

# CCD Correlators for Auditory Models

Richard F. Lyon

Apple Computer, Inc.  
Cupertino, CA 95014  
and

California Institute of Technology  
Pasadena, CA 91125

## Abstract

Surface-channel charge-coupled devices (CCDs) provide a mechanism for analog signal delay that can be built using an ordinary double-poly CMOS digital process, such as offered by Orbit through MOSIS. We have applied this technique to implement the correlation processing needed in auditory models of the sort proposed by Licklider for monaural pitch perception and sound separation. The resulting chips take analog audio in and produce analog video out (moving pictures of the correlogram representation of the sound) in real time. These chips present a variety of interesting analog and digital design challenges, which are addressed in this paper. Our first working experimental chips show real-time correlograms with 84 cochlea taps and 70 CCD delay stages per tap. There remains at least one unanticipated and as yet not understood problem with these circuits, resulting in a serious skew in the displayed correlation levels. Several circuit improvements are planned for the next generation of experimental chips.

## 1 Introduction

Modeling the cochlea has become a popular approach to extending the robustness of speech recognition systems. The cochleagram is a perceptually more relevant version of the commonly used spectrographic or short-time power spectrum family of sound representations. But when smoothed to a time scale comparable with the other representations, the cochleagram does not provide any real new richness of representation, or any particular leverage on the problem of separating out the desired speech from noisy or mixed sound environments.

Before smoothing, the cochlea output has more information, in the form of synchrony to sound. But the information is difficult to use directly, as it is still time-domain. Models that use temporal processing of cochlea outputs to arrive at a slowly changing multidimensional representation offer new possibilities for interpreting real-world sound mixtures.

Two such models, for binaural and monaural processing, were proposed long ago, based on theories of how human and animal auditory systems achieve their remarkable levels of performance. The binaural cross-correlation

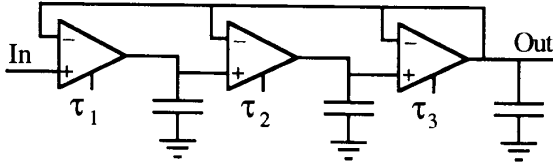
model of Jeffress [1] and the monaural auto-correlation model of Licklider [2] have enjoyed a resurgence of interest in recent years as the high computational requirements of these models have come within reach of experimenters. Patterson's "stabilized auditory image" approach [3] is a more recent example of temporal processing, which may also be described in terms of correlation, and produces results much like Licklider's.

After several years of experimentation with these models using digital computers [4, 5, 6, 7, 8], and after considering the possibilities of custom digital VLSI implementations, we decided instead to pursue analog VLSI as an exciting technology for building low-cost and low-power auditory models. After some work on analog cochleas [9, 10], we spent considerable effort on the problem of analog correlation, eventually focussing on CCDs to implement the required delays [10, 11]. After about a dozen MOSIS test chips, we arrived at the completely working combined cochlea and auto-correlator chip described below. The correlation operation dominates the computation and the silicon area (about 80% of the chip), and is the main topic of interest here.

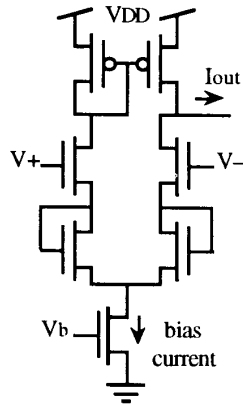
The moving-image output representation of the monaural cochlea and auto-correlation array, called a correlogram, has been described previously [6, 7, 8]. The chips discussed here display the correlogram on a standard TV screen in real time; analog audio input is converted to analog video output using all analog processing, in conjunction with digital timing and addressing.

