TA 1.2: Camera on a Chip

Bryan Ackland and Alex Dickinson

AT&T Bell Labs, Holmdel, NJ

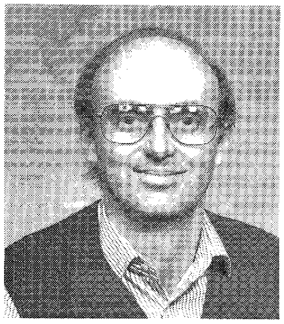
Introduction

Development of low-cost video camera technology has, for many years, been driven almost exclusively by the camcorder (3 million units per year in the U.S.) and security (1 million units per year) markets. The typical product is a multichip camera subsystem consisting of CCD sensor, clock drivers and analog signal-processing devices to provide color balance and exposure control. Output is analog NTSC or PAL. This picture is changing.

Recent advances in video compression and digital networking technology, combined with the ever increasing power of PCs and workstations, are creating enormous opportunities to develop new multimedia products and services built upon sophisticated voice, data, image and video processing. This will create a significant demand for compact, low-cost, low-power electronic cameras for video and still image capture. These cameras will be a standard peripheral on all PCs bundled for multimedia applications. Given that in excess of 60M PCs will be sold this year, a sizable new market for electronic cameras is being created.

Present NTSC cameras are not well suited to digital multimedia applications: (1) They are too expensive - OEM cost typically greater than $100. Experience with the development of CD-ROM players suggests that as the OEM cost of a peripheral falls below about $50, the vast majority of PCs will ship bundled with the peripheral. (2) They are too large to be mounted inconspicuously on (or in) a PC monitor. (3) They consume too much power for portable applications. CCD sensors require high-voltage, high-current clocks - a CCD-based camera subsystem will typically consume 1-2W. (4) Video data is output in encoded analog raster scan format. This limits both the flexibility of the camera and the cost of the overall system. In multimedia systems based on camcorder type cameras, additional circuitry must be included to convert from the NTSC style output of the camera to the digital format required by the application, as shown in Figure 1. This circuitry frequently costs more than the camera itself.

What is needed is a camera technology that can be customized to particular applications such as desktop video, still scene, and document imaging. Cost, size and power constraints require integrating the image sensor along with analog and digital signal processing and interfacing elements onto the same die, as shown in Figure 2. Input to the chip is an image focused onto a sensor array. Output is a (possibly compressed) digital stream that connects seamlessly to the specified multimedia platform.



Bryan Ackland

Bryan Ackland received the BSc in Physics from Flinders University in 1972 and the BE and PhD EE from University of Adelaide, Australia in 1975 and 1979. In 1978 he joined AT&T Bell Labs as Member of Technical Staff. In 1986 he became Head of the VLSI Systems Research Department at Holmdel, NJ. His interests included raster graphics, symbolic layout and verification tools for full-custom VLSI, MOS timing simulation and VLSI layout synthesis.His interests now are VLSI architectures and design tools for high-performance signal processing and communications, particularly multimedia. He is an AT&T Bell Labs Fellow and an IEEE Fellow.

CCD Sensors

Any solid-state imaging device consists of an array of sensing elements combined with some form of transport mechanism to deliver these sensor outputs to the periphery of the die. Sensors used in commercial devices include photodiodes, MOS capacitors, charge injection devices and bipolar phototransistors. All of these devices use essentially the same light sensing mechanism. Photons penetrating a depletion region generate electron-hole pairs. These are swept away by the electric field across the depletion region and generate a small photocurrent.

Except under very bright light conditions, it is not possible to use this photocurrent directly. Even at 100% conversion efficiency, a 10μm sensor illuminated at 1lux will generate a photo current of only 70fA. To achieve reasonable signal-to-noise ratio, these currents are usually integrated (typically for 15ms) to produce an accumulated charge output. For the above example a charge of 10^{-16} coulombs (6000 e-) would be accumulated in this time.

