issues that impact the reliability of NAND Flash memory. For a 4b/cell operation, the programming of 16 states within a limited V_{th} window dramatically decreases read-window margin. Retrying a read can extend the lifetime of a NAND Flash memory, however a performance degradation due to repetitive read operations is inevitable. To minimize the time overhead due to retrying read, a fast read retry scheme is proposed. The concept of detecting charge loss due to retention is illustrated in Fig. 20.3.6(a). The read voltage for detecting retention (V_{DET_RET}) is below the default read voltage (V_{DEF}). Sensing with two different sensing times is equivalent to sensing with two different voltages (V_{DET_DET}, V_{DEF}), therefore the timing overhead for reading with two read levels is negligible. After counting the number of cells between the two voltages with a page buffer operation, the counting result (N_C in Fig. 20.3.6(b)) is compared to the count of reference (N_{TH}). If N_C is larger than N_{TH}, charge loss read errors have been detected. In this case, the read level is lowered according to the predefined look-up table (LUT) of each state's read level. On the other hand, if N_C is smaller than N_{TH} then the default read level is unchanged, since the error bits are correctable. A flow diagram for the proposed fast retry is shown in Fig. 20.3.6(b). Overall, the errors are corrected within two times of a read operation. A die photograph of the fabricated chip showing the chip architecture is shown in Fig. 20.3.7.

References:

[1] K.-T. Park, et al., "Three-Dimensional 128Gb MLC Vertical NAND Flash-Memory with 24-WL Stacked Layers and 50MB/s High-Speed Programming", *ISSCC*, pp. 334-335, 2014.

[2] J.-W. Im, et al., "A 128Gb 3b/Cell V-NAND Flash Memory with 1Gb/s I/O Rate," *ISSCC*, pp. 130-131, 2015.

[3] D. Kang, et al., "256Gb 3b/Cell V-NAND Flash Memory with 48 Stacked WL Layers," *ISSCC*, pp. 130-131, 2016.

[4] C. Kim, et al., "A 512Gb 3b/cell 64-Stacked WL 3D-NAND Fla," *ISSCC*, pp. 202-203, 2017.

[5] N. Shibata, et al., "A 70nm 16Gb 16-level-cell NAND flash memory," *JSSC*, vol. 43, no. 4, pp. 929-937, Apr. 2008.

[6] R. Yamashita, et al., "A 512Gb 3b/cell flash memory on 64-word-line-layer BiCS technology," *ISSCC*, pp. 196-197, 2017.

 978-1-5090-4940-0/18/$31.00

Bits per cell	4
Density	1Tb
Areal Density	5.63Gb/mm^2
Technology	3D NAND with 64 stacked WL layer
Organization	2-plane, 1024Pages/Block, 16KB/Page
I/O Bandwidth	Max. 1Gb/s
VCCQ	1.2V
tBERS	3.5ms (Typ.)
Program throughput	12MB/s
tR	140us

Figure 20.3.1: Key parameter table for the 1Tb 4b/cell 3D-NAND Flash chip.

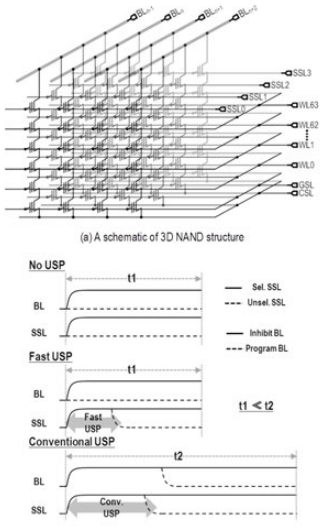


Figure 20.3.2: Introduction to unselect string pre-charge (USP) operation.

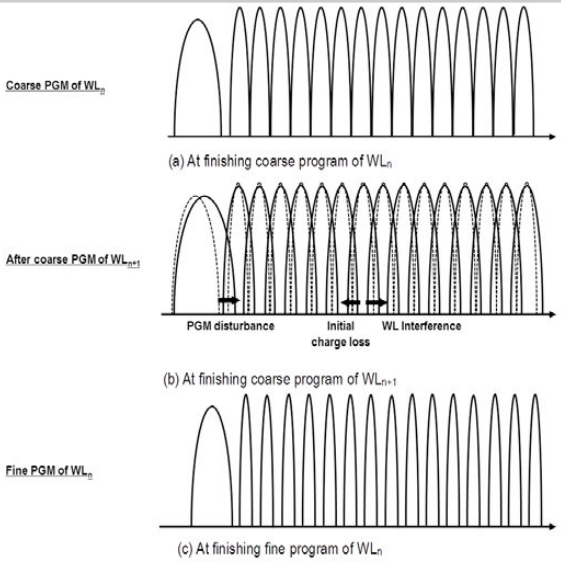


Figure 20.3.3: Transition of V_{th} distribution during re-programing.