## 2 Cochlea Circuits

A progression of variations and improvements to our cochlea circuits has been described before [10], and is continuing. For the chips described here, we selected the third-order "DIF3" stage of Figure 1, in which the first transconductance amplifier is built according to the circuit of Figure 2. The second and third amplifiers of each stage have a second pair of diode-connected transistors under each differential pair, further expanding their range of input linearity, to make the large-signal behavior of the stage compressive, rather than expansive [10]. Due to the more limited linear voltage range of the first amplifier, the output tends to compress toward about 250 mV peak-to-peak (P-P).



**Figure 1:** The circuit for the “DIF3” third-order filter, which corresponds well with our model of short-wave hydrodynamics. Three transconductance amplifiers are biased by the time-constant control voltages to operate in the MOS subthreshold (weak-inversion) region. A cascade of these filters constitutes a model of wave propagation in the cochlea.



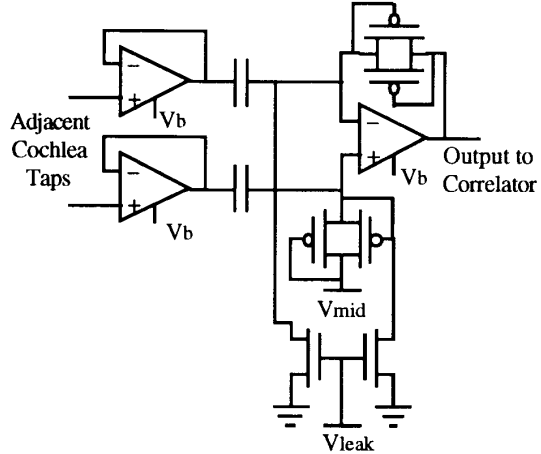
**Figure 2:** The circuit for the transconductance amplifier using diode-connected transistors to extend the linear input range and reduce the transconductance relative to the bias current.

A cochlea with 168 stages fits along one edge of the correlogram chip. For each pair of stages, a spatial differentiator (“spacediff”) circuit is used to provide a bandpass-like signal to each row of the 84-row correlator array. The nonlinear spacediff circuit, shown in Figure 3, is analogous to Mead’s hysteretic differentiator (“hysdiff”) [12], providing high gain and an output that compresses to about 1V P-P. By using a spatial difference instead of a time difference, the time constant of the differentiator scales naturally with the time constant of the cascaded filter stages (a geometric progression of time constants is controlled by a spatial voltage gradient that sets the subthreshold bias currents in the filter stages).

### 3 CCD Delay Lines

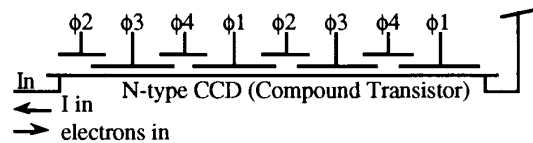
Jeffress and Licklider discussed the use of neural axonal propagation as a possible delay mechanism for implementing delay-dependent correlation operations [1, 2]. This continuous-time discrete-level mechanism works well when there are enough parallel channels (nerve fibers) to

form statistical averages over many fibers, and thereby achieve a nearly continuous-level representation for each frequency region. In VLSI, it is easier to build a continuous-level discrete-time delay mechanism, using clocked CCDs. Compared to continuous-time analog delay mechanisms, the CCD delay line has excellent delay-bandwidth product.



**Figure 3:** The “spacediff” circuit, a nonlinear spatial differentiator that provides high differential gain to small signals and compresses to about 1V P-P about a reference level  $V_{mid}$ . The first two amplifiers are used as followers, so that the cochlea does not see a large nonlinear load.

A CCD is basically just an extended transistor with a compound gate, with different gate segments driven by different clock phases [13]. There are a number of different clocking strategies, including two-phase and three-phase variations, but we chose a simple four-phase scheme with alternating large first-poly storage gates and small second-poly transfer gates, as shown in Figure 4. Charge packets of varying sizes are shifted through the CCD by proper phasing of storage and transfer clocks. For an N-type CCD, the charge carriers being pumped through are negatively charged electrons, so the input (source) current flows outward, as shown. At the far end, the drain terminal is connected to positive  $V_{DD}$  as an electron sink.



