Noise Analysis of Programmable Gain Analog to Digital Converter for Integrated Neural Implant Front End

Amir Zjajo, Carlo Galuzzi, Rene van Leuken

Circuits and Systems Group, Delft University of Technology, Mekelweg 4, 2526 CD, Delft, The Netherlands
a.zjajo@tudelft.nl, c.galuzzi@tudelft.nl, t.g.r.m.vanleuken@tudelft.nl

Keywords: Noise, Programmable Gain Amplifiers, Neural Implant Front End, Analog to Digital Converters.

Abstract: Integrated neural implant interface with the brain using biocompatible electrodes to provide high yield cell recordings, large channel counts and access to spike data and/or field potentials with high signal-to-noise ratio. By increasing the number of recording electrodes, spatially broad analysis can be performed that can provide insights into how and why neuronal ensembles synchronize their activity. However, the maximum number of channels is constrained with noise, area, bandwidth, power, thermal dissipation and the scalability and expandability of the recording system. In this paper, we characterize the noise fluctuations on a circuit-architecture level for efficient hardware implementation of programmable gain analog to digital converter for neural signal-processing. This approach provides key insight required to address signal-to-noise ratio, response time, and linearity of the physical electronic interface. The proposed methodology is evaluated on a prototype converter designed in standard single poly, six metal 90-nm CMOS process.

1 INTRODUCTION

Bio-electronic interfaces allow the interaction with neural cells by both recording, to facilitate early diagnosis and predict intended behavior before undertaking any preventive or corrective actions (Nicolelis, 2001), or stimulation devices, to prevent the onset of detrimental neural activity such as that resulting in tremor. Monitoring large scale neuronal activity and diagnosing neural disorders has been accelerated by the fabrication of miniaturized micro-electrode arrays, capable of simultaneously recording neural signals from hundreds of channels (Frey, 2007). By increasing the number of recording electrodes, spatially broad analysis of local field potentials can be performed that can provide insights into how and why neuronal ensembles synchronize their activity. Studies on body motor systems has uncovered how kinematic parameters of movement control are encoded in neuronal spike time-stamps (Georgopoulos, 1986) and inter-spike intervals (Chae, 2009). Neurons produce spikes of nearly identical amplitude near to the soma, but the measured signal depend on the position of the electrode relative to the cell. Additionally, the signal quality in neural interface front-end, beside the specifics of the electrode material and the electrode/tissue interface, is limited by the nature of the bio-potential signal and its biological background noise, dictating system resources.

For any portable or implantable device, micro-electrode arrays require miniature electronics locally to amplify the weak neural signals, filter out noise and out-of-band interference and digitize for transmission. A single-channel (Yin, 2007) or a multi-channel integrated neural amplifiers and converters provide the front-line interface between recording electrode and signal conditioning circuits and thus face critical performance requirements. Multi-channel, fully differential designs allow for spatial neural recording and stimulation at multiple sites (Shahrokhi, 2010, Gao, 2012, Han, 2013). The maximum number of channels is constrained with noise, area, bandwidth, power (Chae, 2008), which has to be supplied to the implant from outside, thermal dissipation i.e. to avoid necrosis of the tissues even by a moderate heat flux (Seese, 1998) and the scalability and expandability of the recording system. The block diagram of a typical neural recording system architecture is illustrated in Figure 1(a). When a neuron fires an action potential, the cell membrane becomes depolarized by the opening of voltage-controlled neuron channels leading to a flow of current both inside and outside the neuron.
Since extracellular media is resistive (de Zeeuw, 2011), the extracellular potential is approximately proportional to the current across the neuron membrane (Kölbl, 2010). The membrane roughly behaves like an RC circuit and most current flows through the membrane capacitance (West, 1991). The data acquired by the recording electrodes are conditioned using analog circuits. As a result of the small amplitude of neural signals and the high impedance of the electrode tissue interface, amplification and band-pass filtering of the neural signals is performed before the signals can be digitized by a successive approximation register (SAR)-based analog to digital converter (ADC) (Harpe, 2013).

