CHARGE-DOMAIN ANALOG SIGNAL PROCESSING FOR DETECTOR ARRAYS

Eric R. FOSSUM

Department of Electrical Engineering, Columbia University, New York, NY 10027, USA

The use of charge-coupled device structures for analog signal processing is described. By operating in the charge domain, such structures offer the advantages of low-power, compact-layout, and high-speed operation. The circuits are digitally programmable and unlike continuous-time analog circuits, their operation can be fully synchronous. The application of the charge-domain circuits to on-chip signal processing for detector arrays is discussed. A new content-addressable architecture for sparsely illuminated detector arrays is proposed.

1. Introduction

Charge-coupled devices (CCDs) have emerged as a technology of choice for radiation detector array readout. CCDs, which operate in the charge-domain, offer several advantages over other multiplexer technologies. CCDs are real-estate efficient in layout leading to high resolution imagers and are low-power due to their capacitive nature. Since signal charge is conserved, parasitic effects found in direct readout or random-access architectures are avoided. With buried-channel CCD structures, charge loss due to transfer inefficiency is minimal and transfer efficiency in excess of 99.999% is routinely achieved. Noise in these devices is well behaved and often dominated by the output amplifier rather than the multiplexer itself. Fabrication technology for these devices has improved dramatically in the past few years. Multimegapixel CCD imagers are now commercially available at costs comparable to CMOS ASICs (complementary metal-oxide-semiconductor application-specific integrated circuits).

Direct detection of photons and energetic charged particles is achieved by the collection of carriers generated in the semiconductor (typically silicon) within reach of the CCD-potential well. Separation of the CCD and the detector is also possible with the detector transducing the generated carriers into a CCD signal charge. Several schemes have been demonstrated such as direct injection (current-domain), gate-modulation (voltage-domain) and diode cutoff. The detector may be integrated monolithically or in a hybrid configuration. In the latter, the detector is electrically connected to the CCD readout using wire bonding or a "flip-chip" method. For linear arrays, wire bonding may be a simpler technology, but for areal arrays, the flip-chip method provides superior performance and has become well established for aerospace applications.

However, as the image acquisition technology has evolved, a new significant problem has surfaced which is the real-time signal processing of acquired images. Only the most massively parallel digital computer architectures are capable of handling the vast amounts of data collected by a detector array in real time.

In this paper, an approach for overcoming this problem using CCD-like circuits for analog signal processing is described. Such analog charge-coupled computing circuits feature the same low-power, compact-layout advantages of CCDs, while retaining the programmability normally associated with digital computing.

2. Charge-coupled computing

Analog signal processing has traditionally been performed either in the voltage or current domains using active amplifier-based circuits. Charge-domain signal processing is becoming recognized as an alternative approach because of its lower power consumption and its synchronous nature [1–5]. Charge packets are manipulated using capacitive structures to effect arithmetic operations. These structures are referred to as charge-coupled computing circuits and a variety of operations have been demonstrated.

A simple example of charge-coupled computing is addition, as shown in fig. 1. Two charge packets initially confined under the two outer electrodes are added by merging them under the center electrode. The time required for this operation depends upon the geometry of the electrodes and the desired accuracy. The time to add the packets scales as the square of the electrode length and logarithmically with the desired accuracy. For example, with a 5 μm electrode length, approximately 5 ns are required for 1 part in 1000 accuracy. This is equivalent to ten digital bits. The energy required for the operation is partitioned into energy dissipated in the semiconductor and energy dissipated in charging the electrodes. A rough rule of thumb is that the energy is approximately $2QV$, where Q is the size of
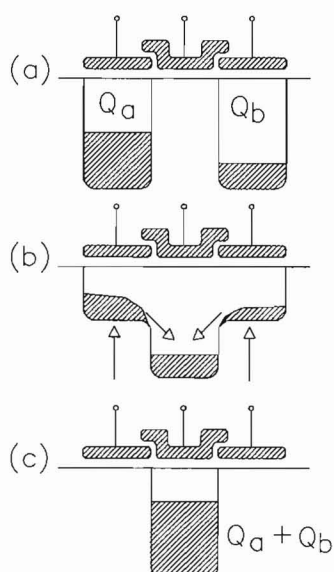
Fig. 1. Illustration of addition in charge-domain.