**Figure 4:** Schematic representation of a four-phase CCD delay line of two stages, showing the relation of poly1 gates for the storage phases ( $\phi_1$  &  $\phi_3$ ) and poly2 gates for the transfer phases ( $\phi_2$  &  $\phi_4$ ). The phase numbering we use is arbitrary and historical.

High-quality CCDs are generally built by adding a special implant to a CMOS or NMOS process, so that the charge is held in a "buried channel" below the imperfect Si-SiO<sub>2</sub> surface interface. The resulting devices are depletion mode, meaning they need clock voltages that are quite negative relative to the normal enhancement-mode N-type transistors. Special off-chip clock drivers and a special process are needed to take advantage of this high-quality approach. We elected instead to employ surface-channel CCDs, which can be built as extended enhancement-mode transistors in any ordinary MOS process that has two levels of poly gates. At the low clock rates needed for audio processing, the charge transfer inefficiency caused by slow trap states at the surface does not seem to be a problem. All the required clock phases are easily generated on-chip, and are driven across the array using limited-speed drivers to prevent charge dispersal.

#### 4 Correlation Detectors

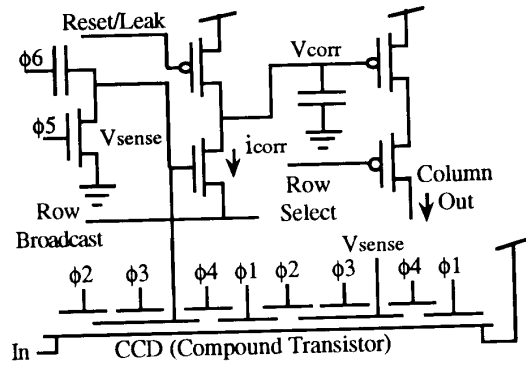
Given a CCD analog delay line, there are still several steps needed to make an auto-correlator. The charge packet must be non-destructively sensed at each delay stage, the signal represented by the delayed charge packet must be multiplied by the undelayed signal, and the instantaneous product must be integrated over a short time interval. All of this must be done with very few transistors, since the core cell is replicated for each pixel of the resulting correlogram output.

Nondestructive charge sensing is achieved using a floating-gate technique. The poly1  $\phi 3$  storage gate is used as a floating "sense node"; it is not driven directly, but rather by capacitive coupling from an overlying poly2 electrode (see Figure 5). The voltage to which the poly1 sense node rises when  $\phi 3$  is driven high on the poly2 electrode depends on how much charge is stored in the channel under the sense node. The voltage on the sense node when the entire charge packet is under it provides one input to the correlation multiplier. To determine the absolute level and timing of this signal, two extra clock phases are used:  $\phi 5$  resets the floating node to ground during the time that only  $\phi 1$  is high, and  $\phi 6$  defines the interval during which we wish to sense the charge, when only  $\phi 3$  is high. The six clock phases are shown in Figure 6.

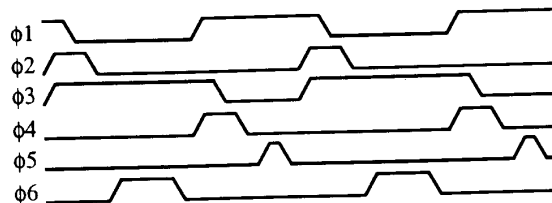
Since we view the output of the cochlea as representing something like a rate of neurotransmitter release, or a probability of neuron firing, we do not need to represent negative values; therefore, the multiplication needed is a one-quadrant operation. The simplest one-quadrant multiplier is a single MOS transistor for which the gate and source voltages are arranged to be logarithmically related to the signals to be multiplied. That is, since the MOS transistor in subthreshold and in saturation has a drain current that is exponentially related to both the gate and source voltages, the drain current may be interpreted as the product of the signals whose log voltages are present at gate and source. But the log nonlinearity does not need to be exact to get a reasonable correlation operation. If it is

omitted entirely, with proper signal polarities the transistor's drain current will be the product of exponentials of the inputs. For inputs swinging a few hundred millivolts, this nonlinear correlation approaches the digital AND operation, or a binary correlation.