To avoid the large capacitive DACs found in the SAR ADC, to lower demands on driving capabilities of the amplifier and relax power, noise and crosstalk requirements, in the alternative architecture illustrated in Figure 1(b), the programmable gain amplifier (PGA) and ADC are combined and embedded in every recording channel. The programmable gain analog to digital converter (PGA ADC) implements simultaneously both signal acquisition and amplification, and data conversion. As illustrated in Figure 2 (Rodríguez-Pérez, 2012), the schematic incorporates a fully-differential operational transconductance amplifier (OTA), a comparator and circuitry for control of the acquisition and amplification operation set by the clock phases $\phi_1$, $\phi_2$ and $\phi_3$, and output generation data conversion operation, controlled by the clock phases $\phi_1$ and $\phi_2$. The recorded signals are capacitively coupled to the input of the amplifier to reject the dc polarization. The differential input signal is sampled, amplified by the capacitance ratio (gain $G^A$ is adjustable by implementing $C_3$ as a programmable capacitor array, $G^A= C_3/C_4$), and transferred to the integration capacitors $C_2$ at the feedback loop of the OTA. At data conversion operation, the differential signal stored in $C_2$ is converted to digital domain by successively adding or subtracting binary-scaled versions of the reference voltage to the integration capacitors. Voltage addition or subtraction is implemented by means of the four cross-coupled switches controlled by the signals $\phi_2p$ and $\phi_2n$. Digital representation of the output signals are then sequentially stored in SAR register for further processing. A low-power monolithic digital signal processing (DSP) unit provides additional filtering and executes a spike discrimination and sorting algorithms (Harrison, 2007). The relevant information is then transmitted to an outside receiver through the transmitter or used for stimulation in a closed-loop framework. Understanding the role of noise in such systems is one of the central challenges in the heterogeneous neural simulation and neural rehabilitation (Harrison, 2008). In this paper, we try to characterize the noise fluctuations on a circuit-architecture level for efficient hardware implementation of neural signal-processing circuitry. This approach provides key insight required to address signal-to-noise ratio (SNR), response time, and linearity of the physical electronic interface (i.e., saturation level).
2 NOISE CHARACTERIZATION

For discrete-time analog signal-processing circuits, analog signals are acquired and processed consecutively, and a sample of the signal is taken periodically controlled by a clock signal. As the sampling circuit cannot differentiate the noise from the signal, part of this signal acquisition corresponds to the instantaneous value of the noise at the moment the sampling takes place. In this context, when the sample is stored as charge on a capacitor, the root-mean-square total integrated thermal noise voltage is $kT/C$, where $kT$ is the thermal energy. This noise usually comprises two major contributions - the channel noise of the switches, which is a function of the channel resistance and the OTA noise. The OTA output noise is in most cases dominated by the channel noise of the input transistors, where the thermal noise and the $1/f$ noise both contribute. If the input transistors of the OTA are biased in saturation region to derive large transconductance $g_m$, impact ionization and hot carrier effect will enhance their thermal noise level (Enz, 2000). Similarly, the $1/f$ noise increases as well due to the reduced gate capacitance resulted from finer lithography and therefore shorter minimum gate length. As a consequence, an accurate consideration of the intrinsic noise sources in such a circuit should have the thermal noise of switches and all amplifier noises readily included. Nevertheless, the input-referred noise $v_n$ (the total integrated output noise as well) still takes the form of $kT/C$ with some correction factor $\chi$.

\[
\bar{v}^2_n = \chi \cdot kT / C_4
\]  

\[
(1)
\]

2.1 Noise Analysis of Programmable Gain Analog to Digital Converter

A fundamental technique to reduce the noise level, or to increase the signal-to-noise ratio of a programmable gain ADC, is to increase the size of the sampling capacitors, by over-sampling or with calibration. However, for a fixed input bandwidth specification, the penalty associated with these techniques is the increased power consumption. Consequently, a fundamental trade-off exists between noise, speed, and power dissipation.

