An OTA Gain Enhancement Technique for low power biomedical applications

R Nagulapalli, K Hayatleh, S Barker, S Zourob, N Yassine

The performance requirement of an operational trans-conductance amplifier (OTA) for the high gain and low power neural recording frontend has been addressed in this paper. A novel split differential pair technique is proposed to improve the gain of the OTA without any additional bias current requirements. The design demonstrates a significant performance enhancement when compared to existing techniques, such as gain-boosting and recycling. A qualitative and quantitative treatment is presented to explore the impact of the split ratio on the performance parameters of gain, bandwidth, and linearity. A prototype implemented in TSMC 65nm CMOS technology achieved 68dB open loop-gain (13dB higher than the conventional circuit) and a 17KHz 3-dB bandwidth. A linearity of -62dB has been achieved with 7mV pk-pk signal at the input. The circuit operates from a 1V supply and draws 0.6uA static current. The prototype occupies 3300um² silicon area.

Key Words: OTA; low power; CMOS; biomedical applications

I. INTRODUCTION

There has been a significant rise in the use of integrated circuit based processing in neuroscience and neuroprosthetic applications [1]. To cope with current diagnostic requirements, the neural recording systems should be fully implantable to record signals from freely moving subjects. It is difficult to implant an integrated circuit and battery due to size and biocompatibility issues. This means that the power needs to be supplied from outside the body, through inductive coupling or energy harvesting [2][3]. Another consideration is that even at moderate levels any heat generated may cause necrosis, hence the devices need to consume as little power as possible. To enable simultaneous recording multiple electrodes are required, which complicates the frontend electrical design because several parallel channels need to be integrated with minimal crosstalk [4].

Fig. 1 shows a multi-electrode neural recording system. Electrodes are used to capture bio-electrical signals which are then fed to the electronic part of the recording system. These electrical signals have a large DC offset voltage, due to DC leakage current in the order of 100pA, and the high impedance at the interface. Neural signals are typically in the frequency range of 1Hz-10KHz [5]. These signals are very low amplitude, varying from 0.2mV to 5mV depending on the electrode size, and hence are very susceptible to background noise. For this reason the front end Low Noise Amplifier (LNA) should have very low input referenced noise, so as not to compromise the system sensitivity. It should have a high open loop DC gain to minimize the settling error, and have an acceptable unity gain bandwidth in order to process high-frequency signals up to 10KHz [5]. As shown in Fig. 1, the front end LNA drives a variable gain amplifier (VGA) to adjust the signal amplitude before the analogue to digital conversion, and subsequent digital signal processing. To process signals from several channels, a MUX is needed at the output to eliminate the need for multiple analogue to digital converters (ADCs). The output from the ADC can then be fed to a digital signal processing system. In summary, the challenges of neural low noise amplifier design are low noise, low power, high gain, and small form factor. Typically signals take ~2ms to reach the digital end from the biological end (depends on the system bandwidth) with ~5% magnitude error (depends on the system gain). In this paper we are presenting a method to increase the gain without changing the bandwidth, hence magnitude error will decrease while keeping the time to reach the digital end constantly. With the proposed technique, we achieved 13dB improvement in the system gain, hence magnitude error will decrease to ~1.21% from 5%. This paper focuses on how to increase the gain without requiring any additional power consumption. The rest of the paper is organized as follows. Section-II describes the LNA design challenges and existing techniques, section-III explains the proposed split differential pair technique, and finally, section-IV summarizes the simulation results of the prototype.
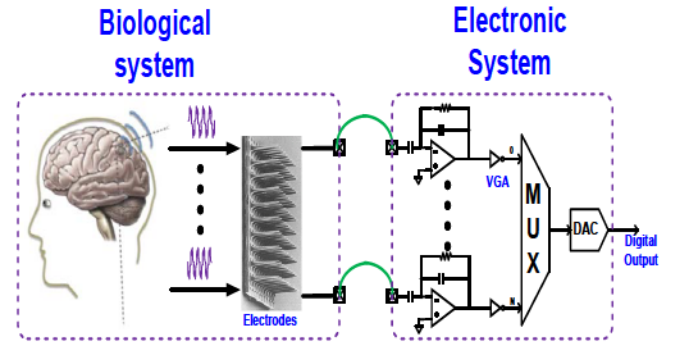

Fig. 1. Multi-Electrode Neural recording system.