In the correlator cell, the floating sense node cannot sink the multiplier's source-terminal current, so it must connect to the (zero-current) gate terminal. The source terminals of all the multiplier transistors in a correlator row are then tied together and driven by the undelayed signal.



**Figure 5:** The circuit for two stages of four-phase CCD delay line showing one correlation/integration/readout cell (there is actually one per stage). The  $\phi 3$  clock couples capacitively to the poly1 floating gate that non-destructively senses the charge in the CCD channel. The correlation current being integrated on the 1 pF storage capacitor is small except during the sense interval when  $\phi 6$  and  $\phi 3$  are both high and the other clock phases are low—then it can be as high as 1 nA. If the broadcast and  $V_{sense}$  voltages are appropriate logarithms of the input, the current  $i_{corr}$  can be an accurate one-quadrant product.



**Figure 6:** The six clock phase waveforms that shift charge through the CCD ( $\phi 1$ – $\phi 4$ ) and control the nondestructive charge sensing scheme ( $\phi 5$ – $\phi 6$ ). Two complete cycles are shown, starting with turning on the first transfer gate  $\phi 2$  to bring charge into the  $\phi 3$  storage site where it will be sensed during  $\phi 6$ .

The drain currents of the multiplier transistors are integrated on capacitors in each cell. The integrators are made leaky by the addition of a Reset/Leak transistor for each cell. These transistors can be driven digitally to reset

the integration, or as we use them, can provide a small leakage current that results in a nonlinear lowpass filter.

Two more transistors are used to sense the voltage on the integration capacitor and provide an output current on a column line when the row is selected.

To debug the concepts and tune up the cell design, we built a series of test chips using the cell of figure 5 and variations. For CCD input voltages from near ground (full charge packet) to about 1.5 V (empty charge packet), the sense voltage follows the delayed input signal with a gain of about one-third, which is adequate to provide plenty of gate-voltage variation.

## 5 Cochlea to CCD Interfaces

Both the charge input terminals of the CCDs and the undelayed "Row Broadcast" signals need to be driven via appropriate interface circuits to represent the approximate logs of the spacediff outputs from the cochlea. The circuit approaches that we have chosen are based more on getting levels that will work than on getting the nonlinearity right; for loud sounds, the 1 V P-P log-compressed output of the spatial differentiators overdrives the correlator array into a nearly digital mode. Combining this large-signal limiting behavior with a small-signal exponential behavior results in a sigmoidal "squashing" function, which at least potentially resembles the response of an auditory neuron.

Notice that the gate and source voltages at the multiplier need to be of opposite signal polarities; a high correlation current corresponds to a high gate voltage and a low source voltage. Therefore an inverting voltage stage is needed in one interface circuit. To get the operating point of the broadcast source voltage into a reasonable range, we decided to regulate the total correlation current while simply AC coupling the spacediff output onto the broadcast line. The total current is regulated through a series connection of P-type and N-type transistors, so that the low-level (expansive) and high-level (compressive) currents are both adjustable (see Figure 7); the DC operating voltage adjusts itself accordingly.

After experimenting with various inverting and level-shifting voltage input circuits for the CCD charge inputs, we decided to try a current-input technique. Using a current source as the input directly sets the charge packet sizes, settling to whatever voltage works. A source offset voltage (V4) on a current mirror scales the CCD input current relative to the current in an AC-coupled nonlinear voltage-to-current converter, or "current inverter." The complete interface circuits are shown in Figure 7. Note that for the CCD charge input, a full charge packet (low voltage) represents a zero signal, and an empty charge packet represents a maximum signal.

