Comprehensive Analysis and Optimization of CMOS Neural Amplifiers for Wireless Recording Implants

Haitao Li and Andrew J. Mason, *Senior Member, IEEE*
Michigan State University, East Lansing, MI 48823, {lihaitao, mason}@msu.edu

Abstract—**Neural amplifiers play a critical role in the bandwidth, power consumption, noise performance and area of wireless neural recording systems. This paper presents comprehensive analysis of neural amplifier design to optimize these performance characteristics. Amplifier architectures and feedback structures are analyzed to identify the best choice for implantable neural recording applications. Bandwidth and phase margin requirements are analyzed to identify an optimization path through relaxing phase margin requirements. To optimize the noise performance and circuit area, noise capacitive matching is applied to optimize the size of neural amplifier capacitors. Following the optimization guidelines identified in this paper, noise efficiency and a figure of merit for neural amplifiers were effectively improved.**

I. Introduction

Neural recording systems can provide unique insight into the function of the human brain. Because of their promise in developing a brain-machine interface (BMI) and potential for monitoring and diagnosing neural disorders, neural recording systems are being explored by many research groups. Neural amplifiers provide the front-line interface between recording electrode and signal conditioning circuits and thus face critical performance requirements. One of the development paths for BMI systems is a fully implanted multi-channel neural recording system with wireless power and data transmission [1]. In such systems, the strict requirements of the neural amplifier are further complicated because of the demand for ultra low power consumption, the presence of a large ripple on transmitted power line and the need for a multi-channel analog front end.

This paper presents a comprehensive analysis of neural amplifier design in order to optimize the bandwidth, noise, power and area for wireless neural recording implants. Fig. 1 illustrates the typical differential configuration for neural amplifiers. Within the reported neural amplifier designs, different feedback path options operational transconductance amplifier (OTA) structure, and capacitor values have been utilized, and these choices must be carefully analyzed to optimize a neural amplifier. This paper provides a thorough analysis of these available design options and identifies the best choice of neural feedback structure and OTA for implantable recording applications. To optimize the OTA design, the open loop phase margin requirement is analyzed. To minimize the noise and area, noise capacitive matching technology is considered and the optimized neural amplifier capacitance is calculated. The optimized neural amplifier is described and shown to provide a significant improvement in noise efficiency factor and a comprehensive figure of merit.

II. Neural Amplifier Feedback Structures

The main duty of a neural amplifier is to provide gain over the limited bandwidth of neural action potentials. Neural spikes typically exhibit amplitudes ranging from 10μV to 500μV and contain data up to ~7kHz. Neural amplifiers should also block low frequency interference, including DC offset from electrode that are typically around 1V, and local field potentials (LFP), typically 0.1mV to 50mV at 300Hz and below. Neural amplifiers typically employ two different feedback path structures, defined as I and II in Fig. 1, to realize a high-pass filter. Style I is realized with two subthreshold-biased transistors [2], while style II is realized with two diode-connected transistors [3].

To compare these two structures, neural amplifiers with both feedback paths were implemented in a 0.5μm CMOS test chip. To accommodate an n-well technology, the substrates of the feedback path transistors were connected to GND, for style I. Experimental results from the feedback structure test chip are summarized in Table I. Style I was shown to be capable of an adjustable low-cutoff frequency, from less than 1Hz to several hundred Hz by changing the voltage V_{tune}, in Fig. 1. In contrast, style II provides a fixed low-cutoff frequency of less than 10Hz even with large size M1 and M2 in feedback path. Thus, for neural recording applications, style II would require an additional filter, e.g., a g_m-C filter or a simpler structure [3], to provide a higher low-cutoff frequency. This additional filtering would occupy more chip area and weaken the total harmonic distortion (THD).

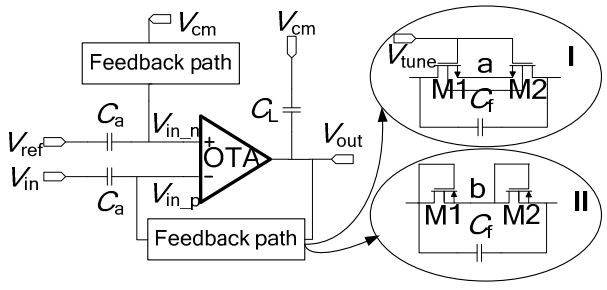

Fig. 1. Neural amplifer with two optional feedback paths. The feedback paths at the two inputs of OTA are the same to improve CMRR.