Fig. 2. Illustration of wire charge transfer.

the charge packet and V is the bucket depth in volts. For a 1 pC charge packet and a 5 V bucket, the energy is 10 pJ. The real-estate required for the adder, assuming an electrode aspect ratio between five and ten, would be approximately 1000 μm^2. In comparison, a 10 bit CMOS adder might be expected to require an order of magnitude more time, two orders of magnitude more energy, and three orders of magnitude more real-estate [6]. Note however that the power remains approximately equal.

The speed, power, and real-estate advantages of charge-coupled computing circuits are clear in the above example. However, one of the disadvantages of using CCD-like circuits is that to achieve accuracy exceeding approximately 10 equivalent bits in more complex operations is difficult in consideration of charge-transfer efficiency, carrier trapping, and fundamental circuit accuracy. Furthermore, since CCDs are dynamic devices, degradation of the signal can occur in time. Thus, applications for charge-coupled computing must be carefully chosen. One particularly promising application is focal-plane image processing [4].

Operations more complex than addition include charge packet regeneration, splitting, routing, differencing, and magnitude comparison. These circuits have been described in detail elsewhere [4,7–10]. The circuits are designed as building blocks for use in more complex functions. Charge is input and output from the block using "wire transfer" as shown schematically in fig. 2. Thus, charge-coupled computing circuits are a hybridization of CCDs and bucket brigade devices [11]. The use of wire charge transfer permits the use of circuit topologies previously denied by CCD designers. Par-
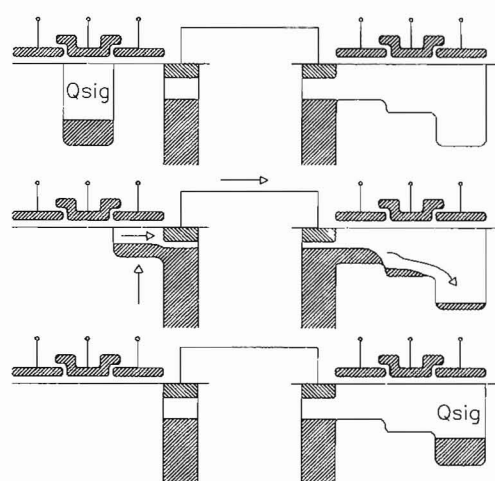
ticularly, the crossing of signal charges is readily achieved with wire transfer, yet nearly impossible with conventional CCD restrictions.

By combining the building blocks, it is possible to construct a simple digitally programmable analog signal processor, or charge-coupled computer. A prototype circuit was fabricated and tested previously [4], and a second generation version is currently under fabrication. This **charge manipulation processor** (CHAMP) integrated circuit is shown in fig. 3. It consists of a charge input section, a charge output section, a splitter, router,
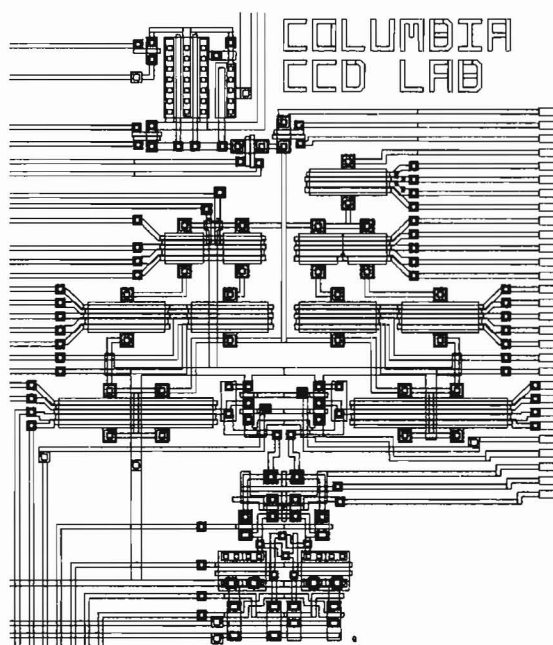

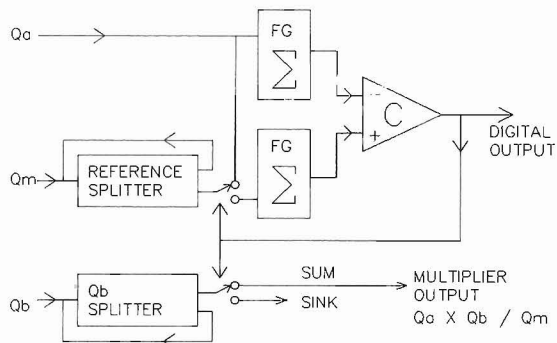Fig. 3. Layout of the charge manipulation processor (CHAMP) chip.