II. NEURAL AMPLIFIER

The previous section explained the need for a high gain amplifier in the front-end to amplify the weak signal, before it is fed to the digital processing section. From the open loop and closed loop architectures existing in the literature, a closed loop architecture has been chosen because of its constant gain across PVT (process, voltage, temperature) variations and its superior linearity due to the negative feedback. The architecture of this amplifier is similar to the implementation shown in [6]. Fig. 2 depicts a capacitive feedback inverting amplifier, with C_f and C_{in} as feedback capacitors. The amplifier should be able to reject the DC offset at the input, so an AC coupling capacitor is necessary. Any negative feedback amplifier should also have DC feedback for maintaining the proper operating point. Therefore there is a need to place a high value resistor in parallel with C_f. M_{P1} and M_{P2} form a sub-threshold region based pseudo resistor (R_b). The amplifier shown here has a band-pass frequency response. The lower cut-off frequency is determined by the RC feedback network, and the upper cut-off frequency is determined by the amplifier unity gain bandwidth (UGB). The OTA non-inverting terminal has been arranged with the same feedback capacitance to maintain the symmetry, which could improve the CMRR and minimise the offset [6].
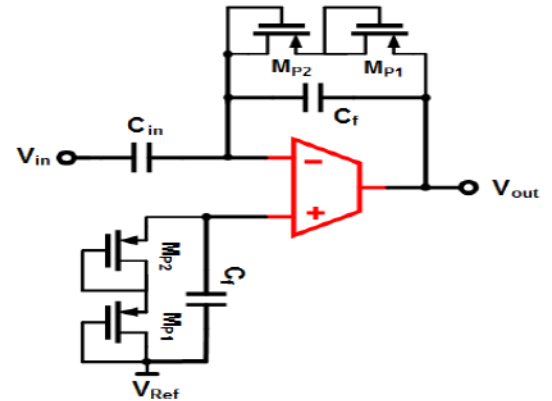

Fig. 2. Multi-Electrode Neural recording system.

Let us say C_p is the input parasitic of the OTA, G_m is the effective transconductance of the OTA, and R_o is the output impedance of the OTA. The feedback factor of the amplifier (β) can therefore be expressed as $\frac{C_f}{C_f+C_p+C_{in}}$. Loop-gain can be expressed as

$$LG(s) = \frac{\beta G_m R_o}{1+SR_o C_{Leff}} \qquad (1)$$

Where C_{Leff} is the effective load capacitance, including the load capacitance and fraction of the capacitance as it appears from the feedback network. This can be expressed as

$$C_{Leff} = C_L + \frac{C_f(C_P+C_{IN})}{C_f+C_P+C_{IN}} \qquad (2)$$

The closed loop gain of the amplifier can be expressed as

$$\frac{V_{out}(S)}{V_{in}(S)} = Idealgain \frac{1}{1+\frac{1}{LG(S)}} = -\frac{C_{in}}{C_f} \frac{1}{1+\frac{SC_{Leff}}{\beta G_m}} \frac{SR_bC_f}{1+SR_bC_f} \quad (3)$$

From equation (3) above, the midband gain can be determined by the ratio C_{in}/C_f. Interestingly, the midband gain is independent of the input parasitic capacitance C_P due to the virtual ground principle of the OTA [7]. The lower cutoff frequency is $1/R_bC_f$, and the upper cutoff frequency is $\beta G_m/C_{Leff}$. In the above analysis the electrode resistance (R_s) has been neglected, because it is very small compared to the feedback resistor (R_b), and it creates a high frequency zero $1/R_SC_{in}$, which is typically much higher than the amplifier bandwidth. Noise contributors in the amplifiers are the pseudo resistor and OTA, However, due to the extremely low bandwidth R_b contribution is much less, He we have to consider noise due to the OTA itself. Input referred noise can be expressed as

$$\overline{V_{n,AMP}^2} = \overline{V_{n,OTA}^2} \frac{(C_{in}+C_f+C_P)^2}{C_{in}^2} \quad (4)$$