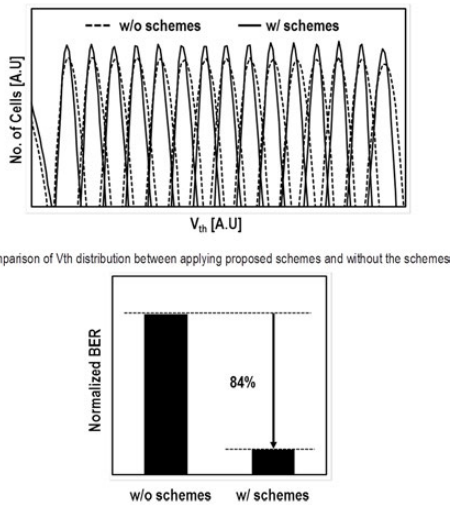


Figure 20.3.4: Measurement results of the proposed fast USP and re-program schemes.

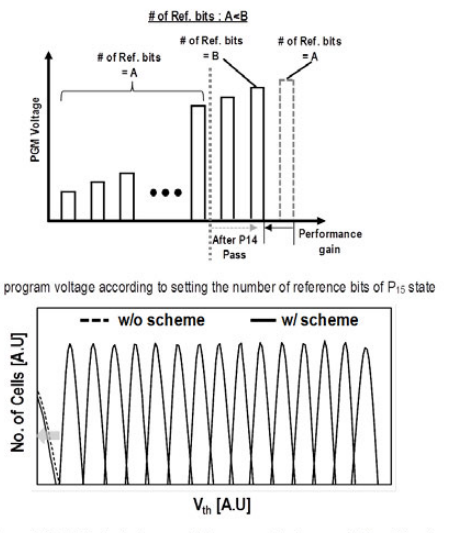


Figure 20.3.5: Illustration of slow bit by-pass scheme.

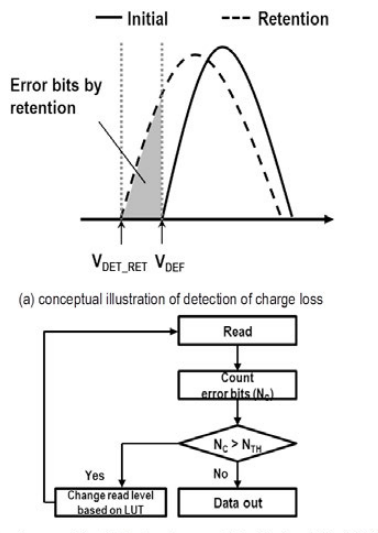


Figure 20.3.6: Illustration of fast-read retry scheme.

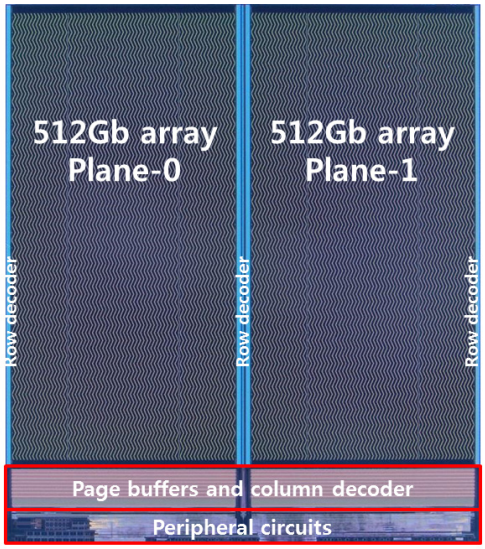


Figure 20.3.7: Die photograph.

978-1-5090-4940-0/18/$31.00

Session 21 Overview:
Extending Silicon and its Applications
TECHNOLOGY DIRECTIONS SUBCOMMITTEE

Session Chair:
Jan Genoe
IMEC, Leuven, Belgium

Associate Chair:
Frederic Gianesello,
STMicroelectronics, Crolles, France

Subcommittee Chair: ***Makoto Nagata,*** Kobe University, Japan

This session includes six papers from the Technology Directions subcommittee at ISSCC 2018. The first two papers present advances in mixed signal processing for machine learning, the third paper describe a 32GHz mechanical resonator achieved for the first time in 14nm FinFET technology, the fourth paper reviews a 10Gb/s Si Photonics transceiver targeting 1Tb/s/mm² die-to-die communication, the fifth paper describes an innovative sensor to detect laser fault injection attack on a cryptographic core in order to avoid any information exposure and the final paper presents an injection-locked VCO array targeting ESR application.

