A 16-channel Analog VLSI Processor for Bionic Ears and Speech-Recognition Front Ends

Michael W. Baker, Timothy K.-T. Lu, Christopher D. Salthouse, Ji-Jon Sit, Serhii Zhak, Rahul Sarpeshkar

rahuls@mit.edu

Analog VLSI and Biological Systems Group, Research Laboratory for Electronics
Massachusetts Institute of Technology
77 Mass. Ave.
Cambridge, MA, 02139

Abstract

We describe a 470μW 16-channel analog VLSI processor for bionic ears (cochlear implants) and portable speech-recognition front ends. The power consumption of the processor is kept at low levels through the use of subthreshold CMOS technology. Each channel is composed of a programmable bandpass filter, an envelope detector, and a logarithmic dual-slope analog-to-digital converter that currently operate over 51dB of input dynamic range. The 16 channels were programmed to cover the entire audio frequency spectrum in a logarithmic or mel-scale fashion and sampled at 312.5Hz with 64 discriminable levels per channel. The processor also includes an on-chip low-power microphone front end that transduces sound to an electrical signal that is input to each of the 16 channels. The processor, implemented in a 1.5μm process on a 9.23mm x 9.58mm chip, with a 2.8V supply, offers an order-of-magnitude power saving over more traditional A-to-D-then-DSP processors implemented in advanced submicron processes. It is thus suited for fully-implanted bionic ear processors of the future or portable speech-recognition front ends.

1. Introduction

Cochlear implants or bionic ears (BE) restore hearing in profoundly deaf patients. They function by transforming frequency patterns in sound into corresponding spatial electrode-stimulation patterns for the auditory nerve. Over the past 20 years, improvements in sound-processing strategies, in the number of electrodes and channels, and in the rate of stimulation have yielded improved sentence and word recognition scores in patients (1). Next-generation implants will be fully implanted inside the body of the patient and consequently have very stringent requirements on the power consumption used for signal processing. Our processor is intended for use in such next-generation implants. The digital outputs of the processor and its programmability ensure ease of use with the other parts of an implant system such as its wireless communication link and programming interface.

In addition to bionic ears, low-power processors such as the one described in this paper are useful in portable speech-recognition front ends of the future. Such systems will need programmable front-ends that take a microphone input, and output bits that represent spectral information.

A common speech-processing strategy, used in implants and in speech-recognition systems, employs a mel cepstrum filter bank with 8-20 channels. The mel scale maps frequencies to a perceptually linear scale (2). Filter banks based on the mel scale use linearly spaced filter center frequencies up to 1kHz and logarithmically spaced center frequencies above 1kHz. In the ubiquitous cepstral techniques, a logarithmic measure of the spectral energy in each filter bank channel is used for further processing.

This paper describes a 16-channel, programmable processor for bionic ears and speech-recognition front ends. A microphone pre-amplifier converts sound into an electrical input that is fed to each of the 16 channels. Each channel is comprised of a programmable Gm-C band-pass filter, an envelope detector, a peak detector with asymmetric attack and release, a logarithmic mapping circuit, and a low-power dual-slope A-to-D converter. We have briefly described the operation of a single channel of processing in (3), presenting data from a similar channel with 51dB of dynamic range. In this paper, we will describe each channel and the overall simultaneous operation of all channels of the processor.

The signal-processing blocks of the system are shown in Fig. 1 and described in the associated caption. Section 2 briefly describes the analog front end of the processor. Sections 3A-3D describe the signal processing of a single channel, namely, the band-pass filtering, the filter programming, the envelope detection, and the logarithmic

0-7803-7842-3/03/$17.00 © 2003 IEEE IEEE 2003 CUSTOM INTEGRATED CIRCUITS CONFERENCE

A-to-D conversion. Section 4 discusses measured experimental results. Section 5 summarizes the conclusions of the paper.