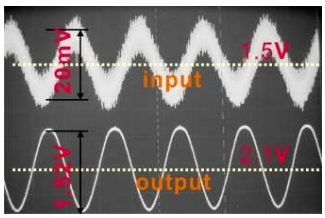

Fig. 2. Feedback structure test chip results for style I. The input is 1kHz, $20mV_{pp}$ sine wave and the output is $1.52V_{pp}$ sine wave with 0.6V dc offset.

TABLE I. PERFORMANCE COMPARISION OF FEEDBACK STRUCTUERS

Performance	Feedback path styles		
	Style I	*Style II*	*Style II + filter*
THD[1]	0.9%	0.37%	0.8%
Low cut-off freq.	Adjustable	Fixed	Adjustable
Ouput offset	Yes	None	None
Area	Small	Small	Medium

1. Simulation result

Style I produces an offset at V_{out} that is proportional to the input amplitude and can be several mV for neural spikes with amplitude over 300μV. This offset voltage in style I is due to the gate voltage of M1 and M2 being fixed by V_{tune} (0.6V larger than V_{cm}), as shown in Fig. 2. When V_{in_p} is equal to V_{cm}, V_{out} is pushed to a voltage higher than V_{cm} to ensure M1 and M2 are on and operated in the subthreshold region. This offset also weakens the THD in style I. In style II, the feedback transistors act as pseudoresistor elements [3]. The MOSFETs and parasitic bipolar transistors in these elements act as converse polarity diodes connected in parallel so the output node will never be pushed to an offset above V_{cm}.

Style II without an additional filter the best choice for low frequency neural signals (e.g., LFP) because it has better THD and no offset. However, for recording neural spikes (300Hz to 7kHz), style II requires a filter to raise the low cut-off frequency, which will weaken the THD and increase power consumption. To accurately compare style I and style II, an additional filter was simulated with style II to match the bandwidth of style I. Table I shows that the THD of style I is similar to style II with a filter. The output offset is not critical because it can be removed by the following variable gain stage. Thus, style I was determined to be better for recording neural spikes because it requires lower power and less chip area.

III. NEURAL AMPLIFIER OTA STRUCTURES

Many different OTA structures have been reported for neural amplifiers, including a 2-stage amplifier [2], a current mirror [3], a folded-cascode amplifier [4], and a telescopic amplifier [5]. To determine the optimal OTA structure for neural amplifiers, OTA performance was analyzed relative to the requirements of neural recording applications.

A. OTA open loop gain

Open loop gain will affect the closed loop gain error as described by

$$Gain\ \ error(\%) = -\frac{100}{T} = -\frac{100}{A_o\beta} \qquad (1)$$

where T is the loop gain of the neural amplifier, A_o is the open loop gain of the OTA and β is the feedback factor, equal to C_f/C_a. (1) shows that the gain error increases if the closed loop gain A_c (represented by $1/\beta$) increases for a fixed A_o. However, a large A_c is needed for weak neural inputs. For a 40dB closed loop gain, the OTA open loop gain should be larger than 80dB to ensure gain error is less than 1%. Higher open loop gain can further reduce the gain error.

B. OTA phase margin and bandwidth

In all the published papers on neural recording systems, the OTA open loop phase margin (PM) is set to a value larger than 50. However, higher closed loop gain inherently provides a larger loop PM [6]. For a neural amplifier with a large closed loop gain of 40dB, the requirement for the OTA PM at 0dB can be greatly relaxed. To illustrate this principle, suppose that the OTA is a two-pole system with dominate and secondary poles at f_d and f_{nd}, respectively. The loop gain T can thus be expressed as

$$T = \beta(f)A_o(f) = \frac{\beta A_o}{(1 + jf / f_d)(1 + jf / f_{nd})} \qquad (2)$$

which results in the gain and phase response illustrated in Fig. 3. For two possible closed loop gains, A_{c1} and A_{c2}, the loop PM and loop gain T, which is the difference between A_o and A_{cx}, are shown. Here, PM_1 and PM_2 are larger than 45º even though PM_0 is almost 0º. Note that PM_1 is larger than PM_2 because T_1 is smaller than T_2. Thus, because the closed loop gain can be as large as 40 dB, the PM of the neural amplifier OTA can be very small. This permits OTA design to tradeoff PM for other performance parameters. Considering also that the desired high-cutoff frequency for neural amplifiers is quite low, ~7kHz, neither bandwidth nor phase margin are critical OTA parameters. However, because frequency characteristics can vary slightly from theoretical models, and because the neural amplifier filter feedback loop will introduce a new pole, it is important to test stability of the closed loop system while setting the OTA PM.