 978-1-5090-4940-0/18/$31.00

10:15 AM

21.1 Mixed-Signal Programmable Non-Linear Interface for Resource-Efficient Multi-Sensor Analytics

K. Badami, KU Leuven, Leuven, Belgium

In Paper 21.1, KU Leuven University presents a highly-configurable non-linear mixed-signal interface for sensor systems enabling energy-efficient and application-tunable translation from analog to non-linear digital. The work achieves 90nm IC proof-of-concept demonstrating up to 33kS/s operation and 50% data-reduction while preserving 9.5b resolution in the region-of-interest

10:45 AM

21.2 A 1μW Voice Activity Detector Using Analog Feature Extraction and Digital Deep Neural Network

M. Yang, Columbia University, New York, NY

In Paper 21.2, Columbia University presents a voice activity detector using analog signal processing for feature extraction and digital deep neural network for classification. Ultra-low power of 1μW is achieved while delivering mean speech/non-speech hit-rate of 84.4%/85.4% for 10dB SNR speech with restaurant noise over 10 dies.

11:00 AM

21.3 32GHz Resonant-Fin Transistors in 14nm FinFET Technology

B. Bahr, Massachusetts Institute of Technology, now at Kilby Labs, Texas Instruments, Dallas, TX

In Paper 21.3, MIT, Purdue University and Globalfoundries present a 32GHz Resonant-Fin Transistors in 14nm FinFET Technology. A periodic array of fins forms a slow-wave structure that achieves confinement of mechanical vibrations in the FEOL.

11:15 AM

21.4 A 10Gb/s Si-Photonic Transceiver with 150μW 120μs-Lock-Time Digitally Supervised Analog Microring Wavelength Stabilization for 1Tb/s/mm² Die-to-Die Optical Networks

Y. Thonnart, CEA-LETI-MINATEC, Grenoble, France

In Paper 21.4, CEA-LETI and ST Microelectronics present a 10Gb/s Si-photonic transceiver for 1Tb/s/mm² die-to-die optical networks. A 10Gb/s 150μW Si-photonic Transceiver with 120μs lock-time has been obtained through digitally-supervised analog micro-ring wavelength stabilization.

11:45 AM

21.5 A 286F²/Cell Distributed Bulk-Current Sensor and Secure Flush Code Eraser Against Laser Fault Injection Attack

K. Matsuda, Kobe University, Kobe, Japan

In Paper 21.5, Kobe University, The University of Electro-Communications and Nara Advanced Institute of Technology present a 286F²/Cell distributed bulk-current sensor that protects against laser fault injection attacks by securely flush erasing the code. The erasing is in the nanosecond order and the overall overhead is only 28%.

12:00 PM

21.6 An 8-Channel 13GHz ESR-on-a-Chip Injection-locked VCO-array achieving 200μM-Concentration Sensitivity

A. Chu, University of Stuttgart, Stuttgart, Germany

In Paper 21.6, The University of Stuttgart, Ulm University and Helmholtz-Zentrum Berlin für Materialien und Energie present an array of eight injection locked VCOs for ESR spectroscopy operating around 13GHz. By using injection locking, this circuit enables a 10-fold increase in sensitive volume and a 10-fold improvement in noise frequency floor.