II. CONTRIBUTED PAPERS

Fig. 4. Block diagram of the CCD serial A/D and D/A circuit.

several buffers and a magnitude comparator. The comparator is configured in a nondestructive voltage-domain mode, with the voltage derived from two reset-floating-gate amplifiers. The programming of CHAMP is achieved by the sequence of clocking waveforms applied to the circuit electrodes.

The same circuits can be combined in a building block fashion to perform other operations as well. An example of this is the serial analog-to-digital (SAND) circuit [12] which uses a successive approximation algorithm to perform the conversion. A block diagram of the SAND chip is shown in fig. 4 and the actual circuit layout is shown in fig. 5. This circuit, which measures approximately 200 μm by 300 μm is expected to perform A/D conversion at the rate of approximately 10^6 conversions/second with an accuracy of 8 bits (1 part in 256). A multiplying digital-to-analog (D/A) converter which is slaved to the A/D is included in the layout. The output charge packet is essentially the product of the A/D input charge packet and D/A input charge packet divided by the reference charge packet.

3. Application to detector arrays

The compact nature of the charge-coupled computing circuits makes it possible to consider placing an array of simple charge-coupled computers on a single chip. Such an array operating as a single-instruction, multiple-data (SIMD) parallel processor can perform image preprocessing functions such as thresholding and edge detection as well as A/D conversion.

In a chip designed and under fabrication [13], a 24 × 24 array of processors is monolithically integrated with a 48 × 48 array of p–n junction photodiodes. Each processor serves four photodiodes and can communicate with its nearest neighbors. Designed for focal-plane image processing, each processor has charge-coupled
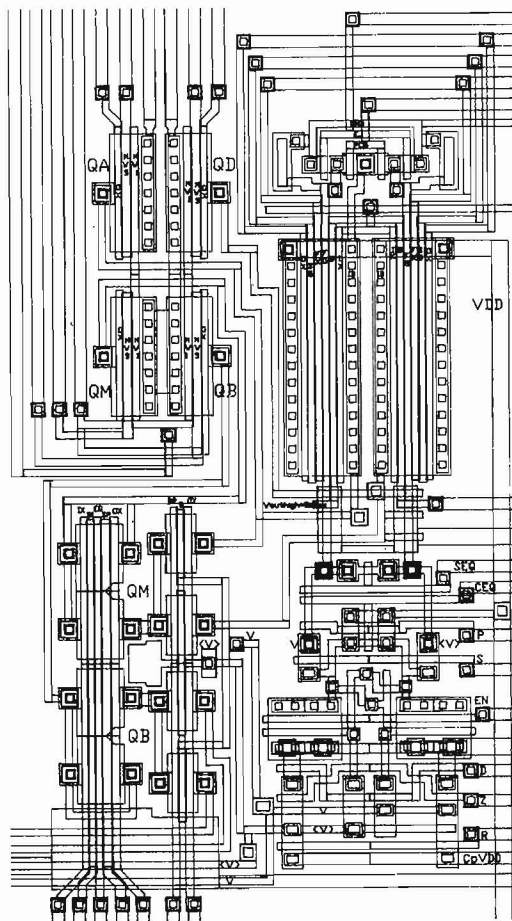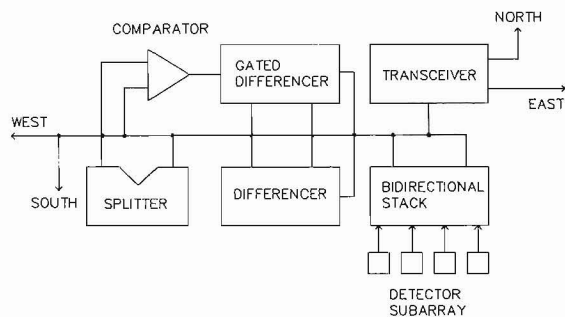

Fig. 5. Layout of the serial A/D and D/A circuit. Upper left quadrant is for forming charge packets Q_a, Q_b, and Q_m. Upper right quadrant contains the floating gate sense amplifiers for the comparator, which is in the lower right quadrant. In the lower left quadrant are the two splitters and routers for generating the reference packets for the A/D and D/A respectively.