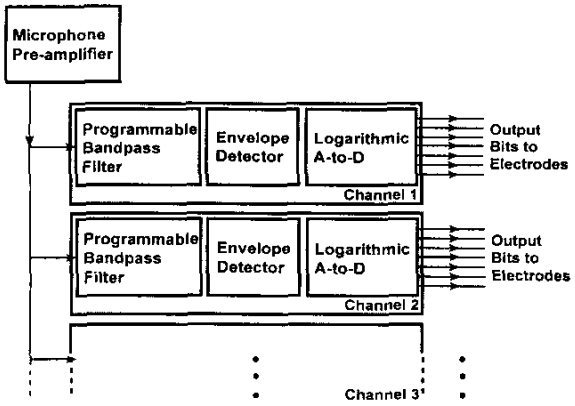

Figure 1. Overall Block Diagram of the Processor. The microphone pre-amplifier converts sound into an electrical input that is fed to each of the 16 channels. Each channel is composed of a programmable bandpass filter, an envelope detector, and a logarithmic A-to-D converter. Each channel converts the log spectral energy of the channel filter into digital output bits.

Subthreshold-MOS, silicon-cochleas, and analog circuits for cochlear-implant processing have been previously proposed (4)-(9) as means for implementing complex signal processing with very low power. This work proves the promise of such prior work by achieving numbers that make an analog processing system commercially feasible in the near term.

2. Analog Front End

The analog front-end circuitry consists of a microphone preamplifier that transduces sound from a JFET-buffered Knowles FG-3329 microphone into an amplified electrical output signal. To achieve good power supply rejection, critical in the highly noisy RF-and-mixed-signal environment of implant electronics, we sense the drain current of the self-biased JFET in the sub-miniature microphone, rather than its source voltage, and convert the current to voltage in a sense-amplifying configuration (10). The dynamic range of our 100Hz-10kHz analog front end is 80dB (50μV rms to 500mV rms at the front end output) although our channels only need a maximum of 60dB of internal dynamic range. A future implementation of our system will include an on-chip automatic gain control (AGC) system to compress the 80dB of dynamic range of the microphone front end into 60dB of internal dynamic range. The analog front end is described in great detail in (10). Note that other non-custom low-power microphone front ends could be used with our processor. We include the 64μW power consumption of the microphone JFET buffer and the 74μW power consumption of the analog front end in the measurements of total power consumption of our processor (Table 1 in section 4).

3. Channel Circuitry

A. Bandpass Filters

In the biological cochlea, high-frequencies excite motion of the basilar membrane in the base, while low-frequencies excite this membrane at the apex. The frequency discrimination of the cochlear system is necessary for the extraction of spectral content in auditory signals. The processor presented here performs the frequency-to-place transformation with a bank of fourth-order bandpass filters (second-order rolloff slopes on the high and low corners).

The bandpass filters were designed using a novel capacitive-divider GM-C topology (11). Traditional subthreshold GM-C filter topologies are limited by their linear range, making them ill-suited for voltage swings greater than a few thermal voltages (75mV). Novel feedback-linearization techniques such as gate degeneration and the use of the well as an input have been employed to attain wide dynamic range (12). The scheme used here, and reported in (11), uses capacitive dividers to obtain wide linear range and wide dynamic range. As the maximum internal dynamic range of an overall channel in our system is required to be 60dB, these filters were designed to obtain a dynamic range of 66dB, leaving room for other sources of noise in the channel.

B. Filter Programmability

The filter corner frequencies were programmed with a set of serially addressable current-mode digital-to-analog converters (DACs). Each channel has a 5-bit DAC to program its filter's low-frequency corner and a 5-bit DAC to program its filter's high-frequency corner. A 5-bit low-frequency-corner bus and a 5-bit high-frequency-corner bus were implemented on chip and shared amongst the channels during programming. The reference currents for these DACs were set with off-chip voltages. Future versions of the processor will use on-chip reference currents, to minimize the need for off-chip components and to provide temperature-and-supply robustness.

An alternative bandpass filter programming scheme, employs a tapered resistive line to bias the gates of subthreshold MOS devices. Each of these device currents may serve as the reference current of a filter DAC,

allowing the construction of a finely controllable and exponentially tapered filter bank.

C. Envelope Detector

To extract the energy in each channel's filter output, we rectify the signal, convert it to a current, and extract the peak value of this current. The envelope detector circuit uses a wide-linear-range transconductor (WLR OTA) described in (12), a class-B mirror, and a simple current-mode peak detector to perform these functions (13). A DC-offset correction loop ensures insensitivity to DC input conditions. The design is optimized for 60dB of dynamic range, consistent with our channel performance goals. A design suited for wider dynamic range operation with a more sophisticated peak-detection strategy is described in (14).