C. OTA structure analysis

For implantable neural amplifiers, the OTA should demand minimal power consumption to avoid damage to human tissue and permit maximum channel density. It should have extremely low noise for high SNR. A very high open loop gain is preferable to minimize the gain error for a large

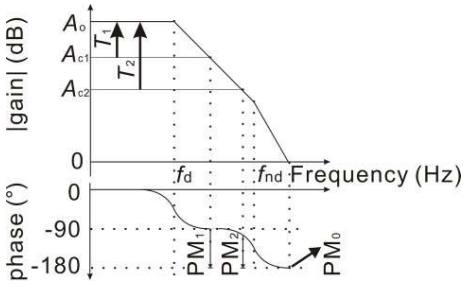

Fig. 3. Bode plot of gain for a two-pole system. PM_0 is OTA's open loop phase margin at 0dB and closed to 0º, whereas $PM_{1,2}$ is closed loop phase margin, larger than 45º.

closed loop gain, and high CMRR is needed to reject common mode noise including common mode neural signals and interference from 50/60 Hz power line noise. For implants with wirelessly transmitted power, a high PSRR is needed to reject the large ripple on the power line, which can be larger than 10mV even if the wireless power receiver is followed by a regulator. For example, a PSRR over 80dB is necessary to ensure 10-bit accuracy for a 10mV ripple, assuming a 10μV input. The bandwidth requirement is not strict because the high cut-off frequency is less than 10kHz and the phase margin does not need to be large. The input and output range does not need to be large; for example, for a closed loop gain of 40dB, the output amplitude for a 10μV to 500μV input is only 0.1mV to 50mV.

Based on theoretical studies and experimental experience, a generalized comparison of key neural amplifier performance parameters is given in Table II for the four common OTA structures. Combining the Table II assessment with the OTA analysis above, it was determined that the telescopic OTA has the best performance for the combination of parameters that are most important for neural amplifiers.

IV. CAPACITANCE OPTIMIZATION FOR NEURAL AMPLIFIER

The analysis in sections II and III defines the preferred feedback path structure and OTA for use in a neural spike recording amplifier. To complete the design, values must be assigned to C_a and C_f. This is not a trivial process because these values have a significant impact on noise and layout area as described in the optimization analysis of this section.

Three primary noise sources must be considered for a neural system: neural signal noise, electrode noise, and electronic noise. For spike recording with a typical microelectrode array, the first two sources can be estimated as 5~10μV_{rms} [7] and 12.6μV_{rms} [8], respectively. The electronic noise should be kept below the total input noise from the neural electrode, which is around 20μV_{rms}. The main source of electronic noise is the neural amplifier because contributions from following stages will be divided by the high gain of the neural amplifier stage and can thus be ignored.

The noise model of a neural amplifier is shown in Fig. 4. The input referred noise V_{ieq} for OTAs in Table I can be expressed as

$$V_{ieq}^{\ 2} = \frac{\left(C_f + C_a + C_p + C_g\right)^2}{C_a^{\ 2}} \times \frac{16kT}{3g_m} \qquad (3)$$

where C_g is MOSFET gate capacitance at the OTA input and

TABLE II. PERFORMANCE COMPARISION OF OTA STRUCTUERS

Performance	OTA structures			
	Two stage	*Current mirror*	*Folded cascade*	*Telescopic*
Gain	good	poor	good	good
Noise	good	poor	fair	good
Power	poor	poor	fair	good
Bandwidth	fair	good	poor	poor
I/O range	good	good	fair	poor
PSRR	poor	poor	good	good

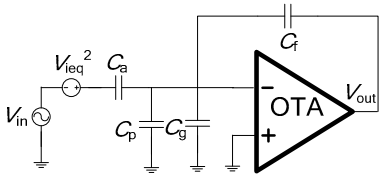

Fig. 4. Noise model for neural amplifier.

C_p is the wiring parasitic capacitance that is negligible compared with C_a. Considering the relationship between C_g and g_m, the theory of capacitive noise matching [6] says that V_{ieq} is minimized when $C_g = (C_a+C_f)$. However, in a practical design this minimum can not be achieved without excessive power consumption and chip area.