Fig. 6. Block diagram of one of the focal-plane image processor processing elements. The differencer and comparator components are described in ref. [4], and the splitter in ref. [8]. The bidirectional stack is a CCD parallel–serial shift register with four extra memory locations. More details may be found in ref. [13].
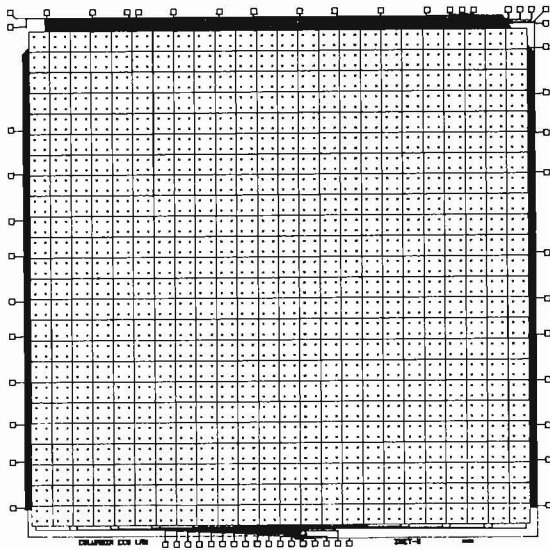
Fig. 7. Layout of the focal plane image processor chip.

computing circuits for servicing the four photodiodes, nearest neighbor I/O, charge packet splitting, magnitude comparison, and differencing as shown in the processor element block diagram of fig. 6.

A partial chip layout illustrating the array of photodiodes is shown in fig. 7. The photodiodes are on a 180 μm pitch and the full chip is 9.4 mm by 9.4 mm.

The throughput of the processor array is expected to be significantly higher than the chip readout rate. With 50 ns clock widths, and allowing for 500 clock cycles to process one pixel (4 pixels per processor), the total time to process a frame is 0.1 ms or 10 000 frames per second. Since a serial readout technique is used for the processor array, the readout rate represents the throughput bottleneck. However, in practice frame rates of only 100 Hz are required for most machine vision applications so that significantly slower clocking rates can be permitted. Total power consumption at a 100 Hz frame rate is estimated to be less than 1 mW.

It can be observed in fig. 7 that the "fill-factor" or fraction of real-estate utilized for photodetection is small. This is due to the monolithic integration of photodiodes and proacessor structures. A hybrid of flip-chip approach can significantly enhance the effective fill factor. The photodiodes sites of fig. 7 can be viewed as sites for the flip-chip solder bumps.

The throughput of the spatially-parallel approach described above exceeds that needed for many applications. Chip real-estate required for pixel processing can be traded for a lower degree of parallelism. (In the chip described above, a factor of 4 has already been traded.) Processors could be located at the bottom of each column resulting in a vector-processor-like architecture. If the processing is restricted to 3 × 3 kernels at a

relatively low data rate, a pipeline processor could be employed after the serial multiplexer prior to readout. Some reformatting of the data is required to reestablish the spatial proximity of the 3 × 3 kernel [3].

4. Content-addressable architecture for sparse illumination

Detector arrays used for high energy particle or photon detection are, in general, sparsely illuminated. Successive frames may be completely empty, and captured events may illuminate only a few dozen pixels. Information regarding the presence, location and magnitude of these illuminated pixels is required by the user in relatively short time. Conventional CCD multiplexers are ill-suited for such applications since the presence of illuminated pixels can only be determined by reading out the full frame (unless generated carriers not collected in the bucket can be detected). The CCD multiplexer provides the sparsely illuminated user with too much information at too slow a rate.

Faster readout may be achieved with random access readout architectures provided it can be determined which pixels to read. Architectures which flag or bound the coordinates have been proposed recently (see other papers in these proceedings). However, these architectures remain inefficient in locating the illuminated pixels. Furthermore, the analog output is susceptible to noise and capacitive loading. A new architecture which is similar to that found in an associative memory is proposed. This is a content-addressable readout structure.