The peak detector produces asymmetric attack and release filtering of the rectified current and is described in (9), (13). The asymmetric dynamics are created by storing the peak current level in a source-follower-buffered current mirror with a linear gate-voltage release. The latter topology produces an instantaneous attack with an adjustable release. Designs to adjust the attack as well are presented in (14). The current-mode output of the peak detector makes the next stage of signal processing in the channel, namely, logarithmic conversion, relatively simple.

D. Logarithmic A-to-D Conversion

The logarithmic portion of this circuit converts the output current of the envelope detector into a logarithmic voltage by passing it into a diode. A reference current corresponding to the minimum output signal from the envelope detector is used to set a reference voltage on another diode. The linear difference between these two diode voltages is converted to a current by a wide-linear-range transconductance amplifier (12) and quantized in a dual-slope converter to obtain a logarithmic A-to-D. The overall scheme is insensitive to temperature variations. Feedback calibration is performed for offset cancellation during the auto-zeroing phase of dual-slope A-to-D conversion. The details are described in (15).

The sampling rate of the conversion is 312.5Hz, providing sufficient timing resolution to capture transients in speech signals. Many circuit innovations were employed to obtain sufficient bit resolution for this converter while maintaining very low power operation (15). Bionic ear patients typically do not exhibit more than 30dB of electrode dynamic range and can rarely discriminate more than a dB of change in electrode stimulation current. Thus, only thirty distinct electrode stimulation levels are necessary and, consequently a 5-6-bit logarithmic converter is sufficient. We use a 6-bit converter in this design.

4. Experimental Results

We fabricated a 9.23mm by 9.58mm die in 1.5um MOSIS AMI SCMOS process. The system was operated from a 2.8V supply. A die micrograph is shown in Fig. 2. To characterize the performance of the integrated circuit, we applied both acoustic and electrical stimulation.

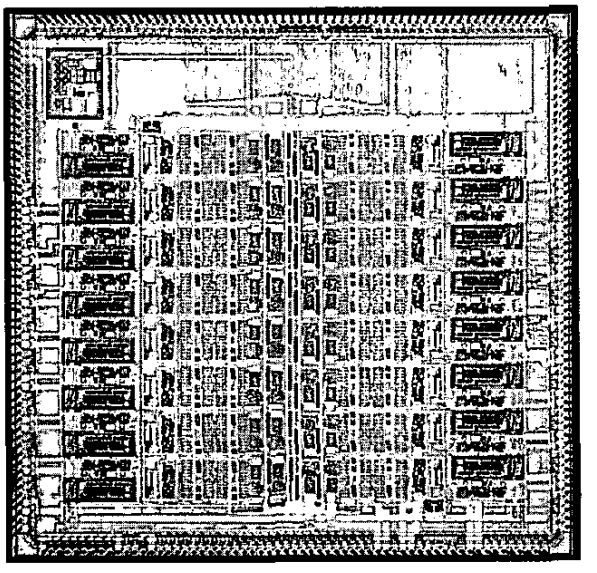

Figure 2. A die micrograph indicates the fabricated layout of the 16-channel processor. The analog front-end is in the top-left corner.

A. Single-Channel Speech Transient Response

Fig. 3 shows the response of successive stages in the signal-processing chain to the word "*test*". The top trace, shows the AC-coupled microphone-front-end output voltage. We see that the speech transient has low-frequency components corresponding to the vowel content of the word. This vowel phase lasts from roughly 0ms to 70ms on the *x*-axis of time. High-frequencies corresponding to the '*s*' sound in the speech are observed from 100ms to 150ms on the time axis.

The second trace from the top is the output of a bandpass filter set to respond primarily to low frequencies. We notice that the vowel information is left mostly intact while the high-frequency '*s*' is largely filtered out. The channel's envelope detector trace is shown next. The instantaneous attack of the peak detector is seen to follow the envelope of the bandpass filter output. The exponential

decay from these peaks corresponds to a release time-constant of roughly 10ms. The release time-constant may be altered by adjusting the leak current of the envelope detector. Finally, the very bottom trace indicates the integrated-and-sampled output voltage of the dual-slope A-to-D converter. We see that, because of the logarithmic nature of the A-to-D conversion, the exponential decay of the envelope detector is converted into a linear decay in the output bits of the converter. Thus, Fig. 3 yields a quick visual confirmation of all the processor's stages of operation.