We intend to do further experiments using lower-gain interface circuits: a simple current-output differential pair instead of a spacediff, explicit log compression using diode-connected transistors for at least the broadcast signals, and a lower-gain charge input to get a smoother analog result.

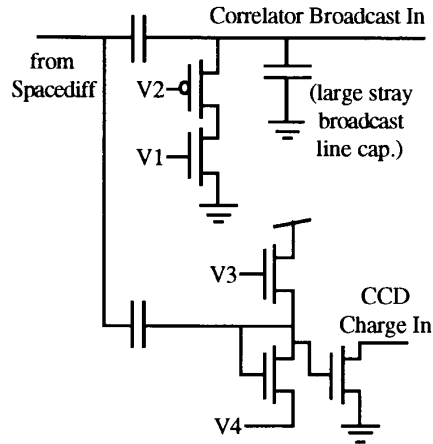


Figure 7: Circuits to interface a cochlea output to a row of the CCD correlator array.

## 6 Digital and Video Output Circuits

The correlogram chip produces real-time NTSC/RS-170 video output in cooperation with an external digital chip that generates NTSC timing (sync and blank) signals and pixel addresses; the digital chip is an Actel FPGA designed for us by Tanner Research. The correlogram chip has row and column address decoders along two edges to select the pixel of interest at any given time, and a sense amplifier and video output buffer amplifier to drive the pixel's analog correlation signal off chip. External bipolar transistors and a few resistors combine the video with the sync and blank signals to produce the final RS-170 compatible video.

The video timing and pixel rate are controlled by an external crystal oscillator. An independent digital clock is divided by 12 on chip to produce the 6 clock phases needed by the CCD correlators. Due to capacitive coupling in the correlator array, the CCD clocks tend to produce diagonal artifacts in the video; these artifacts are visible but not particularly noticeable, except when the CCD clock frequency is being adjusted, in which case the rapidly changing slopes of the artifacts draw attention to them.

## 7 Performance Results

We have arrayed these circuits to produce an 84-channel by 70-lag correlator array, which fits with a cochlea in the MOSIS 6.8 by 6.9 mm die size using 2-micron rules. The large number of free parameters (bias voltage "knobs") makes it a challenge to adjust the chip into a reasonable operating range. We have plans to incorporate circuits with fewer degrees of freedom and some self-adjusting biases in the next version to make the chip easier to set up and more robust. In particular, each cell will have a self-adaptive reset/leak current.

Our current chips produce familiar looking correlograms when the input sound is at a reasonable level. Figure 8 shows a still frame captured from a typical speech sound. Having better compression in the cochlea by integrating the AGC ideas that we have talked about [9, 10] would help it to deal better with loud and soft sounds.

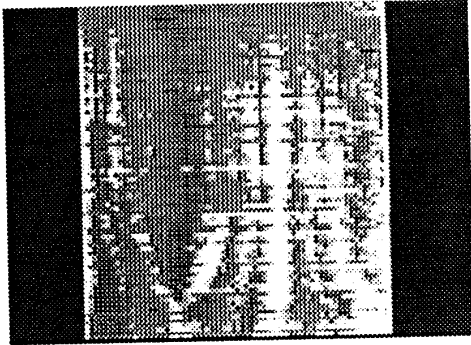


Figure 8: One frame of NTSC video captured from the correlogram chip output.

The current most interesting anomalous behavior is a severe tendency for the correlation currents to be higher toward the long-delay end of the array (toward the right in Figure 8) than they are toward the short-delay end, which makes it hard to get good correlogram images at low sound levels. This skew is consistent with the idea that charges in the CCD are somehow escaping. This idea is the opposite of the more common problem that a CCD accumulates a "dark current" or a "photo current" of additional charges via minority carriers swept up from the bulk of the silicon under the channel.