2.1.1 $kT/C$ Noise

During the acquisition process, $kT/C$ noise is sampled on the capacitors $C_i$ along with the input signal. To determine the total noise charge sampled onto the capacitor network, noise charge $Q_{ns}$ is integrated over all frequencies

\[
\overline{Q_{ns}} = \int_0^{2\pi} \frac{V_n}{|1+j\omega R_m(C_i + C_p + C_{\text{ota}})|} d\omega = \frac{kT(C_i + C_p + C_{\text{ota}})}{\beta C_4}
\]

\[
(2)
\]

where $R_m$ is resistance of the switch, $V_n$ is noise source, $C_p$ is parasitic capacitance and $C_{\text{ota}}$ is the input capacitance of the OTA. Then in the conversion mode, the sampling capacitor $C_s$ which now contains the signal value and the offset of the OTA, is connected across the OTA. The total noise charge will cause an output voltage of

\[
\overline{v_{\text{ac}(\text{con})}}^2 = \frac{Q_{ns}^2}{C_4^2} = kT \left( C_i + C_p + C_{\text{ota}} \right) \frac{1}{\beta C_4} = \frac{kT}{\beta C_4}
\]

\[
(3)
\]

where $\beta$ is the feedback factor. For differential implementation of the circuit, the noise power of the previous equation increases by a factor of 2 assuming no correlation between positive side and negative side, since the uncorrelated noise adds in power. Thus, input referred noise power, which is found by dividing the output noise power by the square of the gain ($G^4 = C_4/C_1$) is given by

\[
\overline{v_{\text{ac}(\text{con})}^2} = \overline{v_{\text{ac}(\text{int})}^2} = \frac{kT}{\beta (G^4)^2 C_4}
\]

\[
(4)
\]

2.1.2 OTA Noise in Conversion Mode

The resistive channel of the MOS devices in OTA also has thermal noise and contributes to the input referred noise of the PG ADC circuit. The noise power at the output is found from

\[
\overline{v_{\text{ac}(\text{con})}^2} = \frac{\gamma}{2} \left[ H \left( \frac{1}{C_{\text{ota}}} \right) \right]^2 x_\text{out}^2 d\omega = \frac{kT \gamma \frac{G_R}{C_{\text{LTR}}}}{\beta C_{\text{LTR}}}
\]

\[
(5)
\]

where $R_o$ is the output resistance and $C_{\text{LTR}}$ is the capacitance loading at the output

\[
C_{\text{LTR}} = C_i + \beta \times \left( C_p + C_{\text{ota}} \right)
\]

(6)

The thermal noise coefficient $\gamma$ depends on the effective mobility and channel length modulation (Jindal, 2006); it is $2/3$ for older technology nodes and between 0.6 and 1.3 for submicron technologies (Ou, 2011). The optimum gate capacitance of the OTA is proportional to the sampling capacitor $C_{\text{OTA}} = \chi_L C_i$, where $\chi_L$ is a circuit-dependent proportionality factor. The drain current $I_D$ yields

\[
I_D = \frac{\chi_L^2 \omega C_i}{\mu x_5}
\]

\[
(7)
\]
where $\mu$ is the carrier mobility, $C_{ox}$ is the gate oxide capacitance, $g_0$ is the gain-bandwidth product, and $W$ and $L$ are the channel width and length. Assuming $G_{o}R_{\beta}$>1, and gain of the conversion operation $G^2=C_{o}/C_{d}$, the input referred noise variance is

$$\frac{\sigma^2_{\text{var}(\text{in})}}{\beta} = \frac{2V}{C_{\text{int}}} \left( \frac{kT}{G^2} \right)^2 \frac{\beta}{G^2} \frac{1}{C_{\text{int}}}$$

(8)

### 2.1.3 Total Input Referred Noise

The noise from acquisition and conversion mode can be added together to find the total input referred noise assuming that two noise sources are uncorrelated. Using the results from (4) and (8), the total input referred noise power for differential input is given by

$$\frac{\sigma^2_{\text{var}(\text{in})}}{\beta} = \frac{2V}{C_{\text{int}}} \left( \frac{kT}{G^2} \right)^2 \frac{\beta}{G^2} \frac{1}{C_{\text{int}}}$$