III. OTA SELECTION

In general, OTA gain depends on the effective transconductance (g_m) and output impedance. There have been several techniques proposed to boost the OTA gain. To increase gain, either g_m or r_o should increase. g_m can only be increased by increasing the bias current for a given device size, hence this is the not a power efficient way. Cascoding is one of the popular output impedance boosting techniques without any additional power overhead, and cascode transistors don't contribute any significant noise [9].
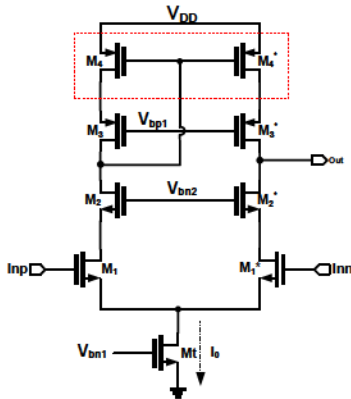

Fig. 3. Telescopic Opamp

But, unfortunately cascoding limits the device headroom, so the maximum number of cascodes will be limited by the linearity requirement, and this is not very suitable for low voltage implementation. Bult, et. al [10] proposed gain boosting by enhancing the cascode transistor g_m through an opamp will significantly improve the gain of the opamp. However, gain boosting adds additional power requirements because of the additional opamp required to bias the cascode transistor, and this technique is also prone to slow step response settling due to the closely spaced pole-zero doublets.
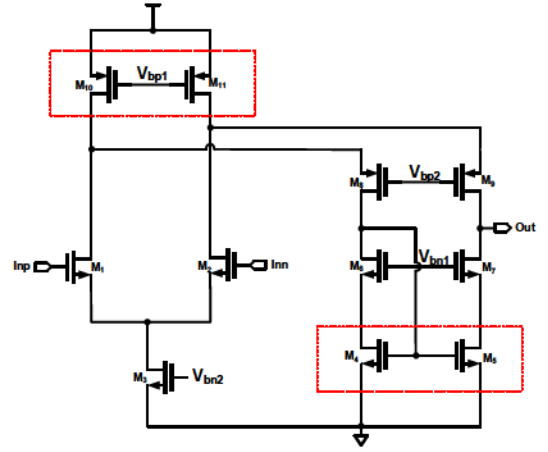

Fig. 4. Folded cascode opamp

Frequency compensation also becomes challenging, so this may not be a suitable technique for the present application. The multi-stage amplifier has been proposed and very commonly used for achieving the gain in the range of 100dB, but this increases the power-consumption, makes compensation very challenging, and demands a large silicon area, hence this is not suitable for low power applications[8][11]. The above techniques can be implemented in either telescopic or folded cascode opamps. Several researchers introduced several derivatives of the above techniques. Recently, current reusing and recycling folded cascode opamps have been demonstrated, which re-use the opamp current to enhance the effective g_m. This technique increases the opamp gain without any requiring additional power [12][13]. Unfortunately, the recycling technique can't be used for telescopic arrangements. It is only useful for folded-cascode. However, the folded cascode opamp is not a very good candidate for bio-medical applications due to high input referred noise and offset, which have been explained as follows. Fig. 3 and Fig. 4 show the telescopic-Cascode and folded-Cascode opamps respectively. While comparing the input referred noise of both architectures, the folded Cascode results in higher noise, because in a Cascode opamp only the input differential pair and PMOS cascode load (M_4) contribute to it, whereas in the folded Cascode opamp, apart from the differential pair, the PMOS Cascode load, and folding current mirror (M_{10}, M_{11}) contribute significant noise. Also, a folded current mirror carries much higher current, so it is noisier than all the other devices. Generally, in any opamp, significant input referred offset is originated from the input differential pair and current mirrors. As shown in fig. 3, telescopic the opamp offset generated by M_1,M_4 pairs can be expressed as follows.

$$\sigma_{vos}^2 = \sigma_{VT1}^2 + \left(\frac{g_{m4}}{g_{m1}}\right)^2 \sigma_{VT4}^2 \quad (5)$$