B. Channel Frequency Response and Programmability

Fig. 4 shows the frequency responses of 8 logarithmically spaced filters with center frequencies programmed to range from 200Hz to 7kHz. To make the plot less cluttered, only 8 filter responses are shown rather than 16. The lowest center frequency corresponds to 100Hz and the highest center frequency corresponds to 11kHz, respectively. The logarithmic A-to-D circuit has a full-scale output of 64, corresponding to 6 bits. The input for this characterization was done at 200mVpp, corresponding to a fraction of the full-scale input. The logarithmic nature of the A-to-D distorts the filter pass-band shapes in the figure from their customary textbook shapes.

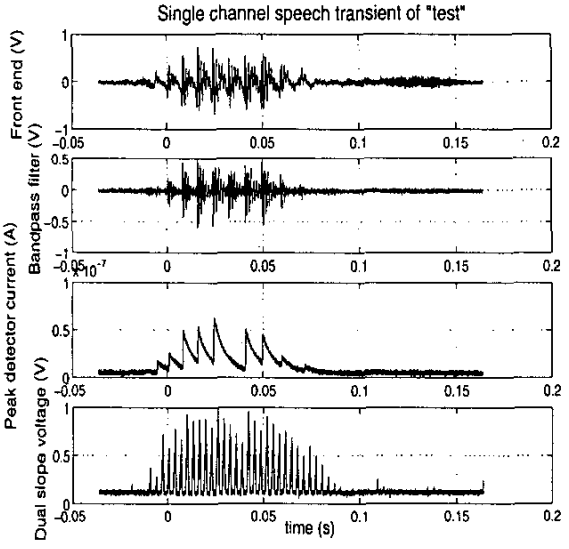

Figure 3. Response of various stages in the signal-processing chain to the word "*test*" are shown. The top trace is the AC-coupled microphone front-end output voltage . The second trace from the top is the AC-coupled output voltage from one of the low-frequency bandpass filters. The envelope detector output current and the dual-slope sampled A-to-D voltage are shown in the bottom two traces.

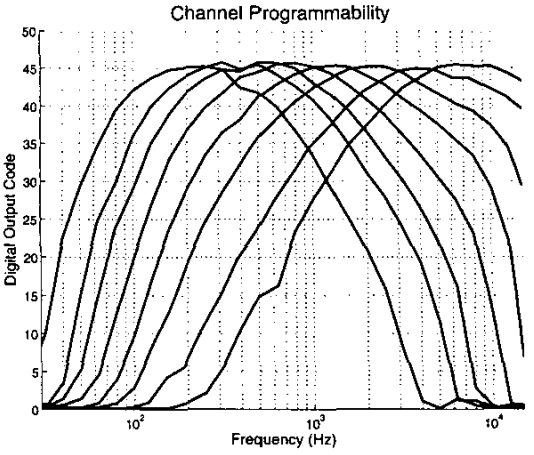

Figure 4. A-to-D response of logarithmically programmed channels to swept sinusoids. Sinusoid amplitude was 200mVpp. Note that the logarithmic nature of the A-to-D converter shows more filter overlap than a linear scaled plot.

C. Channel Dynamic Range

In this implementation, the bandpass filters and logarithmic A-to-D conversion circuits have more than 60dB of dynamic range. The dynamic range of the channel is limited by the dynamic range of the envelope detector and due to non-optimized interfaces between the various processing stages in each channel (3).

The channel dynamic range was characterized at the output of the logarithmic A-to-D in a separate test chip containing just two test channels. At 1kHz, the electrical input level is swept from 2mVpp to 2Vpp with a front-end gain of unity. Fig. 6 shows the A-to-D output as the input intensity is varied. We find that 51dB of purely logarithmic response is observed although there is a monotonic response to the input over the entire 60dB of dynamic range (3).