(9)

For a noise dominated by $kT/C$, the power consumption is found as

$$P_{o} \propto I_{DD} \cdot V_{DD} = \frac{\mu g_{m} \cdot g_{m} \cdot \text{SNR} \cdot 8kT \cdot V_{DD}}{\mu \gamma \cdot V_{\text{rms}}^2}$$

(10)

For a given speed requirement and signal swing, a two times reduction in noise voltage requires a four times increase in the sampling capacitance value and the OTA size. This means that the PG ADC circuit power quadruples for every additional bit resolved for a given speed requirement and supply voltage as illustrated in Figure 3 and Figure 4. Notice that for a small sampling capacitor values, thermal noise limits the SNR, while for a large sampling capacitor, the SNR is limited by the quantization noise and the curve flattens out. Improving the power efficiency beyond topological changes of the OTA and supply voltage reduction require smart allocation of the biasing currents. Hence, techniques such as current reuse (Song, 2013, Zou, 2013), time multiplexing (Chae, 2009, Zou, 2013) and adaptive duty-cycling of the entire analog front end (Lee, 2010, Abdelhalim, 2011) can be used to improve power efficiency by exploiting the fact that neurons spikes are irregular and low frequency. Choosing the OTA bandwidth too high increases the noise and additionally demands unnecessarily low on-resistance of the switches and thus large transistor dimensions. The optimum time constant remains constant regardless of the circuit size (or $I_D$) because $C_L$ scales together with $C_d$ and the parasitic capacitance $C_p$.

![Figure 3: Maximum achievable SNR for different sampling capacitor values and resolutions.](image3)

![Figure 4: SNR versus power dissipation.](image4)

The choice of the hold capacitor value is a trade-off between noise requirements on the one hand and speed and power consumption on the other hand. The sampling action adds $kT/C$ noise to the system which can only be reduced by increasing the hold capacitance $C_H$. A large capacitance, on the other hand, increases the load of the operational amplifier and thus decreases the speed for a given power. The OTA size and its bias current for a given speed requirement and minimum power dissipation are determined using $\tau$-vs.-$C_H$ curves as in Figure 5. Note that for low frequency operation (where $\tau$/$\tau_l$ is large), the $C_{\text{OTA}}$ that achieves the minimum power dissipation for given settling time and noise requirements, usually does not correspond to the minimum time constant point. This is a consequence of setting the $C_H/C_{\text{OTA}}$ ratio of the circuit to the minimum time constant point, which requires larger $C_{\text{OTA}}$ and results in power increase and excessive bandwidth.

Near the speed limit of the given technology (where the ratio $\tau$/$\tau_l$ is small), however, the difference in power between the minimum power point and the minimum time constant point becomes
smaller as the stringent settling time requirement forces the $C_{4}/C_{\text{OTA}}$ ratio (Figure 6) to be at its optimum value to achieve the maximum bandwidth.

2.2 The OTA Noise

The OTA in PG ADC circuit has some unique requirements: the most important is the input impedance, which must be purely capacitive so as to guarantee the conservation of charge. Consequently, the OTA input has to be either in the common source or the source follower configuration. Another characteristic feature is the load at the OTA output, which is typically purely capacitive and as a result, the OTA output impedance must be high. The benefit of driving solely capacitive loads is that no output voltage buffers are required.

![Figure 5: Closed loop normalized time constant versus hold capacitance $C_{4}$ for different biasing conditions; case for $C_{c}=3C_{c}, C_{1}=C_{p}$. The time constant is normalized to the $\tau_{f} (=1/f_{\text{intrinsic}})$ of the device, which is approximately $(C_{4}/R_{\text{in}})$.](image)

The implemented folded-cascode OTA and dynamic latch are illustrated in Figure 7. The input stage of the OTA is provided with two extra transistors $T_{10}$ and $T_{11}$ in a common-source connection, having their gates connected to a desired reference common-mode voltage at the input, and their drains connected to the ground (Bult, 1990). The advantage of this solution is that the common-mode range at the output is not restricted by a regulation circuit, and can approach a rail-to-rail behavior very closely.

