Advances in Microelectronics for Implantable Medical Devices

Andreas Demosthenous

Department of Electronic and Electrical Engineering, University College London, Torrington Place, London WC1E 7JE, UK

Correspondence should be addressed to Andreas Demosthenous; a.demosthenous@ucl.ac.uk

Received 2 December 2013; Accepted 18 February 2014; Published 29 April 2014

Academic Editor: Sebastian Hoyos

Copyright © 2014 Andreas Demosthenous. This is an open access article distributed under the Creative Commons Attribution License, which permits unrestricted use, distribution, and reproduction in any medium, provided the original work is properly cited.

Implantable medical devices provide therapy to treat numerous health conditions as well as monitoring and diagnosis. Over the years, the development of these devices has seen remarkable progress thanks to tremendous advances in microelectronics, electrode technology, packaging and signal processing techniques. Many of today's implantable devices use wireless technology to supply power and provide communication. There are many challenges when creating an implantable device. Issues such as reliable and fast bidirectional data communication, efficient power delivery to the implantable circuits, low noise and low power for the recording part of the system, and delivery of safe stimulation to avoid tissue and electrode damage are some of the challenges faced by the microelectronics circuit designer. This paper provides a review of advances in microelectronics over the last decade or so for implantable medical devices and systems. The focus is on neural recording and stimulation circuits suitable for fabrication in modern silicon process technologies and biotelemetry methods for power and data transfer, with particular emphasis on methods employing radio frequency inductive coupling. The paper concludes by highlighting some of the issues that will drive future research in the field.

1. Introduction

Neuroengineering, the application of engineering techniques to understand, repair, replace, enhance, or otherwise exploit the properties of neural systems, is a topic that is currently generating considerable interest in the research community. The nervous system is a complex network of neurons and glial cells. It comprises the central nervous system (brain and spinal cord) and the peripheral nervous system. Injuries or diseases that affect the nervous system can result in one of the most devastating medical conditions. Conditions, such as stroke, epilepsy, spinal cord injury, and Parkinson's disease, to name but a few, as well as more general symptoms such as pain and depression, have been shown to benefit from implantable medical devices. These devices are used to bypass dysfunctional pathways in the nervous system by applying electronics to replace lost function.

The first implantable medical devices were introduced in the late 1950s with the advent of the heart pacemaker [1, 2] and subsequently the cochlear implant [3, 4]. Both have restored functionality for hundreds of thousands of patients. A pacemaker uses electronics and sensors to continuously monitor the heart's electrical activity and when arrhythmia is detected, electrical stimulus is applied to the heart (via electrodes) to regulate its speed. A cochlear implant uses electronics to detect and encode sound and then stimulate the auditory nerve to enable deaf individuals to hear. Thanks to remarkable advances in microelectronics, electrode technology, packaging, and biomedical signal processing, active implantable medical devices have developed into advanced systems, employing wireless telemetry for transmission of data and sometimes power.

The success of cochlear implants has inspired the development of implantable devices for restoring other basic human sensations. Visual prosthesis translates camera input into electrical stimulation to the visual nervous system to create pixelized vision [5, 6], while vestibular prosthesis connects motion sensors to vestibular nerves to restore balance sensation [7, 8]. Another example is deep brain stimulation (DBS) that has been shown to provide therapeutic benefits for otherwise treatment-resistant neurological disorders such as Parkinson's disease, tremor, and dystonia [9]. In current clinical DBS systems, high frequency stimulation (>100 Hz) produced by a pulse generator (stimulator) is continuously...
applied via deep brain electrodes to the targeted tissue area in the brain (Figure 1). The characteristics of these pulses (e.g., frequency, pulse duration, and intensity) are programmed into the stimulator (implanted in the chest area) and adjusted via an external programmer. DBS emerged from heart pacemaker technology and hence is also called a brain pacemaker. Current developments are focused on the design of cranially-mounted inductively-powered DBS systems [10] to reduce the length of the leads from the stimulator to the electrodes and novel stimulation techniques using high-density segmented electrodes, which can enable current-steering and electric field shaping capability [11, 12]. In addition, recent papers have reported the development of prototype closed-loop (i.e., sense and stimulate) neuroprosthetic devices for applications such as vestibular prosthesis [13] and epilepsy [14, 15]. In the case of epilepsy, the device uses implantable multielectrode arrays and amplifiers to record electrical signals from neurons in the brain. The recorded data is then processed to extract important events, which, for example, predict the onset of an epileptic seizure, and electrical stimulation is applied to inhibit the attack.

This paper provides a review of advances in microelectronics for implantable medical devices and systems. The focus is limited to neural recording and stimulation circuits suitable for fabrication in modern silicon process technologies and biotelemetry methods for power and data transfer. After this introductory part, Section 2 reviews several techniques and circuit topologies for neural recording, including methods for on-chip data reduction to reduce the bandwidth requirements of the wireless transmission link. Section 3 covers advances in neural stimulation circuits and Section 4 discusses biotelemetry methods including wireless power and data transmission by inductive coupling. Finally, conclusions are drawn in Section 5 including some suggestions for future research.

2. Neural Amplifiers

Neural signals are low frequency and low amplitude signals. For example, the amplitude of the electroneurogram (ENG) recorded with implanted cuff electrodes [16, 17] is typically in the region of 1 μV, with most energy concentrated between 300 Hz and 5 kHz. Even when recording neural activity with penetrating microelectrodes such as the Utah Electrode Array [18, 19] or the NeuroNexus penetrating probes (http://www.neuronexustech.com/), the recorded neural action potentials often have amplitudes of only a few tens of microvolts. Hence, circuits for neural signal amplification must have low noise performance and additionally low power consumption so that battery life is prolonged, especially in implantable systems (e.g., implantable loop recorder for long-term monitoring of the heart’s electrical activity [20]). In addition, front-end neural amplifiers are required to reject electrode offsets or common-mode interference. Both clock-based and continuous-time techniques have been used in the design of neural amplifiers.

2.1. Clock-Based Techniques. The noise in CMOS transistors is usually dominated by flicker (1/f) noise up to relatively high frequencies of the order of several tens of kHz [21]. This is particularly troublesome for the design of low frequency, low noise analog circuits. Typically, p-channel transistors have less 1/f noise than n-channel transistors. Various clock-based techniques have been developed to reduce the effects of 1/f noise. Noise reduction based on physical effects (switched biasing), chopper modulation and autozeroing, are amongst these techniques.

The switched biasing technique (Figure 2) reduces the 1/f noise of a MOS transistor by cyclically increasing and decreasing its gate bias so that the device alternates between strong inversion and accumulation [22]. The transistor noise is modulated by the switching signal. The switching operation is represented as a multiplication of the 1/f noise current.
\( I_{\text{noise}} \) with the switching signal \( m(t) \). For a square-wave signal with 50% duty cycle

\[
m(t) = \frac{1}{2} + \frac{2}{\pi} \sin(2\pi f_s t) + \frac{2}{5\pi} \sin(6\pi f_s t)
\]

\[
+ \frac{2}{5\pi} \sin(10\pi f_s t) + \cdots ,
\]

where \( f_s \) is the switching frequency. If \( f_s \) is set sufficiently high, the baseband noise reduces by half and any modulation effects represented by the sine terms in (1) remain outside the bandwidth of interest and can be removed by filtering. Switched biasing improves noise performance at low frequencies but requires a high speed clock applied to the gate (or bulk) of the transistor with potential problems due to charge feedthrough and additional noise originating from the driver circuit. In addition to reducing the intrinsic \( 1/f \) noise, the switched biasing technique reduces power consumption [22].

A switched biasing amplifier demonstrating input-referred noise reduction at low frequencies (<100 Hz) is described in [23]. It uses two operational amplifiers (opamps) configured as buffers linked via resistors. The first stage of each opamp is an operational transconductance amplifier (OTA) circuit of the type shown in Figure 2 where switched biasing is applied to the p-channel transistor supplying the tail current \( I_{\text{bias}} \). Further noise reduction could be achieved by cycling the voltage bias applied to the bulk terminals of the differential pair input transistors.

