A flexible column parallel successive-approximation ADC for hybrid neuromorphic computing

Philipp Dauer, Milena Czierlinski, Sebastian Billaudelle, Andreas Grübl and Johannes Schemmel
Kirchhoff-Institute for Physics, Heidelberg University, Germany
Email: philipp.dauer@kip.uni-heidelberg.de

***Abstract*—This work presents a concept and components for a Nyquist dual mode charge-redistribution SAR ADC in a 65 nm-process. Its fast mode (62.5 MS s^{-1}) offers 7 bit resolution using an interleaving technique and its precise mode 8 bit resolution at 12.5 MS s^{-1}. The ADC is specialized on the requirements of a parallel interface between analog and digital computing blocks in the neuromorphic BrainScaleS-2 (BSS-2) system. Furthermore, it offers a small area footprint of around 1500 µm^2 and low power consumption. It targets at two main applications: A fast result readout for analog vector-matrix-multiplications as well as membrane potentials and other plasticity related observable in spiking neural networks (SNNs).**

***Index Terms*—analog-to-digital conversion, SAR, neuromorphic**

I. Introduction

Analog-to-digital converters (ADCs) are an indispensable building block in mixed-signal systems as they transfer information generated by the nature of analog circuits and stored in the magnitude of physical quantities to the digital domain. The BSS-2 system implements SNNs by emulating accelerated neuron and synapse behavior in analog circuits [1]. The mixed-signal application-specific integrated circuit (ASIC) is based on a low-power 65 nm standard complementary metal-oxide-semiconductor (CMOS) technology and offers multiple domain crossings that make analog signals accessible to its digital periphery (fig. 1). One of these, the column-parallel ADC (CADC), implements an interface between the analog network core and the single instruction, multiple data (SIMD) extension of the embedded microprocessor.

A SNN on BSS-2 works as follows: Each synapse in the synapse array implements a 6 bit current-mode digital-to-analog converter (DAC) which is gated through horizontally distributed addresses and corresponding control pulses. The control pulses symbolize pre-synaptic spikes which are forwarded by the synapses modulated with certain weights. The synapses are column-wise connected to 256 individually configurable neurons which evolve according to either the leaky integrate-and-fire (LIF) model or adaptive exponential integrate-and-fire (AdEx) dynamics [2]. Threshold comparators in the neurons realize a firing mechanism to subsequently trigger spike events. One learning-method for SNNs, spike timing dependent plasticity (STDP), is based on temporal correlations between pre- and post-synaptic spiking activity. The digitization of these correlations to compute synapse weight updates in the microprocessor, is the original purpose of the CADC [3].

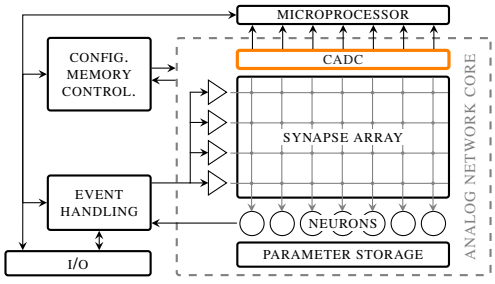

Figure 1. Block-level schematic of the BSS-2's ASIC. This paper presents a redesign of the highlighted CADC.

The system can be reconfigured to operate as an energy efficient analog vector-matrix-multiplication accelerator [4]. In this mode, the synaptic current pulse widths are modulated according to pre-synaptic activation vectors and the neural network core, hence, performs vector-matrix multiplications. The result vectors are represented in the charge accumulated in each column. A following voltage readout is then left to the CADC. However, the current design of the CADC, an 8 bit single-slope ADC, is too slow to yield competitive performance. While the synapses allow a vector processing rate of 125 MHz, the ADC limits the repetition rate to around 1 MHz [5]. A redesigned CADC concept should provide an increased sampling frequency – even at the cost of a slightly decreased resolution – to overcome this bottleneck. Additional constraints are given by general area limitations, the pitch of the synapse array, and the overall power budget. The ADC should, furthermore, provide 8 bit resolution covering an input range of 1.2 V.