The CCD provides a simple mechanism for transporting these small charge packets out of the array. A CCD is a linear array of

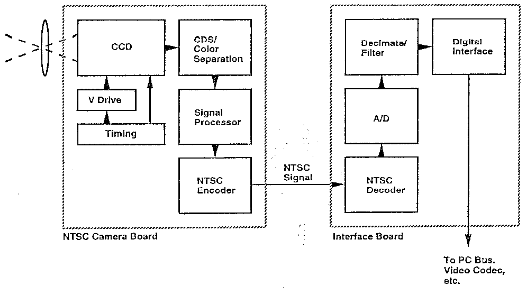


Figure 1: Conventional multimedia camera.

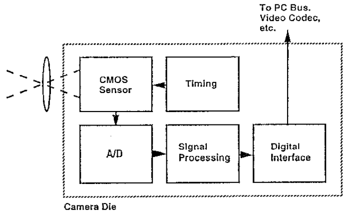


Figure 2: Camera on a chip multimedia camera.

 0-7803-3136-2 / 96 / $5.00 / © IEEE

MOS capacitors that function as a charge-domain shift register when driven by a set of multiphase clocks. In low-cost cameras, interline transfer, as shown in Figure 3, is the most commonly used architecture. Adjacent to each column of sensors is a vertical CCD. Accumulated charge is transferred from the sensor to the corresponding CCD bucket. These charges are then shifted vertically, one line at a time into a horizontal shift register. A single line is scanned out from the horizontal shift register onto a capacitor which converts each charge packet to a voltage for subsequent amplification and buffering.

CCD sensors have improved dramatically since introduction in the 1970s. Scientific arrays of 4096x4096 resolution with dynamic noise levels of 3-5 electrons and a dynamic range of over 80dB have been demonstrated. Low-cost commercial devices typically provide 640x480 pixel resolution with a signal-to-noise ratio of 45dB. One advantage of CCD sensors is the high fill factor (ratio of light sensitive area to total pixel area) that can be obtained, even for small pixels. This is because the only extra area required is that of the CCD register. State of the art sensors use pixel sizes of 5x5μm, which approaches the optical diffraction limit. A second advantage of these systems is their lack of pattern noise. This is caused by variations in offset and gain from one pixel to another. In a CCD, all image information travels through closely matched paths in the charge domain, and then shares the same charge-to-voltage output stage.

CCDs, however, have a number of disadvantages in multimedia applications. Most of these stem from the need to maintain high charge-transfer efficiency (CTE) within the CCD shift register. CTE is a measure of the percentage of electrons that are successfully transferred from one bucket to another in one CCD shift cycle. In a 640x480 sensor, a single charge packet may be shifted 1100 times. Even with a CTE of 0.9995, this will result in a 40% loss of charge by the time the packet reaches the output stage.

One way of increasing CTE is to use 12-15V on the CCD clock lines. This, in turn, leads to high power dissipation. There has been much work recently in trying to develop low-voltage CCD processes better suited to multimedia applications. Low-power consumer-grade sensors have been reported in which most of the high-speed clock circuitry is driven at 3.3V [1, 2].

Another approach is to tune doping profiles to maximize charge carrying capacity and minimize charge loss during transfer. This has lead to processes that provide excellent CCD performance but are not at all suitable for producing standard digital VLSI circuits. Digital camera proposals based on CCDs [3] typically require at least three separate die: one for the sensor, a second for analog clock drivers and A/D and a third for the digital signal processing components [3]. However, a recently reported a 2μm process supports high-quality CCDs, along with npn bipolar and

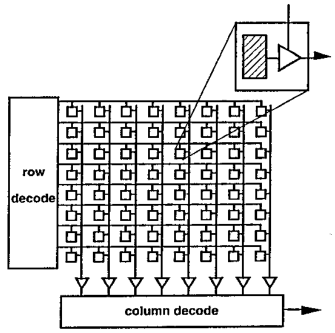


Figure 4: Active pixel sensor array.

conventional CMOS devices [4]. Such a process allows for the possibility of integrating all camera functions onto a single chip at a cost of four extra mask layers and three extra implants.