The chopper technique is also based on signal modulation. The technique enables the design of amplifiers with high common-mode rejection ratio (CMRR) performance. A block diagram of the chopper amplifier is shown in Figure 3 [24]. Before amplification, the amplifier input signal is modulated by a square-wave signal of frequency \( f_s \) that is much higher than the baseband frequencies of interest. The chopping signal may be represented as in (1). The upconverted signal is then amplified and bandpass filtered. The modulated signal spectrum is located at frequencies higher than the \( 1/f \) noise corner. After amplification, the signal is converted back to baseband by multiplication with the same modulation waveform used for upconversion. Lowpass filtering restores the desired signal. The technique reduces both \( 1/f \) noise and amplifier dc offset voltages, but the noise performance is ultimately limited by the noise floor of the amplifier. In addition, practical nonidealities, including the finite amplifier bandwidth, can lead to signal distortion.

Several integrated neural amplifiers employing the chopper technique have been described for recording from both implanted and surface electrodes [25–30]. The design in [28] employs an ac-coupled chopping technique to reject electrode offsets and achieve low noise performance. The concept of this technique is shown in Figure 4. The system consists of a feedforward stage and a feedback stage. To suppress the \( 1/f \) noise of the instrumentation amplifier (IA), the feedforward stage employs an input chopper and an output chopper. To eliminate the electrode offset, the feedback stage employs a lowpass OTA stage followed by a chopper stage. The operation of the circuit is as follows. The input differential electrode offset \( V_{\text{el,offset}} \) is modulated by the input chopper and appears across resistor \( R_1 \). By action of the current-feedback the current through \( R_1 \) is copied to \( R_2 \) and defines the output voltage after demodulation by the output chopper. The lowpass OTA stage filters the dc component of the output and converts it into current. The OTA output current is in turn modulated by a chopper stage. In the steady state, the current supplied by the OTA is \( V_{\text{el,offset}} / R_1 \). As a result, no current is supplied by the IA and the current passing through \( R_2 \) is zero, so the output \( V_{\text{out}} \) is zero. The neural amplifier in [29] uses chopper modulation and switched-capacitor techniques to reduce the \( 1/f \) noise and achieve frequency tuning. The circuit is capable of simultaneous recording of extracellular unit spikes (action potentials) and local field potentials (low frequency signals in the 1 Hz to 100 Hz range). Another publication applies chopper stabilization with a distortion cancelling technique to the design of a front-end transimpedance amplifier for current-mode biosensors [30]. It achieves both low noise and low distortion performance at minimum current consumption. The cancellation technique reduces the distortion and, combined with chopping, substantially reduces the \( 1/f \) noise. For a current consumption of 50 \( \mu \)A, it is shown that the input-referred noise density without cancelling and chopping at 10 Hz is 29 pA/\( \sqrt{\text{Hz}} \) and reduces to 3 pA/\( \sqrt{\text{Hz}} \) when both techniques are employed. The total harmonic distortion for a peak current of 1 \( \mu \)A is –55 dB.
2.2. Continuous-Time Techniques. The noise reduction techniques described above require a clock generation circuit and thus suffer from potential problems associated with high frequency interference and clock feedthrough. In addition, high frequency switching circuits can increase the complexity and power consumption of the design. As an alternative, continuous-time techniques have been extensively used in the design of neural amplifiers. The classic circuit is the ac-coupled OTA-based neural amplifier with capacitive feedback shown in Figure 6 [34]. The circuit is built around a single stage OTA in CMOS technology. The ratio of capacitors $C_1$ and $C_2$ sets the midband gain of the bandpass response. The input is capacitively coupled through $C_1$, so any dc offset from the electrode-tissue interface is removed ($C_1$ should be made much smaller than the electrode impedance to minimize signal attenuation). Transistors $M_a$-$M_d$ implement MOS pseudoresistors with an extremely large incremental resistance ($>10^{12} \, \Omega$). This allows the cutoff frequency of the input high-pass filters (i.e., ac-coupled stage) to be set to the millihertz region. The lower cutoff frequency is set by the product of $C_2$ and the MOS pseudoresistor implemented by $M_a$ and $M_b$. The upper cutoff frequency is a function of the load capacitance ($C_L$), the OTA transconductance ($G_m$), and the midband gain ($G_1/C_2$). To reduce the effect of $1/f$ noise, the OTA input transistors should be p-channel devices with large gate areas. Numerous designs of neural amplifier based on the circuit in Figure 6 or with some variations (e.g., in the realization of the pseudoresistors, use of fully-differential topology with one or two stage OTAs, use of current-reuse techniques to double the transconductance, and so forth [35–39]) have been reported in the literature, including commercial amplifier chips by Intan Technologies, LLC (http://www.intantech.com/). Methods for effective optimization of a recording channel in terms of its power consumption, input-referred voltage noise, silicon area, and technology used are discussed in [40]. The design of nanopower OTAs with enhanced linearity is presented in [41].

In the case of recording from a multielectrode array, the total power consumption of the amplifier array (as well as the silicon area) may be reduced by using the partial OTA sharing structure proposed in [42]. In this technique, each of the $n$ amplifiers in the array share the components corresponding to the reference electrode (i.e., pseudoresistors $M_c$ and $M_d$ and capacitors $C_1$ and $C_2$ connected to $V_{\text{ref}}$ in the amplifier in Figure 6). The silicon area is reduced as a benefit of sharing the bulky capacitor $C_1$. The improvement factor in terms of silicon area depends on the number of shared amplifiers. In general, for an electrode array size of $n$, the OTA sharing technique allows an area saving of $(n - 1)/(2n) \times 100\%$, a power saving of $(n - 1)/(2n) \times 100\%$, and an improvement in noise efficiency factor (NEF) of $[1 - \sqrt{(n+1)/2n}] \times 100\%$, compared to the conventional architecture (i.e., using $n$ neural amplifiers). The NEF is a dimensionless figure of merit (noise-power tradeoff) used to compare different neural amplifier designs. It is defined as [43]:

$$\text{NEF} = \frac{V_{\text{in},\text{rms}}}{\pi \cdot U_T \cdot 4kT \cdot BW}$$ (2)
Advances in Electronics 5

MOS pseudoresistor

Figure 7: Capacitive feedback amplifier with low input capacitance [44].

where $V_{\text{in,rm}}$ represents the integrated input-referred noise, $I_{\text{tot}}$ is the total power consumed by the amplifier, $U_T$ is the thermal voltage, $k$ is Boltzmann’s constant, $T$ is the absolute temperature in Kelvin, and $BW$ is the −3 dB bandwidth of the amplifier. To include the supply voltage $V_{\text{DD}}$, the modified metric $\text{NEF}^2 \times V_{\text{DD}}$ is used [37].

To achieve low noise performance, a large neural amplifier gain is usually required. The conventional ac-coupled neural amplifier (Figure 6) often presents a large input load capacitance (typically 10–20 pF for a midband gain of around 40 dB) to the neural signal source, hence occupying a large silicon area. It suffers from the unavoidable tradeoff between input capacitance and chip area versus the amplifier gain. In the amplifier in Figure 7 [44], this tradeoff is limited by the replacing feedback capacitor with a clamped T-capacitor network. The diodes are used for discharge purposes. Compared to the conventional circuit this amplifier can achieve a given midband gain with less input capacitance, a higher input impedance and smaller silicon area. For a midband gain of 38.1 dB, a neural amplifier employing the topology in Figure 7 used a 1.6 pF input capacitance and a total silicon area of 0.056 mm² in 0.35-μm CMOS technology [44].

Electrode offset removal may also be implemented by means of active feedback loops. An example neural amplifier topology with active feedback for dc rejection is shown in Figure 8 [45]. It consists of a low noise OTA (OTA₁) with an active feedback circuit implemented by a second OTA (OTA₂) configured as a Miller integrator. The time-constant of the integrator is set by capacitor $C_1$ and the MOS pseudoresistor comprising $M_a$ and $M_b$. The midband gain of the amplifier is the same as the gain of OTA₁. The dominant pole of OTA₁ sets the amplifier’s lowpass cutoff and the common-mode voltage is set by voltage $V_{\text{ref}}$. Another example of a neural amplifier with an active feedback loop to bypass any dc offset current generated by the electrode-tissue interface is described in [46]. It predominantly makes use of current-mode circuit techniques.