II. A flexible charge redistribution concept

Successive approximation is one of the most versatile and popular methods in ADC design [6]–[8]. A successive-approximation register (SAR) ADC performs a binary search on the magnitude of a sampled physical quantity using a DAC, a comparator and a digital controller. Thereby, they promise a linear speed-to-resolution ratio at very low power consumption. However, in naive implementations of the commonly used capacitor-based DAC architectures the number of required unit

979-8-3503-0320-9/23/$31.00 ©2023 IEEE

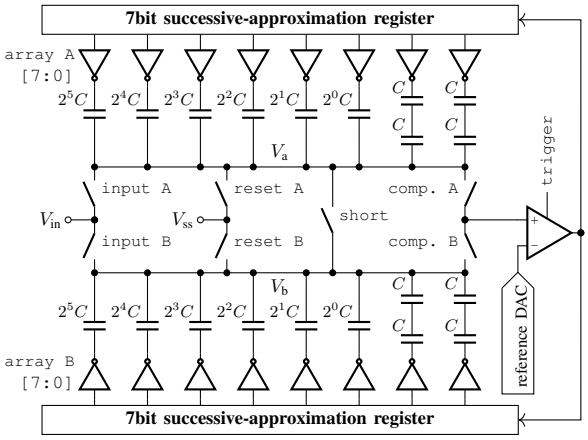

Figure 2. Schematic of the proposed SAR ADC implementing a dual-mode charge redistribution concept. In its fast mode, the two 7 bit halves operate as interleaved ADCs sharing one common comparator and in its precise mode they are joined to one common 8-bit ADC.

capacitors grows exponentially with the required resolution.

We developed a dual-mode SAR concept (fig. 2) that allows choosing between a fast interleaved mode of two 7 bit ADCs "A"/"B" and a joint operation mode yielding 8 bit resolution. Both modes make use of the charge redistribution concept. While one plate of the capacitors is connected to V_a/V_b, the other terminal can be clamped either to ground or supply. This switching is realized by buffered logic levels of `array A`/`array B`. Modifying `array A`/`array B` induces a voltage change on V_a/V_b with a magnitude determined by the relative size of the switched capacitor.

A. fast mode

In the fast mode, the two identical 7-bit DACs alternatively sample and convert the input signal. During the sampling phase of "A", `array A[7]` is connected to the supply node while all other capacitors of "A" are connected to the ground node. With this, the charge of the input signal is equally referred to both rails. While converting, `input A` is open and the comparator is connected and enabled instead. The approximation follows a try-and-reject principle: One adds charge by switching the next bit to supply and reverts this change if the potential was increased beyond half the supply voltage. The path through the decision tree corresponds to the output code. However, the impact of `array A[0]` is identical to the impact of `array A[1]` which limits the resolution to 7 bit. After each decision the charge on V_a is cleared using the `reset A` switch.

B. precise mode

In the precise mode, initially all capacitors of "A" connect to supply while the capacitors of "B" connect to ground. Thus, the equal referring of the charge is not only fulfilled globally, but in every bit and the "try"-phases during approximation are omitted. In each step $i \in [0, 7]$, the control signals `array A[7-`i`]` and `array B[7-`i`]` switch according to comparator's last decision. Since every decision gains one bit of information, we end up with a resolution of 8 bit.

C. implementation

The capacitors are realized as metal-oxide-metal (MOM) capacitors and make use of the inherent capacitance within the metal stack. The capacitor array is build from unit capacitors of 2.36 fF each. Using serial circuits, we halve the unit-capacitor to achieve a compromise between area consumption and capacitor mismatch. The capacitors cover an area of $1000\,\mu m^2$ which at the same time is the total footprint of all analog components as comparator and reference generator are placed underneath the capacitive array. Theoretically, both modes allow a full swing input and lead to perfect conversions. In reality, unintended additional parasitic capacitances reduce the dynamic range, capacitor mismatch limits the linearity and decision offsets yield a static shift.