The basic idea in the content-addressable readout structure is two-fold. First, the array must signal the user as rapidly as possible if at least one pixel has been illuminated above a preset level. The approximate number of illuminated pixels is also of interest. Second, only those pixels which are illuminated should be read out.

It is proposed to use the architecture schematically illustrated in fig. 8. The unit cell for each pixel consists of a resettable but nondestructive low-power charge-to-charge or charge-to-voltage amplifier. Signal amplification as close as possible to the detector will minimize readout noise and loss, as well as allow for multiple operations with each pixel. The amplifier would preferably be linear in gain, though logarithmic compression might be useful in some applications. A possible candidate would be a charge packet replicator configured for gain [9].

The amplifier output is compared to a global threshold level. This threshold level is the "content" to be addressed and what is needed is the address of a pixel with a level above this. The comparison can be implemented in a number of ways, but a cross-coupled flip–flop optimized for low power consumption would work well [10]. The comparator output is used in two
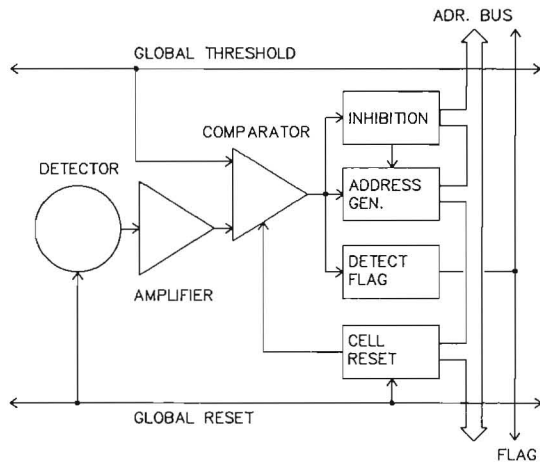
Fig. 8. Content-addressable readout scheme for a detector array.

5. Conclusions

The use of charge-coupled computing circuits for detector array real-time signal processing has been described. Spatially parallel image processing can be achieved by integrating an array of charge-coupled computers with the detector array.

A new content-addressable readout scheme which minimizes readout time and complexity has been proposed. An array of unit cell amplifiers, comparators and address generators can be used to perform implicit A/D conversion and to rapidly scan arrays for illuminated pixels.

ways. First, the user is signalled that at least one pixel compares positively. Outputs might be summed to yield an analog estimate of the number of pixels which compare positively. Second, circuitry for placing the address of the pixel on a global bus is enabled. Simultaneously, all other pixels are inhibited from placing their address on the bus. Alternatively, bus collision can be detected and the global threshold set to a new level which avoids collision. The latter technique is more cumbersome in some ways than the inhibition scheme.

At the end of the comparison, the address of one pixel meeting the global criteria is on the bus. That pixel can then be selectively disabled or reset and the next pixel located.

The global threshold level can be used in a number of ways. It can be used to rapidly determine if any pixels have been illuminated above a certain level. The level can be ramped or scanned to histogram the array. Ramping can also be considered a time-domain "flash" A/D conversion. Pixels readout using a ramp technique are automatically rank sorted.

This scheme works best for sparse illumination in which perhaps a few dozen pixels would be illuminated. It is noted that cell-to-cell process variations will appear as fixed pattern noise but that MOS flat-band voltage shifts induced by the array environment can be cancelled out by straightforward circuit design.

It is estimated that implementation of the above unit cell in a mix of charge-coupled computing circuits and CMOS could lead to a unit cell pitch between 100 and 200 μm, or 10000 to 40000 pixels per cm^2. Power consumption would occur only in illuminated pixels and thus be minimal. Frame rates exceeding 10 MHz would be achievable with detection and readout within a single frame.