D. Microphone Front end Dynamic Range

The microphone front end's maximum undistorted input was measured at a limit of less than 1% total harmonic distortion (THD) to be 530mV rms at 1Khz. Fig. 7 shows the output spectra for this input. Since the microphone front end has a gain of 10 with respect to the JFET buffer's output voltage, this number corresponds to a 53mV rms input from the output of the JFET buffer or 110dB SPL of acoustic input. The output noise from the front end was measured at 52uVrms from 100Hz to 10kHz and is shown in Fig. 8. Referred back to the input this number corresponds to 5.2uV rms or an acoustic input of 29dB SPL.

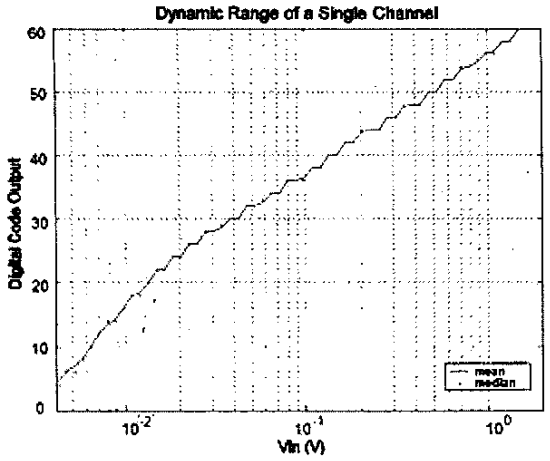

Figure 6. Dynamic range of the channel components is characterized from the logarithmic A-to-D response to sinusoids at 1kHz. This data was obtained from a companion test die (3).

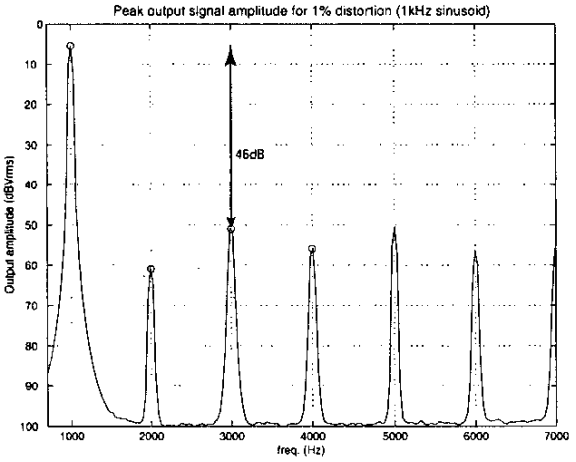

Figure 7. Total output distortion is shown with the largest output at 1% distortion for a 1kHz input tone.

Thus, the microphone front end dynamic range is greater than 80dB in the audio region of the spectrum.

E. Power Consumption

The power consumption of the various parts of the processor are summarized in Table I. We quote the power consumption for logarithmically spaced bandpass filters as that is the mode of usage of the processor in implants and speech front ends.

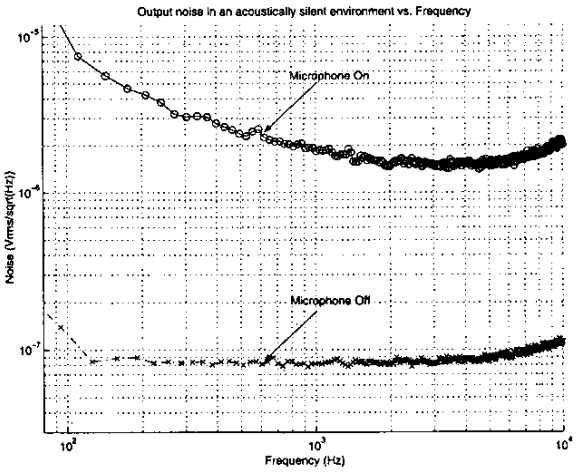

Figure 8. Output noise from the microphone front end is shown. The microphone self-biased JFET contributes most of the output noise during listening.