III. SUCCESSIVE-APPROXIMATION REGISTER

The SAR implements the actual binary search algorithm by taking the comparator decision, controlling the DAC's configuration, and providing the digital output code. In fact, the SAR is a finite state machine (FSM) that due to the two modes becomes more involved. Selecting the optimal state-space partitioning and the choice of representation bits makes the FSM fast, small, and robust. However, with improperly chosen states representations, it becomes hard to meet the timing constraints and avoid glitches.

For our ADC, we suggest separate logic for each 7-bit SAR. The state of each DAC is directly represented in eight flip-flops (FFs). Each comparator decision can be written directly to one of these FFs. Four additional FFs encode a counter functionality belonging to the control logic of the state machine. By using an adopted gray encoding, we avoid glitches in this Moore machine output that for example drives the switches in fig. 2.

Synthesis results indicate a required standard cell area of $482\,\mu m^2$. For one conversion, transistor-level simulations claim for about 1.6 pJ in its fast mode and about 3.0 pJ in its precise mode. A static timing analysis of the synthesized design estimates a maximum clock frequency of 1 GHz at typical conditions.

IV. COMPARATOR

The comparator is intended to compare the potential on the capacitor array to half of the supply voltage. As we focus on the aspects of accuracy, speed, and energy, sense amplifiers are up-and-coming comparator core candidates. Originally developed for memory readout, they are also used in multiple SAR ADC designs [6]–[8]. Their fast regenerative differential pair concept eliminates the need for static bias currents. One option to increase the inherently limited headroom in the differential pair and thus the dynamic range is the use of a second tail separating the differential input stage from the latch stage [9].

A schematic of a double-tail sense amplifier (DTSA) is shown in fig. 3. Nodes V_{d+} and V_{d-} function as implicit

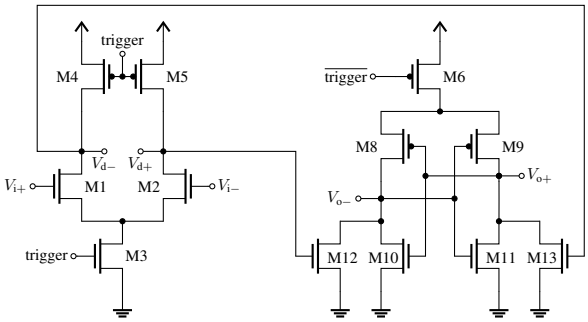

Figure 3. Original concept of a DTSA [9] used as core circuit in the comparator.

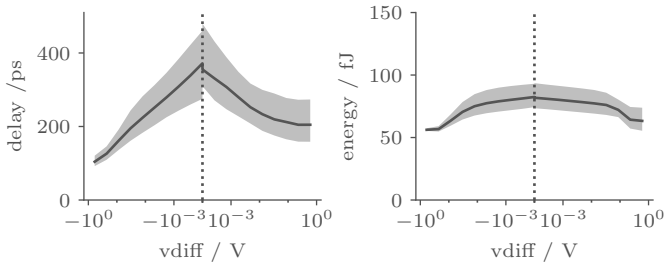

Figure 4. Delay and energy consumption of the DTSA estimated from a post-layout simulation at a nominal supply voltage of 1.2 V and a temperature of 27 °C. In each case, the common mode was set to 0.6 V. The shaded area marks minimal and maximal value over all process corner, the line denotes the typical case.

capacitors which are pre-charged by the transistors M4 and M5. When enabled by the trigger signal at M3, V_{d+} and V_{d-} discharge via the M1 and M2 with input voltage dependent slopes. Two cross-coupled inverters form a latch stage, that is, during the decision phase, enabled by the $\overline{\text{trigger}}$ signal. Initially, both sides of the cross-coupled inverters are pulled to ground, as M12 and M13 are conducting. Caused by the different discharging times of V_{d+} and V_{d-}, either M12 or M13 starts to conduct earlier, pushing the corresponding side of the latch to high. The positive feedback in the cross-coupled inverters then ensures a full-swing signal.