21.1 Mixed-Signal Programmable Non-Linear Interface for Resource-Efficient Multi-Sensor Analytics

Komail Badami, Juan-Carlos Pena Ramos, Steven Lauwereins, Marian Verhelst

KU Leuven, Leuven, Belgium

Tremendous progress has been made in reducing ADC power-consumption, yet, in many portable always-awake and multi-sensor systems, the power consumption is dominated by digital backend processing [1] for feature-computation and classification. Recent Analog-to-Information based innovations (see Fig. 21.1.1) have attempted to alleviate this bottleneck by reducing the amount of data streaming into the digital domain: (a) by extracting sensory features in the analog domain [2] and digitizing these instead of the raw-data; and (b) by compressing the data through an analog non-linearity in order to reduce the required digitization word-length [3-5]. Yet, both approaches suffer from limited applicability and design reuse problems, due to (a) their need for highly application-specific building blocks which are not portable across different sensor-signals in multi-sensor platforms; and (b) their need for complex, performance-sensitive analog building blocks which are not portable across silicon technologies. Overcoming the above shortcomings, this work reports a highly programmable non-linear interface that synergistically combines a digitally-computed, application-tunable non-linearity with a 10b binary DAC in an iterative mixed-signal loop (see Fig. 21.1.1 bottom) to enable a compressive analog to non-linear digital transfer-curve. Such configurable non-linear transfer-curve has wide applicability in multi-sensor analytics for feature-extraction, signal-emphasis/-de-emphasis, signal-correction, etc. A 90nm CMOS proof-of-concept illustrates this versatility with 2 very different application mappings, demonstrating (a) Compressive classification: 2× improvement in rms-error for heart-beat classification from muscle noise corrupted ECG signals, along with 50% backend data reduction; and (b) Analog impairment correction : up-to 20dB distortion correction for analog impairments in the sensory chain.

The wide versatility of the mixed-signal interface is enabled by a highly-flexible analog to non-linear-digital transfer-curve (see Fig. 21.1.2). This non-linear transfer-curve enhances the resolution in the input signal's region-of-interest while allowing for a lower resolution outside this region. Such configurable non-uniform representation creates an emphasis on the information content and enables output data-compression. This non-linearity is realized through 3 programmable parameters, (illustrated in Fig. 21.1.2) which control the shape of the non-linearity: (a) ε control the finest resolution present in the transfer-curve; (b) α dictates the position of this maximum resolution region; and (c) n_{it} defines the width of this region. The latter is directly linked to the digital output word-length and hence the data compression ratio. In order to realize this flexible non-linearity efficiently and without large memory (look-up tables) or area overheads, a new iterative mixed-signal digitization scheme is derived, which non-linearly maps an analog input (v_{in}) to its digital output ($b\,[1{:}n_{it}]$) as:

$$v_{in} = A + B\prod_{i=1}^{n_{it}} W(i)^{b(i)} \quad (1)$$

The parameters A, B and W are precomputed from ε, α, n_{it}, using the formulas shown on Fig. 21.1.3. Based on values of ε, α and n_{it}, Eq. (1) enables a wide range of transfer-curves such as logarithmic, exponential, tangent-hyperbolic/its-inverse, etc. (see measurements in Fig. 21.1.4). As an example, for $\alpha \to 0$ and $\varepsilon << 1$, Eq. (1) can be simplified to a logarithmic transfer-curve (similar to μ / A-Law) given as:

$$\ln(v_{in})/\ln(\varepsilon) \approx k(1 - \textstyle\sum_{i=1}^{n_{it}} 2^{-i}\, b(i)) \quad (2)$$

The iterative search to derive the digital outputs ($b\,[1{:}n_{it}]$) is efficiently implemented through a mixed-signal loop, as illustrated in Fig. 21.1.3, consisting of a standard 10b capacitive binary DAC ($C_{unit} = 10fF$), a dynamic comparator and configurable digital logic. The latter controls the binary DAC with a programmable non-linear mapping scheme as follows: In every iteration i, the digital logic operates in closed loop with the analog to provide the binary threshold for bit $b(i)$ realizing the desired non-linearity (see Fig. 21.1.3): The first iteration compares the sampled analog input (v_{in}) with the parameter α to select the appropriate set of A, B, W as shown in the table in Fig. 21.1.3. In each remaining iteration, i, bit $b(i)$ is first set to '1' and the DAC control signals c [9:0] are digitally updated as follows:

$$c_i[9{:}0] = A + B\prod_{j=1}^{i-1} W(j)^{b(j)} W(i)^{b(i)} \quad (3)$$

The subsequent comparator decision then sets or resets $b(i)$. This iterative approximation enables the DAC output to approach the sampled v_{in} starting from α, with n_{it} steps and a non-linear step-size controlled by ε.