A low noise neural amplifier for implanted cuff electrodes is described in [47]. The circuit schematic is shown in Figure 9. It consists of an input BiCMOS OTA ($Q_1$, $Q_2$, $M_1$, and $M_2$) terminated in the load resistor $R_1$, followed by a first-order bandpass filter (for bandwidth restriction). The upper cutoff frequency is set by the combination of resistor $R_2$ and capacitor $C_1$, while the lower cutoff frequency is set by capacitor $C_2$ with the series combination of transistors $M_6$ and $M_7$, the latter transistor pair forming a high value (~20 MΩ) active resistor. In addition to eliminating low frequencies below the pass-band of the input neural signal, the high-pass section of the bandpass filter also removes some of the low frequency flicker noise voltage tail and ensures a dc offset-free amplifier output ($V_{\text{out}}$). The dc bias voltages of $M_6$ and $M_7$ are provided by the diode-connected transistors $M_8$ and $M_9$, respectively, which are in turn biased by the dc current sources $I_{\text{b2}}$ and $I_{\text{b3}}$. Circuitry is also included ($M_3$–$M_5$ and $Q_3$–$Q_4$) to cancel the base currents of $Q_1$ and $Q_2$. This neural amplifier achieved a measured input-referred root mean square noise voltage of only 290 nV (noise bandwidth of 1 Hz–10 kHz). A variant of this circuit in CMOS technology using lateral bipolar devices for $Q_1$ and $Q_2$ is possible [48]. In general for a given target noise specification the use of lateral npn bipolar devices (available as parasitic devices in CMOS technology) tends to require a larger silicon area compared to using standard npn bipolar transistors in

Figure 8: Amplifier with active dc rejection [45].

Figure 9: A BiCMOS low noise amplifier [47].
BiCMOS technology (but BiCMOS technology might incur higher manufacturing costs). A bipolar transistor structure in CMOS technology featuring high matching characteristics is described in [49].

For biomedical front-ends requiring high CMRR performance and accurate gain setting the use of an IA is desirable. An IA may be realized using the classic three-opamp topology. However, the CMRR of the three-opamp IA depends on the matching of the resistors and the need for low output impedance amplifiers can increase power consumption. Another technique for IA design is to employ switched-capacitor circuits [50], but the fold over of noise above the Nyquist frequency can be a major limitation. A popular IA topology for integrated circuits is the current feedback technique. In a current feedback IA, the gain is accurately set by the ratio of two resistors and the CMRR does not rely on the matching of resistors. Figure 10 shows the simplified circuit schematic of a current feedback IA [51].

The input transconductor stage uses a simple current mirror load and current sink biasing. The sensing amplifier A serves to exactly balance the drain currents of transistors $M_{i1}$ and $M_{i2}$ by adjusting the complementary currents $I_1$ and $I_2$. A direct result of this is that the input differential voltage $V_{in}$ is forced across resistor $R_1$ and hence $M_{i1}$ and $M_{i2}$ of the input stage essentially act as a unity-gain buffer. Similarly, the high gain amplifier $B$ balances the drain currents of transistors $M_{o1}$ and $M_{o2}$ in the output transconductor stage. Since currents $I_1$ and $I_2$ are exact copies of $I_1$ and $I_2$, respectively, the output voltage $V_{out}$ appears across resistor $R_2$. Hence, the dc gain of the IA is given by the ratio $R_2/R_1$. Placing capacitor $C_2$ in parallel with $R_2$ creates a dominant pole, which sets the $-3 \, \text{dB}$ bandwidth of the IA. The CMRR and noise analysis of the circuit are described in [51]. The IA in [52] used the current feedback technique to achieve a CMRR of 99 dB and an input-referred noise of $0.68 \, \mu V \text{rms}$. It was developed to record neural signals from electrodes on the lumbo-sacral nerve roots which can be used to restore lower-body function to patients with paraplegia after spinal cord injury.

Table 1 compares various integrated neural amplifiers reported in the literature. From the table, it is observed that there is a relationship between current demand and input-referred noise. In general, the lower the noise performance, the higher the current consumption. There is a wide variation in the silicon area used.

### 2.3. Data Reduction Techniques for Multichannel Neural Recording Front-Ends

Front-end neural recording interfaces for multichannel (multielectrode) systems are typically based on the two types of architecture in Figure 11 [53]. In the approach in Figure 11(a), the analog front-end circuits which amplify and filter the neural signal acquired from each of the electrodes are grouped in an array of channels. Each of these channels comprises a low noise amplifier (LNA) of the type described in Sections 2.1 and 2.2 and a bandpass filter, followed by a programmable gain amplifier (PGA) to maximize the output swing. The analog outputs from all the channels are then multiplexed in time and converted into digital words by an analog-to-digital converter (ADC). The generated time-multiplexed data frames can be either digitally processed to compress them or sent directly to the output as raw data. In the approach in Figure 11(b), instead of sharing the ADC between the analog outputs of the channels, an ADC is embedded in each channel. Then a common digital processor manages the digitized signals from the channels, classifies (or reduces) the data, and sends it to the output. This solution requires a higher silicon area than the approach in Figure 11(a), but it has some merits in terms of power consumption because of the much lower sampling rate requirement at the digitization stage. In addition, the system in Figure 11(b) has the advantage that is easily scalable by replicating the channels. A very compact circuit implementation for the topology in Figure 11(b) is described in [35], requiring a silicon area of only $0.054 \, \text{mm}^2$ per channel in a 130-nm CMOS process technology.

Bandwidth limitation is a key issue for wireless biomedical devices. Wireless transmission of raw data is a major challenge for high channel count recording front-ends. For
### Table 1: Comparison of integrated neural amplifiers.