CMOS Sensors

An alternative approach is a two-dimensional addressable array of sensors [5-8]. The architecture is similar to that used in conventional random-access memories. A bit line is associated with each column of sensors as shown in Figure 4. A row-enable line allows each sensor in a selected row to place its output onto its bit line. A multiplexer at the end of the bit lines allows for individual column addressing. This is an old idea dating back to the early 1970s, but one that, until recently, has not found commercial application. The difficulty arises from the very small charge generated at each pixel. From our previous example, a charge of 10^{-16} coulombs placed on a 2pF bit-line, generates a voltage change of 50μV. Such a signal is susceptible to noise. What is needed is a simple amplifier at each pixel to provide buffering as shown in Figure 4. This is referred to as an active pixel sensor (APS) array.

In the early 70s, with sensor sizes of 25μm and design rules of 5μm, it was not possible to include an amplifier at each pixel. Today, with sensor sizes of 10μm and process design rules at 0.5μm, a three transistor pre-charge/amplify/select circuit, as shown in Figure 5, can be built and still achieve a fill factor of over 25%. One advantage of APS sensors is that they can be powered from a single 3.3V (or lower) supply and do not require external multiphase clock generators. This leads to reduced system costs and significantly reduced power dissipation. Integrated CMOS cameras with a total power dissipation under 10mW have been reported. Silicon processing costs are also reduced because of the huge volumes associated with generic digital CMOS production. Another advantage is the flexibility that is provided by a random access sensor addressing. No longer is one constrained to access pixels in serial order determined by the architecture of the CCD

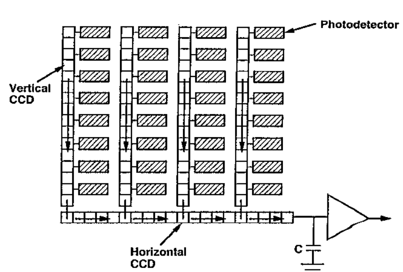


Figure 3: Interline transfer CCD.

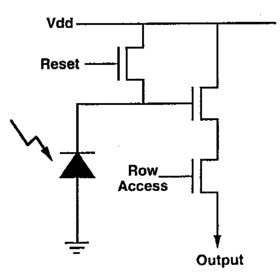


Figure 5: Active pixel circuit.

pipeline. This simplifies introduction of alternate forms of pixel access, such as that required for electronic pan & zoom.

Arguably the most important advantage of the CMOS APS approach, however, is the ability to integrate much of the camera timing, control and signal processing circuitry onto the same silicon die. A CIF (352x288) sensor array can be built in 0.5µm CMOS in under $11mm^2$ active area. Even assuming a $30mm^2$ die, this leaves area for these system functions, plus extra circuitry to customize the camera for a particular application.

One problem associated with APS arrays is the existence of significant levels of fixed-pattern noise. Unlike CCDs, in which all charge packets travel essentially the same signal path and use the same output amplifier, each pixel in an APS array has its own amplifier. Gain and offset variations between these amplifiers lead to a static pattern noise which appears as a background texture on the image. The eye is very sensitive to small amounts of pattern noise (1-2% is clearly visible), particularly if it is aligned along vertical or horizontal lines, as is the case with noise contributed by column output stages.

Fortunately, the ability to integrate signal processing electronics onto the same die provides a number of solutions to this problem. Fixed-pattern noise can be significantly reduced by the use of a simple correlated double sampling (CDS) technique as shown in Figure 6. Each column output stage contains two sample and hold capacitors. One is used to sample the pixel reset level, one is used to sample the signal level (after integration). By subtracting the reset level from the signal level, much of the pixel amplifier offset is removed. A second level of CDS can be applied to eliminate offset in the column amplifier. A combination of these two techniques can reduce fixed pattern noise from 5% to 0.1%. Or digital techniques can be used to reduce the level of pattern noise. The offset level for each column, for example, could be stored in a RAM and digitally subtracted from the output signal column by column. These techniques can reduce the pattern noise to a level imperceptible under the room-light (or brighter) levels of illumination found in most multimedia applications.