Consider the standard model for C_g given by

$$C_g = \frac{2}{3} C_{ox} W L_{eff} (1+\gamma) = \frac{C_f + C_a}{N} \qquad (4)$$

where C_{ox} is the gate oxide capacitance per unit area, γ is a factor relating gate-source and gate-drain capacitances, W and L_{eff} represent the effective gate size, and $N = (C_a+C_f)/C_g$ is a ratio useful in considering capacitive noise matching. C_g is determined by the OTA input transistor size. For neural amplifiers, the sum of C_a and C_f is generally a large value, such that $N \geq 1$. A plot of input referred noise vs. input transistor size [6] shows that smaller N results in better input referred noise when $N \geq 1$. For a fixed input transistor size, i.e. for a fixed C_g, C_f and C_a should kept small to decrease N and thereby improve noise performance. At the same time, this helps to minimize the chip area of the neural amplifier.

V. NEURAL AMPLIFIER DESIGN AND RESULTS

The analysis above can be summarized by the following design guidelines to improve the performance of neural amplifier for multi-channel neural spikes recording implants.

- Style I feedback structure should be used to decrease chip area and power consumption.
- The open loop phase margin of OTA can be set well below 45º to further improve OTA performance.
- A telescopic OTA provides a good tradeoff between noise, power and PSRR.
- The capacitance in the feedback structure should be minimized to improve noise performance and decrease area.

Following the above guidelines, a neural amplifier was designed in 0.5μm CMOS. C_f was chosen to be 100fF to meet the feedback capacitance guideline with a value large enough to overcome stray parasitic capacitance [2]. The closed loop gain of the neural amplifier A_c was set at 100, which is the most common value found in neural amplifiers [2-4, 10]. To match these values, C_a was set to 10pF.

Noise efficiency factor (NEF) is commonly used to qualify the tradeoff between noise, bandwidth and power in neural

amplifiers [3-5, 10]. To set OTA parameters during design, simulations were performed to minimize NEF. Several different bias currents were considered, and 1 µA was found to provide the best NEF. Different input transistor sizes were also tested, and the final size was set to 165µm by 4.5µm because further increase of size beyond this yielded limited improvement in noise performance. PM optimization was found to enable improvements in open loop gain, noise, power performance, CMRR and PSRR. The final neural amplifier layout is shown in Fig. 5 and described by the post-layout simulation parameters in Table III.

For neural amplifiers, layout area and supply voltage are also important factors to maximizing the channel density of a multi-channel neural recording system. To incorporate these parameters into the performance comparison, a new figure of merit (FoM) relating noise, power, and area is given by

$$FoM = NEF^2 \cdot A \cdot V_{dd} = K \cdot V_{ieq}{}^2 \cdot A \cdot P / BW \qquad (5)$$

where A is the chip area, V_{dd} is the supply voltage, P is the amplifier power consumption, BW is the bandwidth, and K is a constant. Lower values of FoM represent a better performance. Table IV shows a comparison of NEF and FoM between reported neural amplifiers. The design reported in this paper consumes the lowest power, has the lowest (best) NEF, and had the lowest (best) FoM. Reported NEF and FoM values, normalized to the results of this work, are plotted in Fig. 6 for easy comparison. Fig. 6 shows that [4] has the closest NEF but has an FoM that is almost two times larger. [5] has the closest FoM but its NEF is 83% larger. This comparison shows that the optimization guidelines developed in this paper enable significant performance improvements in amplifier for wireless neural recordings.

VI. CONCLUSION

A comprehensive analysis of neural amplifiers for design optimization was presented in this paper. Analysis showed that the telescope OTA with voltage tuned feedback is most suitable for a wireless neural spike recording system. OTA performance was found to be optimized with relaxed open loop phase margin requirements. Capacitive noise matching technology was incorporated for noise and area optimization of neural amplifiers. Following the optimization guidelines in this paper, NEF and an FoM for neural amplifiers were effectively improved.

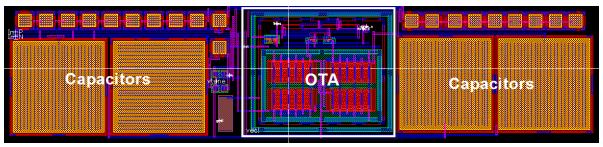

Fig.5. Layouts of neural amplifier following guidelines identified in this paper The area is 507µm×114µm.