<table>
<thead>
<tr>
<th>Parameter</th>
<th>[34]</th>
<th>[109]</th>
<th>[45]</th>
<th>[110]</th>
<th>[111]</th>
<th>[112]</th>
<th>[29]</th>
<th>[42]</th>
<th>[35]</th>
<th>[37]</th>
<th>[44]</th>
<th>[40]</th>
<th>[39]</th>
<th>[46]</th>
<th>[36]</th>
</tr>
</thead>
<tbody>
<tr>
<td>Fully differential</td>
<td>No</td>
<td>No</td>
<td>No</td>
<td>Yes</td>
<td>Yes</td>
<td>Yes</td>
<td>No</td>
<td>Yes</td>
<td>Yes</td>
<td>Yes</td>
<td>No</td>
<td>No</td>
<td>Yes</td>
<td>No</td>
<td>No</td>
</tr>
<tr>
<td>CMOS process ($\mu$m)</td>
<td>0.5</td>
<td>1.5</td>
<td>0.18</td>
<td>0.35</td>
<td>0.13</td>
<td>0.35</td>
<td>0.18</td>
<td>0.18</td>
<td>0.13</td>
<td>0.35</td>
<td>0.18</td>
<td>0.18</td>
<td>0.18</td>
<td>0.18</td>
<td>0.18</td>
</tr>
<tr>
<td>Area (mm²)</td>
<td>0.16</td>
<td>0.107</td>
<td>0.05</td>
<td>N/A</td>
<td>N/A</td>
<td>0.02</td>
<td>N/A</td>
<td>0.063</td>
<td>0.054</td>
<td>0.072</td>
<td>0.065</td>
<td>0.065</td>
<td>N/A</td>
<td>0.76</td>
<td>0.36</td>
</tr>
<tr>
<td>Supply voltage (V)</td>
<td>5</td>
<td>3</td>
<td>1</td>
<td>1</td>
<td>3</td>
<td>1.6</td>
<td>1.8</td>
<td>1.2</td>
<td>1</td>
<td>3</td>
<td>1.8</td>
<td>1/1.8</td>
<td>1</td>
<td>1.2</td>
<td></td>
</tr>
<tr>
<td>Supply current ($\mu$A)</td>
<td>16</td>
<td>38.27</td>
<td>4.67</td>
<td>1.26</td>
<td>1.25</td>
<td>2.8</td>
<td>43.12</td>
<td>4.4</td>
<td>1.6</td>
<td>12.1</td>
<td>2</td>
<td>6.1</td>
<td>1</td>
<td>1.2</td>
<td></td>
</tr>
<tr>
<td>Gain (dB)</td>
<td>39.5</td>
<td>39.3</td>
<td>49.52</td>
<td>45.7</td>
<td>38.3</td>
<td>33</td>
<td>19.1</td>
<td>37.5</td>
<td>47.5</td>
<td>40</td>
<td>37.5</td>
<td>48/60</td>
<td>54.8</td>
<td>60.9</td>
<td>44.5</td>
</tr>
<tr>
<td>Bandwidth (Hz)</td>
<td>25 m–72 k</td>
<td>0–9.1 k</td>
<td>98.4–9.1 k</td>
<td>0.23–78 k</td>
<td>23 m–11.5 k</td>
<td>10–5 k</td>
<td>100–8 k</td>
<td>10–72 k</td>
<td>167–6.9 k</td>
<td>50 m–10.5 k</td>
<td>1–10 k</td>
<td>0.3–9 k</td>
<td>0.38–5.1 k</td>
<td>1–10 k</td>
<td>80–15 k</td>
</tr>
<tr>
<td>Input-referred noise (µV mss)</td>
<td>2.2</td>
<td>7.8</td>
<td>5.6</td>
<td>4.43</td>
<td>1.95</td>
<td>6.08</td>
<td>2.36</td>
<td>3.5</td>
<td>3.8</td>
<td>2.2</td>
<td>10.6</td>
<td>5</td>
<td>4</td>
<td>4.4*</td>
<td>8.1</td>
</tr>
<tr>
<td>Noise bandwidth (Hz)</td>
<td>N/A</td>
<td>0.1–10 k</td>
<td>1–9.1 k</td>
<td>1–12 k</td>
<td>0.1–25.6 k</td>
<td>10–5 k</td>
<td>1–8 k</td>
<td>10–100 k</td>
<td>1–100 k</td>
<td>0.1–105 k</td>
<td>1–10 k</td>
<td>1–8 k</td>
<td>0.3–10 k</td>
<td>N/A</td>
<td></td>
</tr>
<tr>
<td>CMRR (dB)</td>
<td>≥83</td>
<td>N/A</td>
<td>52.68</td>
<td>58</td>
<td>1.3</td>
<td>60</td>
<td>79</td>
<td>70.1</td>
<td>83</td>
<td>80</td>
<td>74</td>
<td>48</td>
<td>48</td>
<td>&gt;60</td>
<td>N/A</td>
</tr>
<tr>
<td>PSRR (dB)</td>
<td>≥85</td>
<td>N/A</td>
<td>51.93</td>
<td>40</td>
<td>&gt;63</td>
<td>N/A</td>
<td>62</td>
<td>63.8</td>
<td>70</td>
<td>≥80</td>
<td>55</td>
<td>55</td>
<td>≥70</td>
<td>N/A</td>
<td></td>
</tr>
<tr>
<td>THD</td>
<td>1%</td>
<td>1.1%</td>
<td>1%</td>
<td>0.53%</td>
<td>1%</td>
<td>1%</td>
<td>1%</td>
<td>1%</td>
<td>1%</td>
<td>1.2</td>
<td>1%</td>
<td>1.2</td>
<td>1%</td>
<td>1.03</td>
<td>0.05%</td>
</tr>
<tr>
<td>Input range (mVpp)</td>
<td>16.7</td>
<td>5</td>
<td>2.4</td>
<td>5.2</td>
<td>1</td>
<td>N/A</td>
<td>N/A</td>
<td>5.7</td>
<td>3.1</td>
<td>1</td>
<td>2.4</td>
<td>1</td>
<td>1.63</td>
<td>**</td>
<td></td>
</tr>
<tr>
<td>NEF</td>
<td>4</td>
<td>19.4</td>
<td>4.9</td>
<td>2.16</td>
<td>2.48</td>
<td>5.55</td>
<td>6.68</td>
<td>3.35</td>
<td>2.16</td>
<td>2.9</td>
<td>5.78</td>
<td>4.6</td>
<td>1.9</td>
<td>5.45</td>
<td>1.52</td>
</tr>
<tr>
<td>NEF² × VDD</td>
<td>80</td>
<td>1129</td>
<td>43.22</td>
<td>4.53</td>
<td>6.15</td>
<td>92.41</td>
<td>71.39</td>
<td>20.20</td>
<td>5.59</td>
<td>8.41</td>
<td>100.22</td>
<td>38.08</td>
<td>3.61</td>
<td>29.7</td>
<td>2.77</td>
</tr>
</tbody>
</table>

* Referred to the input node of the electrode.

** The amplitude range of the input current is 20 nA_pp.
example, a neural interface with 100 channels, a 30 kHz sampling frequency per channel, and an 8-bit sample resolution would generate raw data at 24 Mbps. For a 1024-channel system (for cortical neural sensing), the data rate would increase to a massive 250 Mbps. This data rate is beyond the transmission capabilities of existing (wideband) implantable wireless transmitters [54–56]. In the case of extracellularly recorded action potentials (spikes), spike sorting [57] is an efficient process to achieve on-chip data reduction, thereby enabling lower data rate wireless transmission and low power consumption. Neural recording microsystems with data reduction capability typically employ some sort of thresholding to detect and extract the neural information [58, 59]. Within this scheme, detection occurs when the spike's amplitude crosses a specified threshold. The threshold detector can be implemented with analog or digital circuits. For applications where a more detailed classification of multunit activity into single-unit activity is required, techniques that extract specific biosignal features (e.g., peak-to-peak amplitude, duration, peak-to-zero-crossing time, etc.) are employed. Most spike sorting methods rely on the assumption that each neuron produces a different, distinct shape (as seen by the electrode) that remains constant throughout a recording window. The first step in such techniques is feature extraction, in which spikes are transformed into a certain set of features that emphasizes the differences between spikes from different neurons as well as the differences between spikes and noise. Then dimensionality reduction takes place, in which feature coefficients that best separate spikes are identified and stored for subsequent processing, while the rest are discarded. Finally, using clustering spikes are classified into different groups, corresponding to different neurons, based on the extracted feature coefficients. Implantable spike sorting hardware must be low power and low area. The algorithms implemented in the hardware must be accurate, automatic, real-time, and computationally efficient. A detailed review and comparison of spike sorting algorithms are provided in [57, 60]. An ultra-low power spike sorting digital chip that can perform detection, alignment and feature extraction simultaneously for 64 channels is described in [61]. The chip was implemented in a 90 nm CMOS process and has a power density of 30 μW/mm², which is significantly lower than the power density (800 μW/mm²) known to damage brain cells [62].

Spike sorting can potentially reduce the data rate by several orders of magnitude compared to transmitting raw data. However, the data in the segments without spikes (which contains useful information on the neuronal activities) is lost. To preserve these activities, one solution is to use the discrete wavelet transform to process the data before transmission [63]. This allows the retention of almost all of the data but at the cost of chip area and power consumption. An alternative is the emerging field of compressive sensing [64]. Compressive sensing enables signal reconstruction from a small number of nonadaptively acquired sample measurements corresponding to the information content of the signal rather than to its bandwidth. It has simple compression steps and takes advantage of the signal's sparseness, allowing the signal to be determined from relatively few measurements. Energy-efficient compressive sensing methods for implantable neural recording is a topic of current research [65, 66].

Table 2 compares various multichannel neural recording systems with wireless transmission capability. The design in [55] has the highest number of channels (128) and data rate (90 Mbps). The design in [67] has the lowest power consumption per channel.

3. Neural Stimulators

Electrical stimulus applied to nerves can trigger action potentials. At least two electrodes are required in order to produce current flow. The electrodes are commonly arranged in monopolar or bipolar configuration. In both cases, the active (working) electrode is placed near the nerve to be stimulated. In monopolar stimulation, the indifferent electrode is placed away from the active electrode, whilst in bipolar stimulation the reference electrode is placed near the active electrode.