In a folded cascode opamp there is an additional curent mirror due to the folding node (M_4,M_5), and the offset contributed by M_1,M_4,M_5 is given as follows.

$$\sigma_{vos}^2 = \sigma_{VT1}^2 + \left(\frac{g_{m4}}{g_{m1}}\right)^2 \sigma_{VT4}^2 + \left(\frac{g_{m5}}{g_{m1}}\right)^2 \sigma_{VT5}^2 \quad (6)$$

Where σ_{VT1}^2 is the threshold mismatch coefficient of the respective transistors. The beta mismatch has been omitted because that complicates the analysis, and does not affect the conclusion. Clearly, from equations (5) & (6), the folded cascode is worse when compared to telescopic cascodes. A sample schematic simulation shows the folded cascode is 35% worse. Fig. 5 shows the MC simulation results of both amplifiers. These are similar designs just to prove the offset claim. For both of these example designs the input differential pair size is the same to ensure a reasonable comparison.
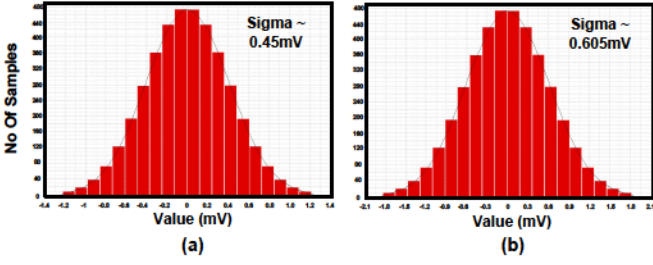
Fig. 5. Monte Carlo results (a) Telescopic (b) folded cascode

Fig. 7. G_m and g_{ds} variation with bias current
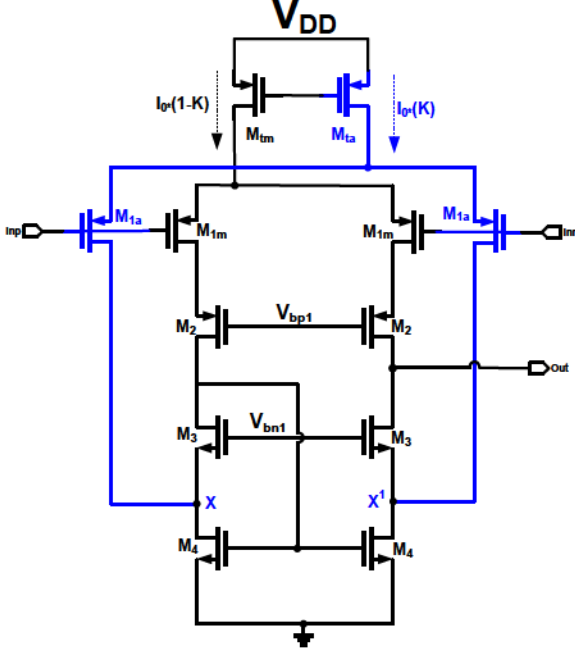
IV.   4. PROPOSED OTA TECHNIQUE

Fig. 6. Split Cascode opamp (1st stage of Neural amp)

Fig. 6 shows the proposed split differential pair based opamp. This technique relies on the principle that the output impedance will increase when the bias current decreases [7]. The simplified square law drain current model of MOS transistors is given by

$$I = \frac{1}{2} u_n C_{ox} \frac{W}{L} \left(V_{gs} - V_{th} \right)^2 (1 + \lambda V_{ds}).$$

Here W is channel width, L is channel length, μ_n is electron mobility, C_{ox} is gate oxide capacitance, and λ is the channel length modulation factor. The small signal parameters can be expressed as follows.

$$g_m = \sqrt{2u_n C_{ox} I_d \frac{W}{L}} \quad \text{and} \quad g_{ds} = \lambda I_d$$