TABLE I

SYSTEM POWER CONSUMPTION

Signal Processing Stage	Power Consumption
Microphone JFET buffer	64μW
Analog Front End	74μW
16 Bandpass Filters	16μW
16 Envelope and Peak Detectors	126μW
16 Logarithmic Mapping A-to-D Converters	190μW
Total	470μW

Our previous work on single-channel processing blocks indicates that the power consumption of the processor could be lowered by as much as a factor of 2-3 below the numbers measured here (3). This integrated multi-channel implementation suffered from transistor mismatches across channels caused by the use of global gate-voltage biasing across channels. Thus, worst-case biasing was required to equalize channel gains and maintain performance. Future versions of this processor will use a low power current-bias-distribution network to remove the latter inefficiency. Even so, our power consumption is an order of magnitude below that of an A-to-D-then-DSP processor: We estimate that such a scheme would use about 0.25mW-0.5mW for the microphone front end + A-to-D, and 250μW/MIP x 20MIPS = 5mW for the other processing, yielding a total power consumption of about 5.5mW.

5. Conclusions

We have successfully demonstrated the operation of an analog VLSI processor for bionic ears and speech recognition front ends that lowers the power consumption of traditional A-to-D-then-DSP designs by an order of magnitude. It is digitally programmable and outputs bits that may be used over a digital communication link to stimulate electrodes or for further back-end processing in speech-recognition systems. The processor is thus suited for use in fully implanted bionic ear processors of the future or in portable speech recognition systems.

References

(1) P. Loizou, "Mimicking the human ear," *IEEE Signal Processing Magazine*, vol. 15(5), pp. 101-130, 1998.

(2) J. W. Picone, "Signal modeling techniques in speech recognition," *Proceedings of the IEEE*, vol. 81(9), pp 1215-1247, 1993.

(3) T. K.-T. Lu, M. W. Baker, C. D. Salthouse, J.-J. Sit, S. Zhak, and R. Sarpeshkar, "A micropower analog VLSI processing channel for bionic ears and speech-recognition front ends," in print, *IEEE International Symposium on Circuits and Systems*, 2003.

(4) R. J. W. Wang, R. Sarpeshkar, M. Jabri, and C. Mead, "A low-power analog front-end module for cochlear implants," presented at the XVI World Congress on Otorhinolaryngology, Sydney March 1997.

(5) W. Germanovix and C. Tourmazou, "Design of a micropower current-mode log-domain analog cochlear implant," *IEEE Trans. on Circuits and Systems II: Analog and DSP*, Vol. 47, pp. 1023-1046, Oct. 2000.

(6) T. S. Lande, J. T. Marienborg, and Y. Berg, "Neuromorphic cochlea implants," In *Proc. of the IEEE International Symposium on Circuits and Systems (ISCAS 2000)*, Vol. 4, pp. 401-404, 2000.

(7) R. F. Lyon and C. Mead, "An analog electronic cochlea," *IEEE Trans. on Acoustics, Speech, and Signal Processing*, Vol. 36, pp. 1119-1134, Jul. 1988.

(8) W. Liu, A.G. Andreou, and M.H. Goldstein, "Voiced-speech representation by an analog silicon model of the auditory periphery," *IEEE Trans. on Neural Networks*, Vol. 3, No. 3, pp. 477-487, May 1992.

(9) R. Sarpeshkar, R. F. Lyon, and C. A. Mead, "A Low Power Wide Dynamic Range Analog VLSI Cochlea", *Analog Integrated Circuits and Signal Processing*, Vol. 16, No. 3, pp. 245-274, Aug. 1998.

(10) M. W. Baker, and R. Sarpeshkar, "A Low-power High-PSRR Current-Mode Microphone Preamplifier," under preparation for publication, 2003.

(11) C. D. Salthouse, and R. Sarpeshkar "A practical micropower programmable bandpass filter for use in bionic ears," *IEEE Journal on Solid State Circuits*, vol. 38(1), pp. 63-70, 2003.

(12) R. Sarpeshkar, R. F. Lyon, and C. A. Mead, "A low-power wide-linear-range transconductance amplifier," *Analog Integrated Circuits and Signal Processing*, Vol. 13, No. 1/2, May/June 1997.

(13) M. Baker and R. Sarpeshkar, "A micropower envelope detector for audio applications", in print *IEEE International Symposium on Circuits and Systems*, 2003.

(14) S, M. Zhak, M. W. Baker, and R. Sarpeshkar, "A low power wide dynamic range envelope detector," under preparation for publication, 2003.

(15) J.-J. Sit, "A low-power analog logarithmic map circuit with offset and temperature compensation for use in bionic ears," S.M. thesis, Massachusetts Institute of Technology, Cambridge, MA, 2002.