There are two main modes of stimulation, namely, current-mode and voltage-mode, as shown in Figure 12, although charge-mode stimulation also exists [68]. Current-mode stimulation (Figure 12(a)) is extensively used in implantable stimulators. Active current sources (and sinks) are used to supply the stimulus current to the load (tissue-electrode impedance). For current sources with high output impedance, the stimulus current amplitude is not affected by changes in the load. Examples of integrated current-mode stimulators in CMOS technology for various applications such as nerve root stimulation for spinal cord injury, vestibular prosthesis for balance disorders, and deep brain stimulation for severe movement disorders are described in [11, 69–73]. The current amplitude range is from 1 mA to 16 mA.

In voltage-mode stimulation (Figure 12(b)), the stimulator output is a voltage, and therefore the magnitude of the current delivered to the tissue is dependent on the inter-electrode impedance. Thus, it is difficult to control the exact amount of charge supplied to the load because of impedance variations. In the system described in [74], the stimulator drives the electrodes with a sequence of voltage steps, charging the electrode metal-fluid capacitance. This applies a voltage waveform that is an approximation of the waveform that would appear at the electrode if a current pulse was applied (see Figure 12(a)). Since the charge is delivered to the electrode directly from the capacitors, it avoids the (substantial) power in the current sources of its current-mode counterpart, as well as providing large voltage compliance. However, this method requires large capacitors (which act as voltage sources) that are difficult to implement on-chip. The design in [74] has five 1 μF capacitors for a 15-electrode stimulation system, which will increase with more electrodes. It is also difficult to achieve fine resolution compared to current-mode stimulation since voltage-mode is an approximate method of producing a current pulse and increasing the resolution requires more capacitors. Another voltage-mode stimulator is described in [75]. Its architecture features energy recovery enabling power savings of 53%.
<table>
<thead>
<tr>
<th>Parameter</th>
<th>[113]</th>
<th>[59]</th>
<th>[55]</th>
<th>[114]</th>
<th>[115]</th>
<th>[56]</th>
<th>[67]</th>
<th>[54]</th>
<th>[116]</th>
<th>[15]</th>
</tr>
</thead>
<tbody>
<tr>
<td>Number of channels</td>
<td>100</td>
<td>64</td>
<td>128</td>
<td>64</td>
<td>32</td>
<td>96</td>
<td>64</td>
<td>100</td>
<td>64</td>
<td>64</td>
</tr>
<tr>
<td>CMOS process (μm)</td>
<td>0.5</td>
<td>0.5</td>
<td>0.35</td>
<td>0.35</td>
<td>0.5</td>
<td>0.13</td>
<td>0.13</td>
<td>0.5</td>
<td>0.13</td>
<td>0.13</td>
</tr>
<tr>
<td>Area (mm²)</td>
<td>27.73</td>
<td>N/A</td>
<td>63.36</td>
<td>8.37</td>
<td>16.27</td>
<td>25</td>
<td>18.4</td>
<td>25.48</td>
<td>12</td>
<td>12</td>
</tr>
<tr>
<td>Supply voltage (V)</td>
<td>3.3</td>
<td>1.8</td>
<td>3.3</td>
<td>3</td>
<td>1.2</td>
<td>1.2</td>
<td>1.2</td>
<td>1</td>
<td>1.2/3</td>
<td>1.4</td>
</tr>
<tr>
<td>Total power (mW)</td>
<td>13.5</td>
<td>14.4</td>
<td>6</td>
<td>172</td>
<td>5.85</td>
<td>6.5</td>
<td>0.38</td>
<td>6/90.6</td>
<td>5.03</td>
<td>1.4</td>
</tr>
<tr>
<td>Power per channel (μW)</td>
<td>135</td>
<td>225</td>
<td>46.87</td>
<td>269</td>
<td>182.8</td>
<td>67.7</td>
<td>5.93</td>
<td>906</td>
<td>78.5</td>
<td>21.8</td>
</tr>
<tr>
<td>Amplifier gain (dB)</td>
<td>40</td>
<td>60</td>
<td>40</td>
<td>65–83</td>
<td>67.8–78</td>
<td>54</td>
<td>47.5–65.5</td>
<td>46</td>
<td>54–60</td>
<td>54–60</td>
</tr>
<tr>
<td>Low frequency corner (Hz)</td>
<td>300</td>
<td>10</td>
<td>0.1</td>
<td>1</td>
<td>0.1</td>
<td>280</td>
<td>200</td>
<td>1</td>
<td>30</td>
<td>1</td>
</tr>
<tr>
<td>High frequency corner (kHz)</td>
<td>5</td>
<td>9.1</td>
<td>20</td>
<td>10</td>
<td>8</td>
<td>10</td>
<td>6.9</td>
<td>78</td>
<td>5</td>
<td>5.1</td>
</tr>
<tr>
<td>Input-referred noise (μVrms)</td>
<td>5.1</td>
<td>8</td>
<td>4.9</td>
<td>3.05</td>
<td>4.62</td>
<td>2.2</td>
<td>3.8</td>
<td>2.83</td>
<td>6.5</td>
<td>5.1</td>
</tr>
<tr>
<td>ADC resolution (bits)</td>
<td>10</td>
<td>8</td>
<td>9</td>
<td>72</td>
<td>8.1</td>
<td>10</td>
<td>7.65</td>
<td>12</td>
<td>78</td>
<td>76</td>
</tr>
<tr>
<td>Sampling frequency (kHz)</td>
<td>15</td>
<td>62.5</td>
<td>640</td>
<td>20</td>
<td>N/A</td>
<td>31.25</td>
<td>27/90</td>
<td>20</td>
<td>57</td>
<td>≤100</td>
</tr>
<tr>
<td>Downlink modulation</td>
<td>ASK</td>
<td>FSK</td>
<td>N/A</td>
<td>N/A</td>
<td>N/A</td>
<td>OOK</td>
<td>40.68</td>
<td>N/A</td>
<td>OOK</td>
<td>915</td>
</tr>
<tr>
<td>Carrier frequency (MHz)</td>
<td>2.64</td>
<td>4/8</td>
<td>N/A</td>
<td>N/A</td>
<td>N/A</td>
<td>N/A</td>
<td>OOK</td>
<td>40.68</td>
<td>N/A</td>
<td>OOK</td>
</tr>
<tr>
<td>Uplink modulation</td>
<td>FSK</td>
<td>OOK</td>
<td>UWB</td>
<td>FSK</td>
<td>FSK</td>
<td>UWB</td>
<td>LSK</td>
<td>FSK</td>
<td>FSK</td>
<td>UWB</td>
</tr>
<tr>
<td>Carrier frequency (GHz)</td>
<td>0.333</td>
<td>0.070–0.2</td>
<td>4</td>
<td>0.4</td>
<td>0.915</td>
<td>N/A</td>
<td>0.004</td>
<td>3.2/3.8</td>
<td>0.915</td>
<td>0–1/3.1–10.6</td>
</tr>
<tr>
<td>Data rate (Mbps)</td>
<td>0.33</td>
<td>2</td>
<td>90</td>
<td>1.25</td>
<td>0.709</td>
<td>30</td>
<td>1.92</td>
<td>48</td>
<td>1.5</td>
<td>10</td>
</tr>
<tr>
<td>On-chip signal processing</td>
<td>Yes</td>
<td>Yes</td>
<td>Yes</td>
<td>Yes</td>
<td>Yes</td>
<td>No</td>
<td>Yes</td>
<td>No</td>
<td>Yes</td>
<td>Yes</td>
</tr>
<tr>
<td>Power source</td>
<td>Inductive</td>
<td>Inductive</td>
<td>Battery</td>
<td>Battery</td>
<td>Inductive</td>
<td>Battery</td>
<td>Inductive</td>
<td>Battery</td>
<td>Battery</td>
<td>Battery</td>
</tr>
<tr>
<td>Stimulation capability</td>
<td>No</td>
<td>No</td>
<td>No</td>
<td>No</td>
<td>No</td>
<td>No</td>
<td>No</td>
<td>No</td>
<td>No</td>
<td>Yes</td>
</tr>
</tbody>
</table>
Common stimulation waveforms are either monophasic or biphasic (Figure 13). A monophasic stimulus consists of a repeating unidirectional cathodic pulse (this type of stimulus is common in surface electrode stimulation). A biphasic waveform consists of a repeating current pulse that has a cathodic (negative) phase followed by an anodic (positive) phase. The cathodic phase depolarizes nearby axons and triggers the action potential. The succeeding anodic phase reverses the potentially damaging electrochemical processes that can occur at the electrode-tissue interface during the cathodic phase by (ideally) neutralizing the charge accumulated in the cathodic phase, allowing stimulation without tissue damage. The application of charge-balanced waveforms is very important, especially for implanted electrodes. Usually the stimulus for the cathodic phase is rectangular, supplied by active circuits, while the stimulus for the anodic phase could be either square or exponentially decaying. The rectangular secondary phase is also known as active discharging and the exponentially decaying phase as passive discharging.