By increasing the bias current, g_m and g_{ds} will increase, so both changes which are in the same direction will not increase the gain up to a great extent. Fig. 7 shows the gm and g_{ds} variation with bias current. To increase the gain, g_m must increase and g_{ds} must be decreased. However, these changes cannot be achieved by simply changing the bias-current. For example, in the telescopic Cascode amplifier (Fig. 3), to increase the gain we would ideally increase the current in the differential pair (M_1 and $M*_1$), since its g_m is important, and then decrease the current through Cascode devices (M_2, M_3, M_4) to improve g_{ds}. The approach presented in this paper is to increase the DC gain without affecting the Unity Gain Bandwidth. This technique improves g_{ds} without affecting g_m.
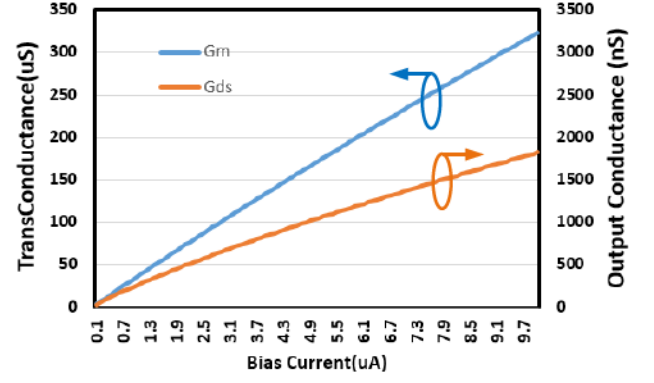
In the proposed split Cascode amplifier, we have chosen to keep the total current constant while decreasing the current through the Cascode devices load to eliminate the extra current requirement. As shown in Fig. 6, the tail current source and differential pair are split into two sections, with currents in the ratios of (1-k):k. The main Cascode differential pair (M_{1m}) carries (1-K) times the main current, and auxiliary differential pair (M_{1a}) will carry K times current. To improve g_{ds}, the current flowing through the Cascode must decrease. This means the part of the differential pair which is carrying less current will be connected to the Cascode loads. To preserve the bandwidth the effective g_m should be constant, so the auxiliary differential pair is connected to the low impedance node x such that the effective g_m is constant. As the current through the Cascode devices has decreased by (1-k), g_{ds} will also decrease by a factor of (1-k). The effective g_m and voltage gain of the proposed amplifier are expressed as follows.

$$g_m = g_{m1a} + g_{m1m}$$

$$A_v \sim \frac{g_{m1a} + g_{m1m}}{\frac{g_{ds1m} g_{ds2}}{g_{m2}} + \frac{g_{ds3} g_{ds4}}{g_{m3}}} \tag{7}$$

$$\frac{\text{Gain of the proposed}}{\text{Gain of the tradiational}} = \frac{1}{(1-K)^2} \tag{8}$$

The loading of M_{1a} is neglected in the equation (8) because the node x is a low impedance node compared with M_{1a}. Therefore, the current splitting ratio (K) should obey the following equation, otherwise there will be a large discrepancy between the simulated and calculated gain.

$$\sqrt{2u_n C_{ox} K I_d \frac{W}{L}} > \lambda (1-K) I_d. \tag{9}$$

Fig. 8 shows the simulated and calculated gain of the opamp, which concur within 2dB error at a smaller K value, but once K is greater than 0.5 the deviations become large because M_3 can no longer carry significant current.
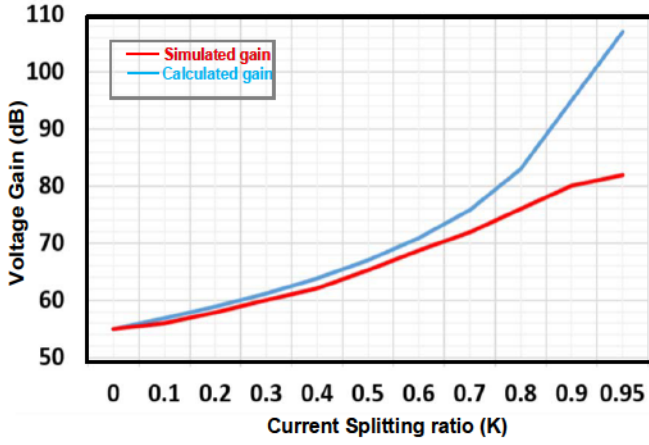
Fig. 8. Voltage gain versus current splitting ratio (K)