We estimated the DTSA performance in post-layout simulation. Figure 4 illustrates the extracted delay time and energy consumption over the different corners. Compared to the original results of [9], [10], our DTSA is slower but competitive in terms of its area and energy footprint. The maximum power consumption appear in the slow-fast corner (91 fJ conv^{-1}); in the typical corner the DTSA requires 80 fJ conv^{-1} at 0.7 mV voltage difference. The maximum required delay time at 0.7 mV input voltage difference occurs in the slow-slow corner (388 ps conv^{-1}); in the typical corner we simulate a delay time of around 307 ps conv^{-1}. However, the comparator is fast enough to yield a correct decision within half of a clock cycle at a frequency of 1 GHz. The DTSA covers an area of 67.6 µm^2 while it only uses up to the second metal layer to fit below the MOM capacitor array.

The reference voltage $V_{i-} = V_{dd}/2$ is generated by a capacitive voltage divider based on MOS capacitors realized in the form of transistor gates. To adjust the reference voltage and

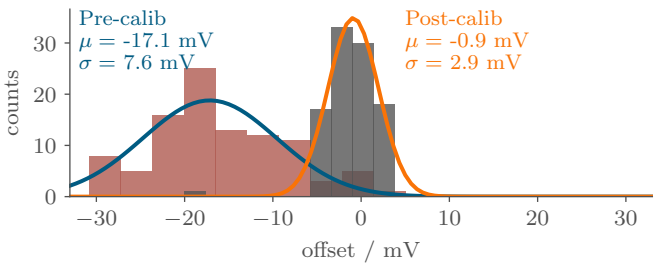

Figure 5. Histogram of the simulated comparator offset evaluated using Monte Carlo methods and its compensation by the capacitive reference DAC. The solid lines annotate a Gaussian fit. The systematic shift in the pre-calibrated comparator offsets comes from an input impedance mismatch between V_{i+} and V_{i-}.

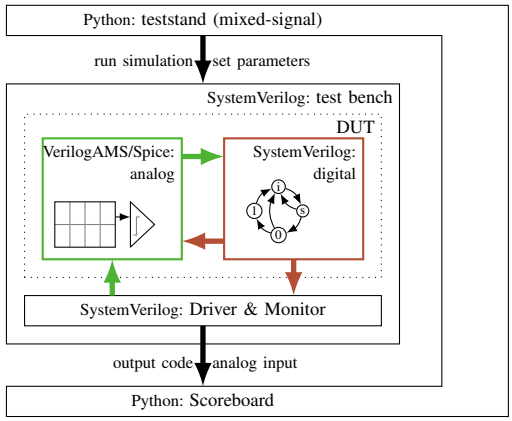

Figure 6. The custom mixed-signal verification framework is controlled by Python using an extended teststand module. Teststand wraps Cadence® simulators.

to calibrate for the comparator's input offset, the capacitor size can be adjusted with 5 bit resolution by selectively attaching individual capacitors. An intended static offset in the capacitance reduce its gain to 3.3 mV LSB^{-1}. Due to leakage currents through switches and capacitors gate oxide, the capacitors in the voltage divider discharge over time. To mitigate this issue, the ADC relies on two interleaved reference circuits, allowing a periodic refresh while still ensuring uninterrupted operation. We simulated the reference generators together with the DTSA to demonstrate the offset compensation (fig. 5). Each reference circuit covers around 220 µm^2.

V. Simulation and Verification

To verify the ADC performance, we use the mixed-signal simulation framework illustrated in Figure 6. It bases on our custom Python module `teststand` [1], which interfaces the Cadence® Spectre® simulator. We extended it to cover mixed-signal designs as well. Controlling simulations from Python enables programmatic parametrization and data analysis.