3.1. Charge-Balanced Stimulation. Charge imbalance can be caused by many reasons including semiconductor failure, leakage currents due to crosstalk between adjacent stimulating channels (sites), and cable failure. A blocking capacitor in series with each electrode is used for electrical safety against single-fault conditions [76]. The blocking capacitor is also used to achieve (passive) charge balance. Figure 14 shows three current-mode stimulator configurations, each employing a blocking capacitor [69]: (a) dual supplies with both active phases, (b) single supply with both active phases, and (c) single supply with active cathodic phase and passive anodic phase. The programmable current sink $I_{\text{stimC}}$ and current source $I_{\text{stimA}}$ generate the cathodic and anodic currents, respectively. These currents are driven through the load by the control of switches $S_1$ and $S_2$. When only a single supply is available (Figure 14(b)), the anodic and cathodic currents are generated from a single current sink ($I_{\text{stim}}$) by reversing the current paths by switch $S_3$. Both configurations in Figures 14(a) and 14(b) are (ideally) designed to be charge-balanced to avoid charge accumulation. However, achieving exactly zero net charge after each stimulation cycle is not possible due to mismatch or timing errors and leakage from adjacent stimulating sites. Therefore, it is important to include switch $S_3$ to periodically remove the residual charge by providing an extra passive discharge phase in which the voltage on the blocking capacitor drives current through the electrodes to fully discharge them. Given the necessity for the third phase in the circuits in Figures 14(a) and 14(b), it is possible to use the passive discharge phase as the main anodic phase as shown in the circuit in Figure 14(c) and the corresponding waveform in Figure 13(c). Note that the use of capacitive
electrodes may also drive the passive discharge phase. However, blocking capacitors may still be deemed necessary to ensure that dc current cannot flow into the electrodes in the event of semiconductor failure due to breakdown voltage or leakage current. Due to the large value required for the blocking capacitors (which can be a few microfarads in the case of stimulators for lower-body applications), they are typically realized as off-chip surface-mount components. For multichannel stimulation implants, a single blocking capacitor per channel may not be sufficient to guarantee safety due to a single-fault failure [77].

For applications where the use of blocking capacitors is not possible due to physical size limitations (e.g., retinal implants with several hundreds of stimulating electrodes), other methods for (active) charge balancing exist (Figure 15).

1. Dynamic current balancing [78]. In the circuit in Figure 15(a), a current sink is used to generate the cathodic phase and a pMOS transistor ($M_1$) to generate the anodic phase. Before generating a biphasic current pulse, $S_{\text{cathodic}}$ and $S_{\text{anodic}}$ are open, and the two sampling switches, $S_{\text{samp}}$, are closed. When the circuit settles, the amplitude of the drain current of $M_1$ is the same as the current sink (due to the feedback), and the resulting bias voltage ($V_{\text{bias}}$) on the gate of $M_1$ is sampled and held. Then the two $S_{\text{samp}}$ switches open and $S_{\text{cathodic}}$ closes to form the cathodic current. Following this, $S_{\text{cathodic}}$ opens and $S_{\text{anodic}}$ closes. Because the gate voltage of $M_1$ is held at $V_{\text{bias}}$, an anodic current equal to the amplitude of the (cathodic) sink current passes through $S_{\text{anodic}}$ to the load. By optimizing the S&H circuit to reduce errors due to charge effects, a residual dc current error of 6 nA is claimed [78].

2. Active charge balancer [79]. In the circuit in Figure 15(b), the residual voltage after a biphasic pulse is measured and compared to a safe reference voltage. If the electrode voltage exceeds the safe window, additional short stimulation current pulses are applied to steer the electrode voltage towards a balanced condition. The charge balancer is typically activated for less than 5% of the operation time. This method has the advantage of providing feedback information on the electrode condition after stimulation.

3. H-bridge with multiple current sinks [80]. This approach assumes an asymmetric biphasic waveform. By way of example, consider the H-bridge circuit in Figure 15(c). During the high-amplitude cathodic phase, identical current sinks $I_1, I_2, \ldots, I_N$ act in parallel to pass current through the electrodes for a time $T$. Then during the low-amplitude anodic phase, one of the $N$ current sinks is used to pass current through the electrodes (in the reverse direction) for a time $N \cdot T$. On a single stimulation cycle this would give inaccurate charge balance as in practice the current sinks would not be perfectly matched. However, if the current sink that is active in the anodic phase is sequentially changed after each stimulus waveform, then after $N$ cycles each sink would have been active for the same amount of time during the cathodic and anodic phases, yielding accurate charge balance. The method is claimed to achieve a maximum charge mismatch of 0.45% [80].

4. Multiphase compensation [71]. The multiphase compensation technique is illustrated by the waveform in Figure 15(d). To generate an asymmetric biphasic pulse, the width of the anodic phase is extended to $N$ times the cathodic width, $T$. Ideally, the anodic current amplitude should be $N$ times smaller than the cathodic amplitude. The amplitude of the currents is controlled through an ADC. Due to the finite resolution of the ADC, the amplitude of the cathodic...
current, $A_c$, may not be an integer multiple of $N$, where $A_c = M + \alpha$ and $M$ is the integer multiple of $N$ that is the closest to $A_c$. Thus, there will be charge balance error between the two phases equal to $(\alpha/N) \cdot N \cdot T = \alpha \cdot T$. In the multiphase compensation technique, an additional shorter anodic pulse of width $T$ and amplitude $\alpha$ is initiated after the anodic phase to compensate for the error. Subsequently, a passive discharge phase can be applied to further reduce any remaining charge imbalance. The method is claimed to achieve a residual dc current error of 4.5 nA [71].

### 3.2. Other Stimulator Circuits

A stimulator circuit that is failsafe with no off-chip blocking capacitors is shown in Figure 16 [72]. The circuit generates an active stimulation phase by high frequency current switching (HFCS), followed by a passive discharge phase. During the active stimulation phase, current $I_{stim}$ (generated by a current generator circuit) is switched alternately through the left and right branches of the charge transfer block (Figure 16(b)). This high frequency switching mechanism allows the size of the blocking capacitors $C_1$ and $C_2$ to be significantly reduced. The circuit operation is as follows. During the low-state of the control signal $\Phi_{stim,left}$, switch $M_1$ is closed and $M_3$ is open. In this phase, diode $D_1$ is reverse biased, diode $D_2$ is forward biased, and current $I_{S1}$ flows and charges up $C_1$. In the same phase, on the right branch of the charge transfer block, the control signal $\Phi_{stim,right}$ is high, and hence switch $M_2$ is open and $D_3$, $C_2$, and $M_4$ form a closed path which discharges $C_2$ to one-diode-drop voltage. During the high-state of $\Phi_{stim,left}$, $\Phi_{stim,right}$ is turned low which causes $C_2$ to be charged up and $C_1$ to be discharged. The complementary high frequency currents $I_{S1}$ and $I_{S2}$ generated during the stimulation phase are summed at the anode of the electrode-tissue load. After the stimulation phase, the load is passively discharged via an ac-coupled discharge switch (Figure 16(c)) using depletion-mode transistors $M_5$–$M_7$ (connected in parallel for redundancy) which conduct most of the time. The operation of this circuit is as follows. At each negative edge of the control pulse $\Phi_{discharge}$, negative charge is injected into capacitors $C_3$ and $C_4$. On the following positive edge diode $D_5$ is reverse biased so the charge on
C₄ is retained (apart from the leakage via resistor R₁), and positive charge is injected into C₃ balancing out the injected negative charge on C₄. When a second negative charge arrives it adds to the charge in C₄ and the cycle is repeated. After a couple of cycles enough negative charge is built up on C₄ to switch off M₅–M₇. While the pulses continue, the gate-source voltage of M₅–M₇ remains negative, so they are held off. When the pulses cease the negative charge decays via R₁. An ac-coupled switch is also used for the cathode as shown in Figure 16(d). Its operation is exactly the same as described above, except now it is the positive edge that provides the charge to C₆. Implementation of the HFCS technique requires silicon-on-insulator (SOI) technology which features fully-isolated active and passive devices. The design in [72] used the X-FAB XT06 process technology (http://www.xfab.com/) which has trench isolation. HFCS is suitable for stimulus current amplitudes up to about 1 mA.