Our CCD is built using N-type channels in a P-Well CMOS process, because we figured that putting it in an isolated well would give it an environment free of minority carriers injected from the peripheral circuits or from stray light or thermally generated carrier pairs. For a CCD in a well, it may be that some charges that wander off from overfilled storage sites get collected by the well-substrate junction, rather than wandering back as they would in a CCD in substrate. Until we study it more, this is just a speculation. So far we have not been able to avoid the problem by being careful about not using full charge packets or by reducing clock edge rates.

## 8 Summary and Conclusions

The analog approach to auditory model implementation supports the possibility of putting a cochlea and a correlator array on a single chip, and operating them at a power low enough to consider for a battery-powered portable device. Using CCDs for the correlation delay provides a continuous-level delay line with precise delay timing, excellent delay-bandwidth product, and adequate signal-to-noise ratio. However, it is still not clear that digital custom VLSI approaches are not also competitive

for this particular application; the silicon area and power requirements will be higher with a digital approach, but the results can be made cleaner and more predictable. We will continue to pursue the analog path, while comparing both the quality and the implementation to digital alternatives.

Subsequent processing of correlograms to interpret them, to separate and classify sounds, etc., will be explored digitally initially, due to the flexibility of programmable supercomputers. But ultimately, as VLSI densities continue to increase, the in-place local computation approach of Mead's "Silicon Retina" and other "Neural Systems" [12] will probably be the most efficient way to implement the next level or two of processing.

## References

- [1] Jeffress, Lloyd A., "A Place Theory of Sound Localization," *J. Comp. Physiol. Psychol.*, **41**, pp. 35-39, 1948.
- [2] Licklider, J.C.R., "A Duplex Theory of Pitch Perception," *Experientia* **7**: 128-133, 1951.
- [3] Patterson, R.D. and Holdsworth, J., "A Functional Model of Neural Activity Patterns and Auditory Images," in Ainsworth, W.A. (ed.), *Advances in Speech, Hearing and Language Processing*, Vol. 3. JAI Press, London, in press.
- [4] Lyon, R.F., "A Computational Model of Binaural Localization and Separation," *Proceedings IEEE International Conference on Acoustics, Speech, and Signal Processing*, Boston, April 1983.
- [5] Lyon, R.F., "Computational Models of Neural Auditory Processing," *Proceedings IEEE International Conference on Acoustics, Speech, and Signal Processing*, San Diego, March, 1984.
- [6] Slaney, M. and Lyon, R.F., "A Perceptual Pitch Detector," *Proceedings IEEE International Conference on Acoustics, Speech, and Signal Processing*, Albuquerque, April, 1990.
- [7] Duda, R.O., Lyon, R.F., and Slaney, M., "Correlograms and the Separation of Sound," *24th Asilomar Conference on Signals, Systems and Computers*, IEEE, Maple Press, 1990.
- [8] Slaney, M., and Lyon, R.F., "Apple Hearing Demo Reel," Apple Technical Report #25, Apple Computer, Inc. (report and video available from Apple corporate library), Cupertino, 1991.
- [9] Lyon, R.F., and Mead, C., "An Analog Electronic Cochlea," *IEEE Trans. ASSP* **36**: 1119-1134, 1988.
- [10] Lyon, R.F., "Analog Implementations of Auditory Models," *DARPA Workshop on Speech and Natural Language*, Morgan Kaufmann, San Mateo CA, 1991.
- [11] Lyon, R.F., "Analog VLSI Hearing Systems," in Brodersen and Moscovitz (eds.), *VLSI Signal Processing III*, IEEE Press, 1988.
- [12] Mead, C., *Analog VLSI and Neural Systems*, Addison-Wesley, Reading MA, 1989.
- [13] Séquin, C.H., and Tompsett, M.F., *Charge Transfer Devices*, Academic Press, New York, 1975.