V. IMPLEMENTAION DETAILS

A prototype of the neural amplifier shown in Fig. 1 has been implemented in 65nm CMOS technology. A two stage OTA used here, with the first stage as proposed, and the second stage a traditional common source amplifier. A closed loop gain of 30dB has been achieved with a chosen 3-dB bandwidth of 17KHz. This bandwidth is good enough to amplify any bio-medical signal without having any significant filtering effect. The ratio of feedback capacitors is 30, with C_{in}=10pF and C_f=0.333pF. Miller compensation has been used to make the opamp stable. The current splitting ratio (K) has a negative impact on linearity, because as K increases the Cascode device bias currents decreases, hence the transistors' operating point changes significantly, and the percentage of signal current to bias current increases. Hence, linearity will decrease.
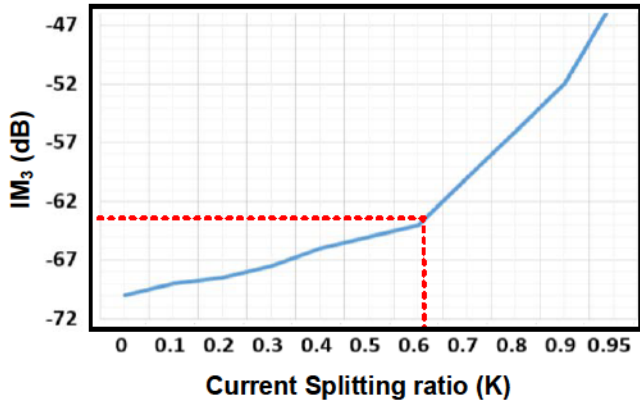

Fig. 9. Linearity versus current splitting ratio (K)

A two-tone test was carried out with 9.5KHz and 10.5KHz signals, and the 3rd order intermodulation products (IM3) at 11.5KHz have used a figure of merit of linearity [12]. IM3 is -69dB when K=0 and dropped to -46dB when K=0.95 (due to the very low bias current in the Cascode devices) as shown in Fig. 9. As a balance between linearity degradation and gain boosting we have chosen the current splitting ratio of 0.6.
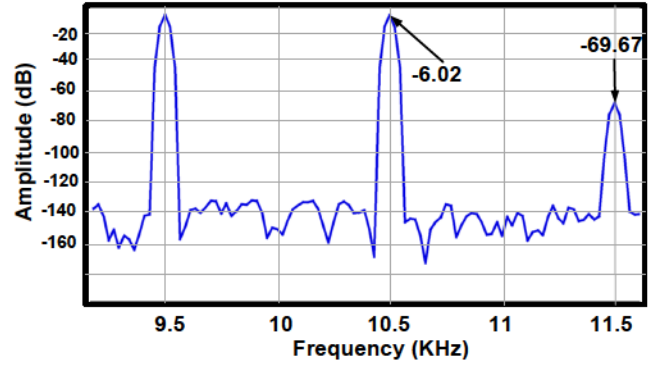

Fig. 10. Two tone FFT spectrum@k=0.6

Fig. 10 shows the FFT of the two tone test when the current splitting ratio is 0.6. Fig. 11 shows the frequency response of the OTA first stage. A conventional technique exhibits 55dB, and the proposed technique achieves 68dB, thus providing an almost 13dB improvement without any additional power requirement and area overhead. From this figure it is very clear that Unity Gain Bandwidth is very close to the original design. To achieve a 17KHz 3-dB bandwidth with the feedback factor of 30 (ratio of capacitors), the open loop gain should have unity gain bandwidth of 570KHz.
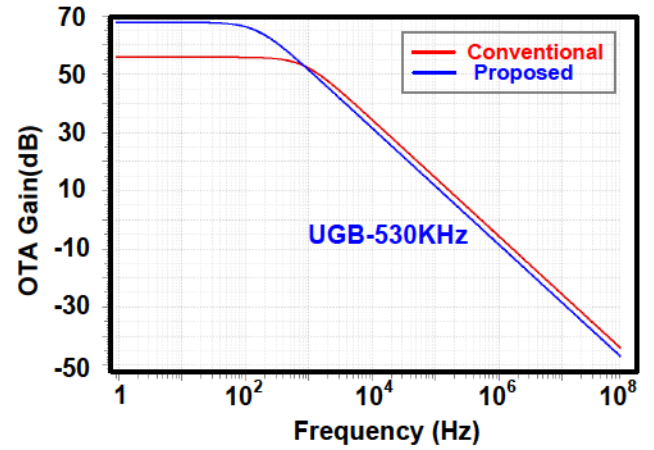