A Single-Chip Multimedia Camera

The basic architecture of a camera on a chip for multimedia applications is shown in Figure 7. The detailed functionality of each module depends on the nature of the application. Of particular significance is the degree of autonomy required of the camera. At one extreme, represented by a conventional video camera, the camera operates in a stand-alone mode, calculating exposure times and color balance and producing a stream of video information at a predetermined frame rate. A more flexible model, however, is one in which camera functionality is partitioned between the camera hardware (simple, per-pixel operations) and software (complex, per-frame operations) in an intelligent host (e.g., PC), thereby reducing the cost of the camera hardware and increasing the functionality of the overall system. The camera is no longer autonomous, instead having a symbiotic relationship with the host and the host application as shown in Figure 7. Exposure control may be partitioned between the camera that maintains summary statistics on intensity derived from examining each pixel and the host that uses the summary statistics to calculate exposure time that is passed to the camera.

The sensor array is comprised of a two-dimensional pixel array that can be randomly addressed through adjacent row and column decoders. An emerging standard for video-telephony is CIF at 288 by 352 pixels, somewhat less than NTSC resolution, but sufficient for many compressed video and snap-shot applications.

The standard video frame rate is 30 frames per second. Multimedia applications, however, typically operate at 10 - 15 frames per second. These numbers suggest per frame exposure times of approximately 60 to 100ms. These are however maximum exposure times that are only used in moderate to low-light conditions. As lighting intensifies, it is necessary to reduce the effective exposure time to ensure the sensors do not saturate (exceed their maximum charge capacity). Long exposure times may also lead to high levels of dark current pattern noise. The exposure control block generates timing signals to set the interval between resetting a row of pixels and reading the accumulated charge. Exposure time is calculated to maintain a specified distribution of pixel brightness levels. This calculation can be performed either by on-board circuitry, or software on the host.

A specific region of the sensor may be read simply by presetting the counters that drive the row and column decoders to values that represent the origin of the region of interest at the start of every new frame. Electronic panning may then be implemented by altering the preset values to the new window origin under user (or application) control.

Traditionally, digital cameras have used a single A/D converter operating at the pixel rate (3Mpixels/s for CIF, 30frames/s). Eight-bit resolution is sufficient provided that AGC and gamma correction have already been performed in the analog domain. Having the converter on the same die as the sensor, however, allows alternative A/D solutions. A CMOS camera in which a ΣΔ A/D converter has been effectively integrated into each pixel, allowing digital readout at the cell level [9]. The sensor capacitance conveniently performs summing and integrating. This simplifies signal read-out but leads to larger pixel sizes and reduced sensitivity since the raw sensor current is now the A/D input.

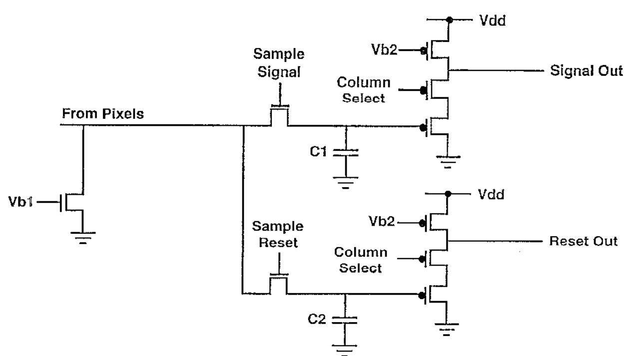


Figure 6: CDS fixed pattern noise reduction circuit.

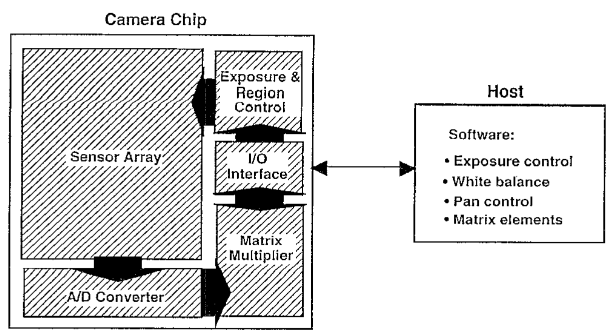


Figure 7: Camera on a chip hardware/software partitioning.