The results shown in figs. 7 and 8 are based on post-layout simulations for the analog part of the circuit including a transistor-level description of the SAR logic. In this first attempt, we sweep the desired dynamic input range and extracted the observed code transitions. Codes, where the code transition was not sharp but accompanied with some jitter are annotated by orange arrows. On one hand, this jitter

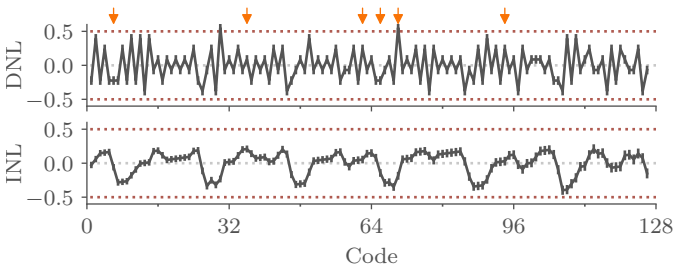

Figure 7. DNL and INL of one 7 bit ADC array, extracted from a post-layout simulation with a transistor-level controller at an interleaved sampling frequency of 62.5 MS s^{-1} (500 MHz clock frequency). Arrows denote output code jitter at code transitions.

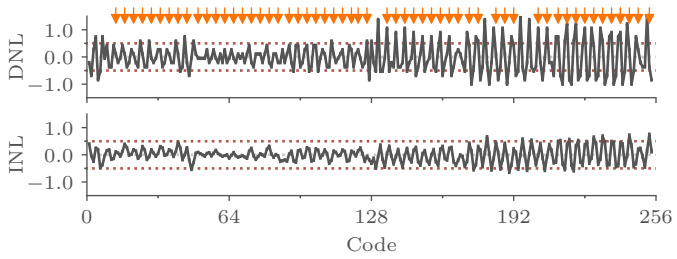

Figure 8. DNL and INL in the precise mode, extracted from a post-layout simulation with a transistor-level controller at an sampling frequency of 12.5 MS s^{-1} (200 MHz clock frequency). Arrows denote output code jitter at code transitions.

comes from even-odd effects at the two reference generator, on another hand, it is caused by simulation artifacts. Generally, we observed static even-odd effects when interleaving the two 7 bit DACs. In the fast mode (7 bit, 62.5 MS s^{-1}), the maximum DNL is 0.55 LSB and the maximum INL is 0.40 LSB for one individual array. The pattern in the INL-curve is an artifact caused by restarting the simulation. The precise mode (8 bit, 12.5 MS s^{-1}) yields a maximum DNL of 1.53 LSB and a maximum INL of 0.73 LSB. Simulations of the same kind allow estimating a power consumption of 2.8 pJ to 3.3 pJ per conversion in the fast mode and 4.6 pJ to 5.0 pJ in the precise mode, respectively.

VI. Conclusion and Outlook

Compared to other Nyquist SAR ADCs (table I) with similar speed and process nodes, our ADC offers a good balance between the different objectives. By its flexible design, the presented ADC is a milestone in the ongoing development process of the BSS-2 system. The ADC's gain and offset error is limited by parasitic capacitance and its accuracy by capacitor matching, charge injection, switch resistance, and the reference voltage's quality. In further design iterations we would like to overcome the even-odd effects in the reference generator and between the two ADCs in the fast mode by an improved layout. Also the transmission gates's parameters should be adopted to optimize for on- and off resistance. Using metal-isolator-metal (MIM) capacitors instead of MOM capacitors promise an improved capacitor accuracy, but increases the area of the unit-capacitor. To test for Nyquist sampling, we will repeat our simulations on randomized input voltages. Finally, we plan to verify our results in a silicon prototype.

Table I
Comparison of our simulated ADC with other implementations.
*Post-layout simulation, †Pre-layout simulation.