TABLE III. Post-layout simulation result of neural amplifiers.

Parameters	Performance
N	6.67
Open loop PM of OTA (°)	5
Input referred noise (μV_{rms})	5.38
Bandwidth (Hz)	331-7.1k
CMRR (1-7kHz) (dB)	>43
PSRR (1-7kHz) (dB)	>81
OTA Bias current (µA) @3V	1
Close loop gain (dB)	40
C_L (F)	1.8p
Area (mm^2)	0.058

TABLE IV. Comparison of neural amplifier characteristics

	Process (µm)	Power (µW)	Gain (dB)	Supply (V)	Area (mm^2)	NEF	FoM
[3]	1.5	40	40	5	0.16	4	12.8
[4]	0.5	7.7	40	2.8	0.16	2.67	3.19
[5]	0.35	4.2	34	3	0.02	4.6	1.26
[10]	0.18	7.9	39.4	1.8	0.063	3.35	1.28
This work	0.5	**3**	40	3	0.058	**2.51**	**1.11**

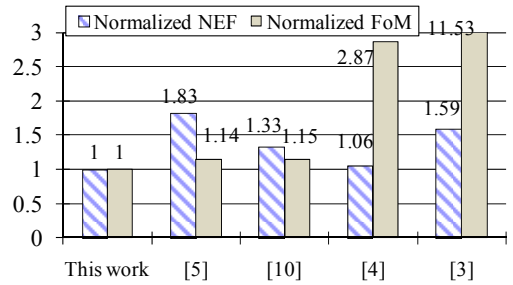

Fig. 6. Normalized NEF and FoM comparison.

REFERENCES

[1] A. Kamboh, M. Raetz, K. Oweiss, A. Mason "Area-Power Efficient VLSI Implementation of Multichannel DWT for Data Compression in Implantable Neuroprosthetics," IEEE Trans. Biomedical Circ. Systems, vol. 1, no. 2, pp. 128-135, Jun. 2007.

[2] R. H. Olsson, III, D. L. Buhl, A. M. Sirota, G. Buzsaki, and K. D. Wise, "Band-Tunable and Multiplexed Integrated Circuits for Simultaneous Recording and Stimulation with Microelectrode Arrays," IEEE Trans. Biomedical Eng., vol. 52, no. 7, pp. 1303-1311, Jul. 2005.

[3] R. R. Harrison, , and C. Charles, "A Low-Power Low-Noise CMOS Amplifier for Neural Recording Applications," IEEE J. Solid-State Circuits, vol. 38, no. 6, pp. 958-965, Jun. 2003.

[4] W. Wattanapanitch, M. Fee, and R. Sarpeshkar, "An Energy-Efficient Micropower Neural Recording Amplifier," IEEE Trans. Biomedical Circ. Systems., vol. 1, no. 2, pp. 136-147, Jun. 2007.

[5] J. N. Y. Aziz, K. Abdelhalim, R. Shulyzki, G. Roman, B. L. Bardakjian, M. Derchansky, *et al.*, "256-Channel Neural Recording and Delta Compression Microsystem with 3D Electrodes," IEEE J. Solid-State Circuits, vol. 44, no. 3, pp. 995-1005, Mar. 2009.

[6] W. M. C. Sansen,. Analog Design Essentials. Dordrecht, The Netherlands: Springer, 2006, pp. 144-146, 155-155.

[7] M. S. Fee, P. P. Mitra, and D. Kleinfeld, "Variability of Extracellular Spike Waveforms of Cortical Neurons," J. of Neurophysiology, vol. 76, no. 6, pp. 3823-3833, Dec. 1996.

[8] M. C. Chae, W. L, and M. Sivaprakasam. "Design Optimization for Integrated Neural Recording Systems," IEEE J. Solid-State Circuits, vol. 43, no.9, pp. 1931-1939, Sep. 2008.

[9] R. R. Harrison, "The Design of Integrated Circuits to Observe Brain Activity (Invited)," Proc. IEEE, vol. 96, no. 7, pp. 1203-1216, Jul. 2008.

[10] V. Majidzadeh, A. Schmid, and Y. Leblebici, "Energy Efficient Low-Noise Neural Recording Amplifier With Enhanced Noise Efficiency Factor," IEEE Tran. Biomedical Circ. Systems., vol. 5, no. 3, pp 262-271, Jun. 2011.