The transistors of the output stage have two constrains: the $g_{m}$ of the cascading transistors $T_{5,6}$ must be high enough, in order to boost the output resistance of the cascode, allowing a high enough $dc$ gain and the saturation voltage of the active loads $T_{3,4}$ and $T_{7,8}$ must be maximized, in order to reduce the extra noise contribution of the output stage. These considerations underline a tradeoff between fitting the saturation voltage into the voltage headroom and minimizing the noise contribution. A good compromise is to make the cascading transistors larger than the active loads: in such a way the $g_{m}$ of the cascading transistors is maximized, boosting the $dc$ gain, while their saturation voltage is reduced, allowing for a larger saturation voltage for the active loads, without exceeding the voltage headroom. The output SNR is equal to

$$\text{SNR}_{\text{out}} = \frac{C_{1}}{2kT} \frac{A^{2} \times g_{m1} R_{\text{out}}}{g_{m2} + g_{m3} + g_{m4}}$$

where $A$ is the amplitude of the input signal and $R_{\text{out}}$ denotes the open-loop output resistance of the OTA. From (11) can be concluded that in order to maximize the output SNR, $C_{1}$ must be maximized, which means that bandwidth must be minimized.

The noise contribution of the individual transistors are shown in Figure 8. The input-referred noise of the OTA input pair is reduced by increasing the $g_{m}$, increasing the current, or increasing the aspect ratio of the devices. The effect of the last method, however, is partially canceled by the increase in the noise excess factor. When referred to the OTA input, the noise voltages of the transistors used as current sources (or mirrors) in the first stages are multiplied by the $g_{m}$ of the device itself and divided by the $g_{m}$ of the input transistor, which again suggests that maximizing input pair $g_{m}$ minimizes noise. It can be further reduced by decreasing the $g_{m}$ of the current sources. Since the current is usually set by other requirements, the only possibility is to decrease the aspect ratio of the device. This leads to an increase in the gate overdrive voltage, which, as a positive side effect, also decreases $\gamma$.

![Figure 6: Optimum gate capacitance $C_{\text{OTA, opt}}$ versus hold capacitance $C_{4}$ for different loading and parasitic conditions.](image)
Increasing $L$ to avoid short channel effects is also possible, although with a constant aspect ratio it increases the parasitic capacitances.

The dynamic latch illustrated in Figure 7(b) consists of pre-charge transistors $T_{14}$ and $T_{17}$, cross-coupled inverter $T_{12-13}$ and $T_{15-16}$, differential pair $T_{10}$ and $T_{11}$ and switch $T_9$, which prevent the static current flow at the resetting period (Kobayashi, 1993). When the latch signal is low (resetting period), the drain voltages of $T_{10-11}$ are $V_{DD}-V_T$ and their source voltage is $V_T$ below the latch input common mode voltage. Therefore, once the latch signal goes high, the $n$-channel transistors $T_{11-13}$ immediately go into the active region. Because each transistor in one of the cross-coupled inverters turns off, there is no static power dissipation from the latch once the latch outputs are fully developed. A large portion of the total comparator current is allocated to the input branches to boost the input $g_m$. Similarly, the noise from the non-gain element i.e. the load transistor, is minimized, by applying small biasing current. Additionally, small width and large length for their gate dimensions is chosen.

![Figure 7: a) OTA schematic, b) Comparator schematic.](image)