Fig. 11. Frequency response of the first stage

Fig. 12 shows the closed loop frequency response of the neural amplifier, as explained in the introduction. The feedback capacitor, C_f, and pseudo resistor form a high pass filter, with a cutoff frequency of 220Hz. This number is large enough to isolate the DC mismatch and small enough to pass band signals.
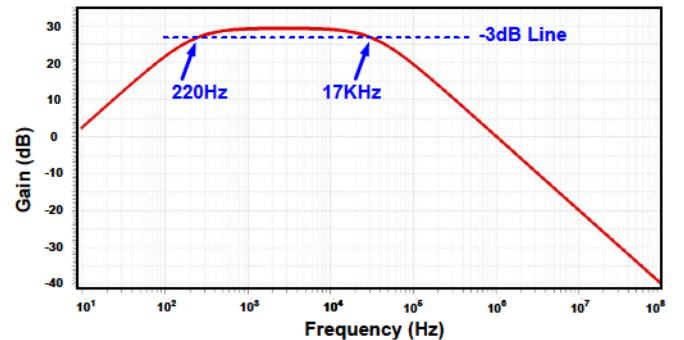

Fig. 12. Neural Amplifier Frequency Response

Fig. 13 shows the input referred spectral density. The RMS voltage is 7.81uV, which is obtained by integrating the spectral density from 10Hz to 1MHz. A dominant noise contributor is the flicker noise of the input differential pair. The achieved noise level is sufficient to detect all biological signals [6].
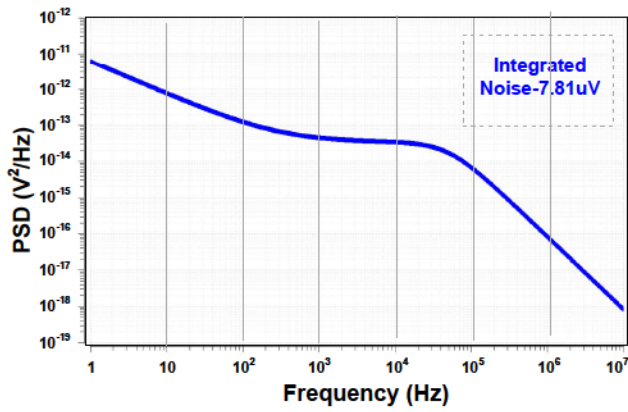
4

Fig. 13. Input referred Spectral Density

Fig. 14 shows the layout of the proposed amplifier. Every transistor has been laid with proper care with respect to mismatch, like extensive interleaving for the differential pair with the ABBA layout pattern. Special care has been taken to avoid well proximity effect (WPE) and shallow trench isolation (STI) effect by adding enough dummies for each device, and keeping transistors away from the nwell [14]. Table-II summarizes the performance parameters.
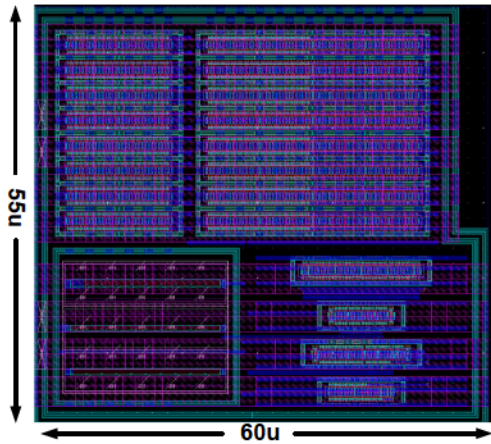

Fig. 14. Layout of the proposed Circuit.