Alternatively, a simple low-speed single-slope converter may be placed at the output of every column as shown in Figure 8. A comparator compares the column output against a reference. The reference voltage is a ramp derived from a counter feeding a D/A converter. When the column comparator senses equality between the column signal and the reference, the counter output is recorded in a register associated with each column. At the end of a single conversion cycle (one row of video) the registers contain digital values representing the analog output of each column, and may be sequentially (or randomly) addressed to generate a digital output signal from the chip. Additional signal processing (AGC, gamma correction) may be performed by modifying the slope and the shape of the reference ramp.

Additional signal processing is usually required to compensate for variations in processing and operating conditions. These may be performed in either the analog or the digital domain. Automatic gain control (AGC) is required in moderate to low-light conditions where the camera is running at its maximum exposure time. Gain may be added into the signal path to increase the apparent brightness of the scene. This has the effect, however, of also amplifying noise, so the total gain will be limited by the signal-to-noise ratio of the array. Gamma correction is required to compensate for display non-linear response. One implementation simply uses the digital value of the video signal to address a lookup table containing the corrected signal values.

It is usually necessary to color correct (white balance) to compensate for non-ideal response in the sensors and color filters. This may take the form of a simple gain control on each component channel. Digital processing, however, provides a more flexible solution in the form of a color transformation matrix:

$$\begin{bmatrix} R' \\ G' \\ B' \end{bmatrix} = \begin{bmatrix} X_{1,1} & X_{1,2} & X_{1,3} \\ X_{2,1} & X_{2,2} & X_{2,3} \\ X_{3,1} & X_{3,2} & X_{3,3} \end{bmatrix} \begin{bmatrix} R \\ G \\ B \end{bmatrix}$$

This transformation is applied to each R/G/B sample to allow for crosstalk and response variation between the three color channels. It can be used to correct significant errors in sensor spectral and color filter responses and therefore simplifies color filter array specification. In addition, such a matrix can generate alternative color space formats (e.g., YCrCb, YUV).

To maximize the value of integration it is necessary to select an appropriate digital interface and implement it on-chip. Clearly the interface is highly application dependent, with choices from a simple proprietary unidirectional data stream for connection directly to a video codec, to more complex bi-directional standard such as IEEE 1394 ("Firewire") for connection to a PC.

Back-End Manufacturing Steps

Although the concept of a single chip camera is based on fabricating the entire device on a single die in a standard CMOS process, a number of additional or back-end steps must be included to complete the overall manufacturing process.

Chip packaging costs are a significant portion of the complete costs of a sensor. A sensor package must: (1) have a transparent, hermetically sealed lid, (2) be able to adequately dissipate sensor power, (3) be manufactured (including chip placement) to tolerances sufficient for inclusion in an optical system.

The relatively high power dissipated by CCD sensors and their sensitivity to thermal stress often requires use of costly ceramic or precision plastic packages. The much lower power requirements of CMOS arrays permits the use of conventional plastic packages, considerably lowering the cost of the completed part.

Testing optical devices requires more complex facilities than those used for standard digital parts. Additional test issues include: (1) providing a controlled optical source for photodetector stimulation, (2) digital tester inputs capable of verifying that a good output lies in a range of values rather than a single value.

To construct a single-sensor color camera it is necessary to pattern the sensor surface with a suitable mosaic of color filters. There are two color systems available: additive (red, green, blue) and subtractive (magenta, cyan, yellow). Subtractive filters have the advantage that they let more light onto the sensor and provide greater sensitivity. Additive filters lead to simpler color processing. Filters are typically made from dyed polyamide, each color being lithographically defined and etched in turn to create an individual filter square over each pixel. Various patterns, such as the one shown in Figure 9, have been proposed. Typically patterns are chosen to emphasize the luminance (or green) resolution while chrominance (or red/blue) resolution is sampled at lower resolution. Depending on the pattern chosen, various amounts of buffering (typically one to two lines) and processing (addition/subtraction) may be needed to derive the required per-pixel output data from the raw array data.