The technique can be employed to greatly reduce the size of high-reliability, multichannel stimulator implants sited close to the target nervous tissue (e.g., in the spinal canal or on the brain surface).

There is a demand for high efficiency stimulators in applications which have limited power availability. The basic current-mode stimulator is inherently inefficient and attempts have been made to increase stimulator efficiency. A design is described in [81] which achieves a 2x to 3x reduction in energy consumption compared with the basic current-mode stimulator. It is based on a dc-dc buck voltage converter efficiently providing a variable output for biphasic pulses. The voltage output drive is adjusted by feedback from a sensor detecting the current in the electrode so providing a controlled current independent of the value of the load impedance. The conventional series resistor employed for current sensing wastes energy and is not used. At the
output of the dc-dc converter, there is a smoothing capacitor which is in parallel with the electrode load. This capacitor is temporarily disconnected from the converter and the rate of decay of voltage across the capacitor due to the current in the electrode load is detected and used in feedback. This sensing method avoids wasting energy. The system has appreciable sawtooth noise superimposed on the current pulse. Biphasic 400 μA current pulses with widths of 1 ms and rise times of about 100 μs are shown. Components external to the integrated circuit control unit include an inductor (39 μH) and capacitors (up to 10 μF).

Another example of minimizing power dissipation while retaining the basic current-mode generator is presented in [82]. It assumes the use of a high frequency (1 MHz) inductive power link. On its secondary coil, zero voltage switching and adjustment of the conduction angle provide a variable voltage supply to the electrode load and current generator. Feedback adjusts the variable voltage supply so that the current generator operates at just above its compliance limit, minimizing the power it uses. The biphasic current is generated using a single sink current generator with switches similar to Figure 15(c). The supply voltage update is near real-time so the quality of the current pulses is high, irrespective of how the load changes during stimulation. Depending on load conditions, 20% to 75% power saving compared to a conventional current-mode stimulator is claimed. In a prototype design, stimulation currents of 20 μA to 1 mA with pulse widths of 20 μs to 200 μs are quoted.

Where cross-coupling between closely spaced, simultaneously operated, stimulating sites must be avoided, floating power supplies are needed. A successful example design, capable of supporting parallel stimulation to electrodes on three semicircular canals for vestibular prosthesis, is described in [83].

### 4. Delivery of Power and Data to Implants

The requirements imposed on medical devices operating in the body are application specific, but there are a common set of constraints in size, power, and functionality. The interplay between these constraints determines the available processing bandwidth for the electronics, the operating time (in the case of battery-operated devices such as pacemakers) and the communication range and bandwidth of the wireless telemetry link. In addition, the location of the device in the body, data rate, frequency, and regulatory standards influence the design complexity and power dissipation of telemetry links.

Long range telemetry links (typically greater than 2 meters) are mainly battery-operated and must conform to strict regulatory standards. They require both high sensitivity receivers and high output power transmitters, both of which result in high power dissipation. In addition, unlike most near-field (short range) inductive links (discussed below) where a stable reference clock can be extracted from the external carrier frequency, long range telemetry links require stable crystal references and frequency synthesizers to generate a local carrier with good frequency stability. These transceivers are typically operated in dedicated frequency bands such as the U.S. federal communications commission (FCC) approved 402–405 MHz for medical implant communication service (MICS) band (e.g., the ZL70102 Microsemi transceiver). This band has a 300 kHz maximum bandwidth and a maximum output power of ~16 dBm [84] and the data rate is limited to about 200 kbps. The industrial, scientific, and medical (ISM) radio bands are also frequently used for medical telemetry transmitters. These include the 902–928 MHz, 2.4–2.4835 GHz, and 5.725–5.875 GHz frequency bands and have transmission ranges up to 10 meters. Lower frequencies require larger antennas, while higher frequencies have higher losses due to tissue absorption. The optimum frequency band for wireless transmitters located in the body is reported to be approximately 900 MHz [85]. Ultra wideband (UWB) is an alternative wireless data transmission method used at very low energy levels for short range, high bandwidth communications. Recently, UWB communication in the 3.1–10.6 GHz band was used to develop low power wireless transmitters in CMOS technology for implantable medical devices [15, 55, 56] for high date rate (10–90 Mbps) transmission from the implant to the external device. Modulation schemes such as on/off keying (OOK) and pulse position modulation (PPM) are used to generate the short pulses.

For short range links (typically up to a few centimetres), low frequency inductive links are used for both power supply and bidirectional data transmission. Examples include cochlear and vestibular implants [3, 86]. These near-field systems can be made small and highly integrated. Modulation schemes such as OOK, amplitude shift keying (ASK), and frequency shift keying (FSK) are often used for data transmission from the external unit to the implanted device [87, 88]. To minimize radio frequency heating due to tissue absorption, these systems are typically operated below 15 MHz with an achievable data rate up to a few Mbps. Transferred power to the implant typically ranges from 10 mW to 125 mW [89–91]. The power transmitted through the tissue should comply with safety standards.

A basic block diagram of a typical architecture for transmitting power/data to an implant via an inductive link is shown in Figure 17. The external transmitter typically consists of a class D or class E power amplifier capable of providing large currents in the tuned primary coil (L1) from a relatively low voltage. In the implant, the induced voltage that appears across the secondary coil (L2, tuned by a capacitor) is rectified and regulated to provide a power supply for the electronics. The data link from the external transmitter to the implant (the downlink) is often achieved by modulating the envelope of the power carrier to create detectable changes across the secondary coil. The data link from the implant to the external circuit (the uplink) is commonly implemented by load modulation techniques. These techniques utilize the property of the coupled coils in which a change in the load of the secondary circuit is reflected back as changing impedance in the primary, through their mutual inductance M. Examples of modulation techniques for uplink and downlink data transmission are discussed later.

The basic equivalent circuit of the inductive link is shown in Figure 18 [92, 93]. The primary circuit (R1, L1, C1) is
tuned in series in order to provide a low impedance load to the driving transmitter. Resistor $R_1$ includes the loss of the inductor $L_1$ and the output resistance of the voltage source $V_s$ (representing the transmitter driver). At resonance, the voltages related with $C_1$ and $L_1$ cancel each other and thus the primary circuit requires small voltage swings at its inputs. Thus, the primary circuit loads the secondary circuit with a small load. The topology of the secondary circuit ($L_2, C_2, R_2$) is tuned in parallel in order to amplify sufficiently the induced voltage to drive a nonlinear (rectifier) load. Resistor $R_2$ includes the loss of the inductor $L_2$ and the load resistance of the implant circuits. Both RLC circuits are tuned to the same resonant frequency. The gain factor of the link at resonance is [93]

$$\frac{V_2}{V_s} = \frac{\sqrt{R_2/R_1}}{(k/k_{\text{crit}} + (k/k_{\text{crit}}))}; \quad k_{\text{crit}} = \frac{C_1 R_1}{C_2 R_2}, \quad (3)$$

where $k$ is the coupling coefficient and $k_{\text{crit}}$ is the critical coupling coefficient. The relative dimensions of the coils and the air gap cause a low coupling coefficient ($k < 0.1$). In addition, the inductances of the primary $L_1$ and secondary $L_2$ coils are small and can be generated using coils with a few turns. The typical variations of the gain factor are illustrated in Figure 19. The gain factor is not constant with respect to coupling variations. The link transfers maximum power when the resistance in the primary is equivalent to the reflected secondary resistance, assuming that reactive components are being cancelled. It can be shown that efficiency ($\eta$) at critical coupling is equal to 50% (Figure 20) [93]. For the purposes of data transfer, the link behaves as a narrow-band bandpass filter. At resonance and assuming that the coils have the same quality factor $Q$, the bandwidth of the link is $B = \sqrt{2}(f_c/Q)$ where $f_c$ is the carrier frequency [87].