	Xu [6]	V. d. Plas [7]	Harpe [8]	This work	
Process node	65 nm	90 nm	65 nm	65 nm	
Architecture	async. SAR	hybrid SAR	SAR	interl. SAR	SAR
Resolution / bit	7	7	10	7	8
Sampling / MS s^{-1}	40	150	30	62.5	12.5
max. DNL / LSB	1.08	≤1.00	0.55	0.55*	1.53*
max. INL / LSB	-1.20	≤0.52	0.39	0.40*	0.73*
Energy / pJ conv^{-1}	7.46	0.89	2.39	3.1†*	4.8†*
thereof analog	1.72	-	1.67†	1.4*	1.7*
thereof digital	5.75	-	0.50†	1.7†	3.2†
Supply / V	1.0	1.0	1.0	1.2	
Area / μm^2	17,000	50,000	1,000	≥1,400	

Acknowledgements

We thank Y. Stradmann, J. Ilmberger, the Electronic Vision(s) group, and all others involved for their contributions.

This work has received funding from the European Union research and innovation funding H2020 945539 (HBP SGA3) and DFG project EXC 2181/1 – 390900948 (STRUCTURES).

Autor Contributions

This work recap on the master theses of P.D. and M.C. Both theses were supervised by J.S. P.D. developed the SAR, set up the verification framework and designed the comparator as well as the reference generator. M.C. elaborated the dual-mode capacitor concept. The simulation results were performed in a joint effort of P.D. and M.C. S.B. supported the analog aspects and A.G. the digital ones. P.D., S.B. A.G. and J.S. edited the text.

References

[1] J. Schemmel, S. Billaudelle, P. Dauer, *et al.*, "Accelerated analog neuromorphic computing," in *Analog Circuits for Machine Learning, Current/Voltage/Temperature Sensors, and High-speed Communication*, Springer, 2022, pp. 83–102.

[2] S. Billaudelle, J. Weis, P. Dauer, *et al.*, "An accurate and flexible analog emulation of adex neuron dynamics in silicon," in *2022 29th IEEE International Conference on Electronics, Circuits and Systems (ICECS)*, IEEE, 2022, pp. 1–4.

[3] S. Friedmann, J. Schemmel, A. Grübl, *et al.*, "Demonstrating hybrid learning in a flexible neuromorphic hardware system," *IEEE transactions on biomedical circuits and systems*, vol. 11, no. 1, pp. 128–142, 2016.

[4] J. Weis, P. Spilger, S. Billaudelle, *et al.*, "Inference with artificial neural networks on analog neuromorphic hardware," in *IoT Streams for Data-Driven Predictive Maintenance and IoT, Edge, and Mobile for Embedded Machine Learning*, Springer, 2020, pp. 201–212.

[5] Y. Stradmann, S. Billaudelle, O. Breitwieser, *et al.*, "Demonstrating analog inference on the brainscales-2 mobile system," *IEEE Open Journal of Circuits and Systems*, 2022.

[6] Y. Xu and T. Ytterdal, "A 7-bit 40 ms/s single-ended asynchronous sar adc in 65 nm cmos," *Analog Integrated Circuits and Signal Processing*, vol. 80, no. 3, pp. 349–357, 2014.

[7] G. Van der Plas and B. Verbruggen, "A 150 ms/s 133 μw 7 bit adc in 90 nm digital cmos," *IEEE Journal of Solid-State Circuits*, vol. 43, no. 12, pp. 2631–2640, 2008.

[8] P. Harpe, "A compact 10-b sar adc with unit-length capacitors and a passive fir filter," *IEEE Journal of Solid-State Circuits*, vol. 54, no. 3, pp. 636–645, 2018.

[9] D. Schinkel, E. Mensink, E. Klumperink, *et al.*, "A double-tail latch-type voltage sense amplifier with 18ps setup+ hold time," in *2007 IEEE international solid-state circuits conference. Digest of technical papers*, IEEE, 2007, pp. 314–605.

[10] M. Miyahara, Y. Asada, D. Paik, *et al.*, "A low-noise self-calibrating dynamic comparator for high-speed adcs," in *2008 IEEE Asian Solid-State Circuits Conference*, IEEE, 2008, pp. 269–272.