3 EXPERIMENTAL RESULTS

The time series representation of an neuron signal (Figure 9) are composed of a spike burst, plus additive Gaussian white noise (grey area with 1000 randomly selected neural channel compartments). In typical electrode-tissue interface, we are relying on the current measurement to sense these neural signals. Hence, by maintaining a constant current density, the relative uncertainty of the current becomes inversely proportional to the square of the interface area. The electrode noise spectral density has an approximate dependence of -10 dB/dec for small frequencies. However, for frequencies higher than 1-10 kHz, capacitances at the interface form the high-frequency pole and shape both the signal and the noise spectrum; the noise is low-pass filtered to the recording amplifier inputs. After band-pass filtering and amplification, the noisy neural signal is further processed with programmable gain analog to digital converter. The fluctuation of the voltage on the sampling capacitor is inversely proportional to the capacitance (the variance of the capacitor voltage is $kT/C$ at any given time). This implies that with scaling, the uncertainty of the sampled voltage increases. It can be seen that in both scenarios, in electrode-tissue interface and in PG ADC, the noise in the neural interface front end greatly increases as the interface size reduces. The interface’s input equivalent noise voltage decreases as the gain across the amplifying stages increase, e.g. the ratio of the square of the signal power over its noise variance can be expressed as

$$SNR = A^2 \left[ \frac{V_{	ext{noise}}}{V_{	ext{signal}}} + \frac{V_{	ext{noise}}}{V_{	ext{signal}}} + \sum G_j \rho_{	ext{amp,ij}} \right]$$  \hspace{1cm} (12)$$

where $\rho_{	ext{amp,ij}}$ represents the variance of the noise added by the $i$th amplification stage with gains $G_j$. 

![Figure 8: Noise contribution of the individual transistors in the OTA](image)
The variance of the electrode is denoted as $v_{\text{ns(electrode)}}^2$ and $v_{\text{ns(neural)}}^2$ is variance of the biological neural noise. The observed SNR of the system also increases as the system is isomorphically scaled up, which suggests a fundamental trade-off between SNR and speed of the system. This lower bound on the speed in a converter loop is primarily a function of the technology’s gate delay and $kT/C$ noise multiplied by the number of SAR cycles necessary for one conversion.

All PG ADC simulations were performed with a 1.2 V supply voltage at room temperature (25ºC). Spectral signature of PG ADC is illustrated in Figure 10. The circuit offers a programmable amplification of 0-18 dB by digitally scaling the input capacitance $C_3$. As shown in Figure 11, the Signal-to-Noise and Distortion Ratio (SNDR), Spurious-Free Dynamic Range (SFDR) and Total Harmonic Distortion (THD) remain constant at different gain settings. The THD in the range of 10-100 kS/s is above 54 dB for $f_{\text{in}}$ of 5 kHz (Figure 12). Within the bandwidth of neural activity of up to 5 kHz, SNDR is above 44 dB and SFDR more than 57 dB. The degradation with a higher input signal is mainly due to the parasitic capacitance, clock non-idealities and substrate switching noise.

Parasitic capacitance decreases the feedback factor resulting in an increased settling time constant. The non-idealities of clock such as clock jitter, non-overlapping period time, finite rising and fall time, unsymmetrical duty cycle are another reason for this degradation. The three latter errors reduce the time allocated for the setting time.

4 CONCLUSIONS

The high density of neurons in neurobiological tissue requires a large number of electrodes for accurate representation of neural activity. To develop neural prostheses capable of interfacing with single neurons and neural networks, multi-channel neural probes and the electrodes need to be customized to the anatomy and morphology of the recording site. The increasing density and the miniaturization of the functional blocks in these multi-electrode arrays, however presents significant circuit design challenge in terms of area, bandwidth, power, and the scalability, programmability and expandability of the recording system. In this paper, for one such functional block, programmable analog to digital converter, we evaluate trade-off between noise, speed, and power dissipation and characterize the noise fluctuations on a circuit-architecture level.
This approach provides key insight required to address SNR, response time, and linearity of the physical electronic interface.

ACKNOWLEDGEMENTS

This research was supported in part by the European Union and the Dutch government as part of the CATRENE program under Heterogeneous INCEPTION project.

REFERENCES


Han, D., et al., 2013, A 0.45 V 100-channel neural-recording IC with sub-μW/channel consumption in 0.18 μm CMOS, IEEE Transactions on Biomedical Circuits and Systems, vol. 7, no. 6, pp. 735-746.


Ou, J., 2011, g_{m}/I_{b} based noise analysis for CMOS analog circuits, Proceedings of IEEE International Midwest Symposium on Circuits and Systems, pp. 1-4.