The 90nm CMOS fully dynamic non-linear interface can operate up-to 33kS/s from a 1.2V supply with an input signal swing of 1.8Vppd. Measured transfer-characteristics in Fig. 21.1.4 shows the system's versatility through a logarithmic (α = 0, Fig. 21.1.4 top-left) and exponential (α = 1, Fig. 21.1.4 top-right) non-linearity for a several settings of n_{it} and ε. The parameter n_{it} can be set from 5 to 9 (see Fig. 21.1.4 top-left) and enables a trade-off between the data-compression ratio and the width of the high-resolution region, while the parameter ε influences the steepness of the curve (see Fig. 21.1.4 top-right), and hence the maximum resolution of the transfer-curve. Different digital non-linearities can also be combined on both sides of the transfer-function ($v_{in} < / > \alpha$), to enable a hyperbolic tangent (see Fig. 21.1.4 bottom-left) and its inverse (see Fig. 21.1.4 bottom-right) transfer-function, which allows one to focus the high resolution in an application-required region-of-signal-interest through tuning α. The total mixed-signal system power-consumption for the different operation modes and multiple sampling frequencies is highlighted in Fig. 21.1.4. Fig. 21.1.5 illustrates the quantization error for the non-linear transfer-curves of Fig. 21.1.4. The quantization error has a DAC limited 10b minimum at $v_{in} = \alpha$ and gradually increases away from α to enable a non-linear compressive transfer-characteristic. While only a few combinations of ε, α and n_{it} are shown in Figs. 21.1.4 and 21.1.5, these parameters can be independently tuned to benefit a wide range of applications.

Two applications are mapped to illustrate the system's versatility across sensor interfaces and tasks:
(a) Signal classification using non-linearly compressed data is illustrated through heart-beat detection from muscle noise corrupted ECG signals (Fig. 21.1.6 top). The setup uses MIT-BIH Noise Stress Test database [6], which contains noisy ECG signals with SNR from 24dB to -6dB. Measured results show that the non-linear mixed-signal mapping with α = 0.8, ε = 0.05 and n_{it} = 5 emphasizes the peaks while suppressing the muscle noise. Compared to a classical linear sensor interface, this reduces the rms error in heart-beat detection by 2× while also reducing the required digital word-length by 50% which decreases the dominant backend digital power by ~50%.

(b) Analog impairment correction of sensory signals is illustrated through front-end amplifier distortion correction (Fig. 21.1.6 bottom). Here, the non-linear interface is used in the inverse-tangent-hyperbolic mode with ε = 1.1 and n_{it} = 8 to compensate the tangent-hyperbolic non-linearity of efficient open-loop amplification. Measured results show that the mixed-signal non-linear interface corrects the amplifier distortion from as high as 5% THD (SNDR 25dB) to 0.5% THD (SNDR 46dB) thus eliminating the need for power expensive backend digital calibration/correction.

Figure 21.1.7 highlights the different sections on the die micrograph and compares the reported mixed-signal interface to the state-of-the-art designs. The proposed design uses an iteratively computed digital non-linearity in a mixed-signal loop and hence is the first one to allow complete programmability to match its transfer-characteristics with the varying requirements of multi-sensor applications. This enables a true analog-to-information conversion, while reducing the digital word-length up to 50% at comparable power efficiency with the existing non-flexible implementations. While this work reports two application mappings, the highly configurable non-linear interface is widely applicable to many sensory analytics systems, such as gait/posture-detection from accelerometers, stress/fatigue detection in machine monitoring, etc.

Acknowledgements:
Authors would like to thank Wannes Meert for helpful discussions and IMEC-IC link for backend support during tape-out.

References:
[1] M. Yip, et al., "A fully-implantable cochlear implant SoC with piezoelectric middle-ear sensor and energy-efficient stimulation in 0.18μm HVCMOS," *ISSCC Dig. Tech. Papers*, pp. 312-313, Feb. 2014.
[2] K. Badami, et al., "Context-aware hierarchical information-sensing in a 6μW 90nm CMOS voice activity detector," *ISSCC Dig. Tech. Papers*, pp. 430-431, Feb. 2015.
[3] M. Judy, et al., "Nonlinear Signal-Specific ADC for Efficient Neural Recording in Brain-Machine Interfaces," *IEEE Trans. on Biomedical Circuits and Systems*, vol. 8, no. 3, pp. 371-381, June 2014.
[4] J. Lee, et al., "A 2.5 mW 80 dB DR 36 dB SNDR 22 MS/s Logarithmic Pipeline ADC," *IEEE Journal of Solid-State Circuits*, vol. 44, no. 10, pp. 2755-2765, Oct. 2009.
[5] Ji-Jon Sit and R. Sarpeshkar, "A micropower logarithmic A/D with offset and temperature compensation," *IEEE Journal of Solid-State Circuits*, vol. 39, no. 2, pp. 308-319, Feb. 2004.
[6] G. B. Moody, et al., "A noise stress test for arrhythmia detectors," *Computers in Cardiology*, vol. 11, pp. 381-384, 1984.

 978-1-5090-4940-0/18/$31.00