Power transfer (gain factor) optimization and data transfer (bit rate) optimization have contradicting requirements. To illustrate this, the circuit in Figure 18 was simulated in Advanced Design System (Agilent EEsof EDA) with ($R_1 = 1\Omega$, $L_1 = 1\mu\text{H}$) and ($R_2 = 1k\Omega$, $L_2 = 1\mu\text{H}$) tuned to 1MHz by $C_1$ and $C_2$, respectively [94]. The relationships between gain factor (power transfer), carrier frequency, and coupling coefficient are plotted in Figure 21. The optimum gain factor is at the resonant frequency (1MHz) and coupling coefficient ($k$) of around 0.05. The gain factor can be improved by increasing the quality factors. The bandwidth of the inductive link can be increased by lowering the quality factors.
It also increases as the coupling coefficient increases (i.e., the gap between the coils decreases). In practice, the inductive link should be designed to account for the potentially wide variation in the gap between the coils. Coupling compensation techniques (incorporating feedback) regulate the voltage across the secondary coil [95–97]. Power and data links can use separate sets of coils because this allows independent optimization to maximize performance [98]. However, using separate coils has the drawbacks of an increase in the implant footprint and electromagnetic interference. The latter requires the use of complex modulation techniques to minimize its effects and increases system complexity.

The most commonly used technique for uplink data transmission is passive signalling [99] also known as load shift keying (LSK) [100]. This modulation is based on the reflection of the implant’s load to the transmitter via the inductive link. A typical implementation is shown in Figure 22. The binary Data stream shorts the implant coil and the change in impedance is reflected in the transmitter because the implant load is much larger than the on-resistance of the switch transistor. Where only two coils are used (for both power and data) there is a risk of disruption in power delivery if the short is applied for too long to the implant coil. When communication is in idle mode, the link should be optimized for maximum power transfer. The bandwidth of LSK is limited by the coupling factor, the parameters of the coils, and the transient response of the inductive link. Multilevel LSK may be used to increase the data rate [101]. An alternative communication technique for uplink is passive phase shift keying (PPSK) [102]. Unlike LSK, the switch across the secondary coil (Figure 22) closes synchronously with the carrier for half the carrier cycle. The transient response in the primary coil current is detected as a logic “1” signal. An integrated implementation of PPSK in CMOS technology is presented in [103]. The circuit was designed to work at 13.56 MHz with a single set of coils for both data transmission and power delivery. The link can reach a data rate of up to 1/16 of the carrier frequency, that is, 847.5 kbps in this case, making it the fastest data rate achieved by a single wireless (inductively coupled) link used simultaneously for power delivery and communication for implantable devices.

Another method for uplink transmission is pulse harmonic modulation (PHM). It achieves a data rate of 20 Mbps for 1 cm coil separation using a carrier frequency of 66.7 MHz [104]. A pattern of very narrow pulses with specific time delays and amplitude is used, which minimizes intersymbol interference across the receiver coil. PHM is one of the fastest data transmission methods currently known via inductively coupled coils. Unfortunately, to implement it, a separate power link is required as this method is carrier-less.

For downlink data transmission, various digital modulation schemes are used. The most common are binary amplitude shift keying (BASK), binary frequency shift keying (BFSK) and binary phase shift keying (BPSK). The simplest in implementation is BASK but it is sensitive to amplitude fluctuation and the bit rate is typically limited to about 10% of the carrier frequency [87, 88]. In BFSK the binary data is represented by constant amplitude, using two different frequencies, where logic “1” is assigned to one frequency and logic “0” to the other. Compared to BASK, BFSK can provide higher data-rate-to-carrier-frequency ratio [105] but it requires a wide passband in the inductive link to allow...
for the different frequencies, which limits power transfer. The advantage of BPSK over BASK and BFSK is the use of a carrier with fixed amplitude and fixed frequencies [106], enabling efficient and stable power transfer. However, BPSK requires a complicated demodulator, implemented by some kind of phase-locked loop, usually a Costas loop [87]. A comprehensive survey and comparison of modulation techniques is presented in [88].

5. Conclusion and Future Directions

Since the 1950’s remarkable efforts have been undertaken in the development of implantable medical devices. Initially most of the successful applications focused on cardiac rhythm management. Today’s implantable medical devices provide therapy to treat numerous health conditions. Exciting new applications, for example, in electrical neuromodulation, can be used to treat Parkinson’s disease, epilepsy, bladder control, gastrointestinal disorders and numerous psychological disorders such as obsessive-compulsive disorder. Implantable medical devices can now provide a range of pharmacological therapies enabling precise dosage and interval delivery of drugs to more effectively treat patients’ conditions while minimizing side effects.

There are many challenges when creating an implantable medical device. These include microelectronic design, electrode technology, packaging, and biomedical signal processing. This paper has focused on the advances in microelectronics over the last decade or so for implantable medical devices and systems. Several examples of neural amplifiers featuring low noise and low power to monitor the small electrical potentials produced from living neurons via electrodes have been discussed. Both clock-based and continuous-time techniques are used in the design of neural amplifiers. Nowadays, implantable neural recording devices include sophisticated signal processing functions on-chip for data compression, such as spike sorting. Analog and mixed analog/digital circuits are used in their implementation. New advanced techniques to further reduce the power and bandwidth requirements of the wireless data transmission link, for example, based on the concept of compressive sensing, will continue to emerge.

Neural stimulation is a key function performed by implantable medical devices for many applications. Some common principles and design techniques for neural stimulators have been presented, including new techniques which avoid the need for off-chip blocking capacitors and methods for achieving charge balancing. There is a need for energy-efficient stimulator circuits to reduce power consumption (two such examples have been discussed) and further developments in this area are anticipated, particularly for applications requiring many stimulation sites (e.g., retinal prosthesis for the blind).

Wireless power and data operation of implantable medical devices are important because they avoid the need for implantable batteries and offer more flexibility to patients. Although major advances have been achieved in the field of wireless communications and wireless powering for implants, further improvements in terms of new techniques that allow better optimization of the entire system are anticipated. The basic principles of inductively coupled telemetry links commonly used to wirelessly power implants and to provide them with a medium for bidirectional communication have been discussed. Recent developments include the introduction of advanced modulation schemes (e.g., PHM) that allow wideband transmission of data over inductively coupled coils. In addition, transceivers based on conventional wideband wireless radio technology are emerging. These are expected to continue to offer improved performance in terms of an increase in output data rate with lower power consumption requirements, as smaller geometry silicon process technologies are used for the implementation of the implantable circuits. On the subject of providing power to implantable medical devices, it is expected that systems using energy harvested from outside and inside the human body will evolve. An example of such a system that uses the cochlear in the inner ear as a battery source to supply a 2.4 GHz radio transmitter is described in [107]. For such systems ultra-low power circuits are essential.

An important research topic involves closed-loop sense and stimulate systems which will continue to develop, possibly combining both sensing of electrical and chemical responses, for applications such as DBS, epilepsy, and other neurological conditions. Such systems will allow better management of the clinical condition allowing systems to adapt to the varying pathological characteristics. In addition, when new highly miniaturized implantable neural interfaces are developed, a step change in micropackaging techniques will be needed to protect the active area of the integrated circuit from hostile environments experienced in the body. As an example, a recent publication describes a technique for integrated circuit micropackages, dedicated to neural interfaces, based on gold-silicon wafer bonding [108].

The expectation of longer life and a progressively increasing knowledge base will place further dependence on modern healthcare technologies to improve the quality of life of a very large number of patients (both young and old). This will undoubtedly provide a fertile ground for future research.

Conflict of Interests

The author declares that there is no conflict of interests regarding the publication of this paper.

References

1-2, pp. 3–21, 2008.


20 Advances in Electronics


D. C. Galbraith, M. Soma, and R. L. White, "A wide-band efficient inductive transdermal power and data link with coupling
Submit your manuscripts at http://www.hindawi.com