References

[1] E.R. Fossum, Charge-Coupled Analog Computing Elements and Their Application to Smart Image Sensors, Ph.D. thesis, Yale University (1984).
[2] J.D. Joseph, P.C.T. Roberts, J.A. Hoschette, B.R. Hanzal and J.C. Schwanebeck, Proc. SPIE 501 (1984) 238.
[3] G.R. Nudd, P.A. Nygaard, G.D. Thurmond and S.D. Fouse, Proc. SPIE 155 (1978) 15.
[4] E.R. Fossum, Opt. Eng. 26 (1987) 916.
[5] A.M. Chiang, Proc. SPIE 827 (1987) 126.
[6] See, for example, M. Shoji, CMOS Digital Circuit Technology (Prentice Hall, 1988).
[7] T.L. Vogelsong and J.J.Tieman, IEEE J. Solid-State Circuits SC-20 (1985) 562.
[8] S.S. Bencuya and A.J. Steckl, IEEE Trans. Electron Devices ED-31 (1984) 1494.
[9] E.R. Fossum and R.C.Barker, IEEE Trans. Electron Devices ED-31 (1984) 1784.
[10] R.E. Colbeth, N.A. Doudoumopoulos, E-S. Eid, S.E. Kemeny, A. Montalvo and E.R. Fossum, Columbia University, CTR Tech. Rep. (1987).
[11] F.L.J. Sangster, IEEE Int. Solid-State Circuits Conf., Dig. Tech. Papers, XIII (1970) 74.
[12] S.E. Kemeny and E.R. Fossum, unpublished.
[13] E-S. Eid and E.R. Fossum, to appear in Proc. SPIE 977 (1988).

Questions

Q. (R. Chase, LAL, Orsay): I am concerned about the splitter, because there you need very good precision, better than 0.5%. How do you do that?

A.: The charge packet splitting was demonstrated in thesis work in RPI by Bencuya and Steckl (see ref. [8]). Experimentally they showed the splitter is very accurate. Splitting occurs due to field oxide or isolation pattern. The accuracy of mask making tool is within 0.1 μm. The main cause of noise is a ripple at the surface (within the order of kT/q) but also this effect is small as one typically operates at 5 or 10 V.

Q. (E. Gatti, Politecnico Milano): You steer your charges with voltages. In a cascade of operations like successive approximation ADC, you have to look at the result of each of your operations and go back from charges to voltages. What is the cost of this in terms of power?

A.: The comparator is the biggest user of power on that circuit and it is the flip–flop that does the comparison. We turn the flip–flop on so that current flows from the positive to the negative rail only over a very short time interval. The design is such that the current that flows, is enough to charge the capacitance of the router stage. It is almost as if we are doing it in the charge domain.

Q. (N. Allinson, York): In the pixel processor it seems to me that the communication to the nearest neighbours is not symmetrical, while I would think that, if you do a 3×3 convolution you want a symmetric neighbourhood.

A.: It only seems that way, but it works. We can organize clocking cycles such that we can transmit in all directions and diagonally as well.

Q. (N. Allinson, York): So the circuit is synchronously clocked. How many clock cycles are needed to do a 3×3 convolution?

A.: We are running at 50 ns clock cycles and typically one needs 200 operations to process one picture element, so this gives 10 μs. The bottleneck is the time to read the information out of the array.

Q. (R. Horisberger, CERN): Could you explain the principle of the differencer you mentioned?

A.: We can regenerate exact copies of input signal charges. By bringing the charge to be subtracted onto a second gate we can generate the difference. There turns out to be good linearity.

Q. (H. Williams, Univ. Pennsylvania): Does your estimation for power consumption include the power to drive the clock lines? The interconnects seem to represent a lot of capacitance.

A.: The rough rule of thumb for that is to multiply by 2. The buckets are made on a 60 nm thick oxide whereas the Al clock lines run over a 1 μm thick oxide. It turns out that the total contribution of both is about the same.

Q. (E. Heijne, CERN): Can the optical coupling you do with GaAs also be done with Si? This would be useful to avoid problems of clock feedthrough to neighbouring detectors in large experiments.

A.: It can be done but for a good performance on Si the fibers should come in laterally, on the sides of the chip. Such techniques are becoming widely used. You can get the speed that you need, but you may be limited by the power you can burn in your amplifier, and the capacitance of the light detector. Please see R. Ade et al. IEEE Trans. Electron Devices ED-34 (1987) 1283 and R. Ade et al., Proc. SPIE 881 (1988) 199.

Q. (S. Shapiro, SLAC): What speed can be achieved?

A.: The possible speed is 50 MHz (maybe 100 MHz) and it is mainly limited by the detector of the light. One may choose different wavelengths, e.g. green for optimal operation.