TABLE II
Performance Summary

Parameter	Value	Units
Minimum Supply Voltage	1	V
Current consumption	0.6	uA
Closed Loopgain	30	dB
3-dB bandwidth	17	KHz
Linearity (IM$_3$)	-63	dB
Integrated Noise	7.81	uV
CMRR	65	dB
Technology	65	nm
Area	3300	um^2

VI. CONCLUSION

An OTA voltage gain enhancement technique has been proposed, which is suitable for bio-medical applications, specifically a neural recording amplifier. The proposed design depends on the splitting of the differential pair rather than increasing the power, as in previous techniques. We have demonstrated the impact of splitting the ratio on the performance of the OTA, and have given instructions for optimal performance. We have explained why the telescopic Cascode OTA is superior to the folded Cascode OTA from the input referred offset voltage and noise point of view. A transistor level circuit has been implemented in 65nm CMOS technology, post layout simulations show >4X voltage gain improvement while drawing 0.6uA from the 1V power supply.

VII. REFERENCES

1. M. A. L. Nicolelis, "Actions from thoughts," Nature, vol. 49, pp. 403–407, 2001.
2. N. Neihart and R. Harrison, "Micropower circuits for bidirectional wireless telemetry in neural recording applications." IEEE Trans. Biomed. Eng., vol. 52, no. 11, pp. 1950–1959, November 2005.
3. M. S. Chae, W. Liu, and M. Sivaprakasham, "Design optimization for integrated neural recording systems." IEEE J. Solid-State Circuits, vol. 43, no. 9, pp. 1931–1939, September 2008.
4. T. Jochum, T. Denison, and P. Wolf, "Integrated circuit amplifiers for multi-electrode intracortical recording," J. Neural Eng., January 2009.
5. R. Harrison, "A low-power integrated circuit for adaptive detection of action potentials in noisy signals," in Proc. IEEE lnt. Conf. Eng. Med. Biol. Soc., Sep. 2003, pp. 3325–3328.
6. V. Chaturvedi, B. Amrutur, "An area-efficient noise-adaptive neural amplifier in 130 nm CMOS technology", *IEEE J. Emerg. Sel. Topics Circuits Syst.*, vol. 1, no. 4, pp. 536-545, Dec. 2011
7. T.H. Lee, The Design of CMOS Radio-Frequency Integrated Circuits, Cambridge University Press, United Kingdom, 1998.
8. W. Wattanapanitch, R. Sarpeshkar, "A low-power 32-channel digitally programmable neural recording integrated circuit", *IEEE Trans. Biomed. Circuits Syst.*, vol. 5, no. 6, pp. 592-602, Dec. 2011.
9. E. Sackinger, W. Guggenbuhl, "A high-swing high-impedance MOS cascode circuit", *IEEE J. Solid-State Circuits*, vol. 25, no. 1, pp. 289-298, Feb. 1990.
10. K. Bult, G. J. G. M. Geelen, "A fast-settling CMOS op amp for SC circuits with 90-dB DC gain", IEEE J. Solid-State Circuits, vol. 25, pp. 1379-1384, Dec. 1990.
11. S. Pernici, G. Nicollini, R. Castello, "A CMOS Low-distortion fully differential power amplifier with double nested miller compensation", *IEEE J. Solid-State Circuits*, vol. 28, pp. 758-763, July 1993.
12. Assaad, R., & Silva-Martinez, J. (2009). 'The recycling folded cascode: A general enhancement of the folded cascade amplifier'. IEEE Journal of Solid-State Circuits, 44(9), 2535–2542.
13. Yan, Z., Mak, PI. & Martins, R.P. Analog Integr Circ Sig Process (2012) 71: 137. https://doi.org/10.1007/s10470-011-9762-y
14. P. G. Drennan, M. L. Kniffin, and D. R. Locascio, "Implications of proximity effects for analog design," in Proc. IEEE Custom Integrated Circuits Conf., Sep. 2006, pp. 169–176.