Because the fill factor of a pixel is typically around 30%, considerable gains in sensitivity can be achieved by constructing a lens over the entire pixel to focus light onto the active area. These microlenses may be fabricated by patterning and etching polyamide to form a cylinder over each pixel, then heating the material until it flows to form a spherical lens over the pixel.

In addition to the single silicon die, a complete camera will require a simple PC board (perhaps including a voltage regulator), a plastic housing, lens, cabling and connector. Recent developments in the production of precision injection molded plastic asphere lenses suggest that a high-quality multi-element lens can be produced at a fraction of the cost of glass spherical lenses. For simple fixed-focus applications these may be attached directly to the camera chip package.

Evolution of Single-Chip Cameras

The concentration of work on single-chip cameras is presently aimed at the production of low-cost color video cameras for multimedia applications. As the CMOS sensor array technology evolves, and process line widths continue to decrease, we expect to see: (1) CMOS image sensor arrays will become standard cells much like any other chip layout component. For example an image sensor and a video encoder may be combined with some ASIC glue to create an application-specific product. (2) more sophisticated use of application/camera interaction resulting in increased image quality and production values (e.g., automatic head tracking supported by electronic pan and zoom). (3) higher resolution sensors for document image capture applications such as fax. (4) high-resolution sensors for all-electronic consumer still camera applications. (5) low-cost, low-resolution sensors that allow intelligent machine vision functions to be added to consumer items such as automobiles and home appliances.

The product life-cycle of the digital clock may provide some indication of the future of cameras. Initially a relatively costly stand-alone device, digital clocks became cheap enough to be combined with radios, and then eventually became standard features of a wide range of products from microwave ovens to VCRs. Similarly, the high levels of integration enabled by the development of CMOS image sensor technology will drive consumer electronic cameras from their present stand-alone (camcorder) form to being a ubiquitous feature of everyday life.

Figures 8 and 9 and *References*: See page 412.

(Continued from page 25)

References:

[1] Fujikawa, F., et al., "A 1/3-inch 630k-pixel IT-CCD Image Sensor with Multi-Function Capability," IEEE ISSCC Digest of Technical Papers, pp. 218-219, Feb., 1995.

[2] Bosiers, J., et al., "Design Options for 1/4"-FT-CCD Pixels," Proc. IEEE Workshop on Charge-Coupled Devices and Advanced Image Sensors, April, 1995.

[3] Wang, S., et al., "A Real-Time Signal Processor for use with the Interline Transfer Color CCD Imager," Proc. IEEE Workshop on Charge-Coupled Devices and Advanced Image Sensors, April, 1995.

[4] Guidash, R., et al., "A Modular, High Performance, 2µm CCD-BiC-MOS Process Technology for Application Specific Image Sensors and Image Sensor Systems on a Chip," Proc. of IEEE Intl. ASIC Conf. & Exhibit, pp. 352-355, April, 1994.

[5] Renshaw, D., et al., "ASIC Vision," Proc. IEEE Custom Integrated Circuits Conf., pp. 7.3.1-7.3.4, 1990.

[6] Fossum, E., "Active Pixel Image Sensors - Are CCD's Dinosaurs?," Proc. SPIE, vol. 1900, pp. 2-14, 1993.

[7] Jansson, C., et al., "An Addressable 256x256 Photodiode Image Sensor Array with an 8-bit Digital Output," Analog Integrated Circuits and Signal Processing, pp. 37-49, Vol. 4, 1993.

[8] Dickinson, A., et al., "Standard CMOS Active Pixel Image Sensors for Multimedia Applications," Proc. Conf. on Advanced Research in VLSI, pp. 214-224, March, 1995.

[9] Fowler, B., et al., "A CMOS Image Sensor with Pixel Level A/D Conversion," ISSCC Digest of Technical Papers, pp. 216-217, Feb., 1994.

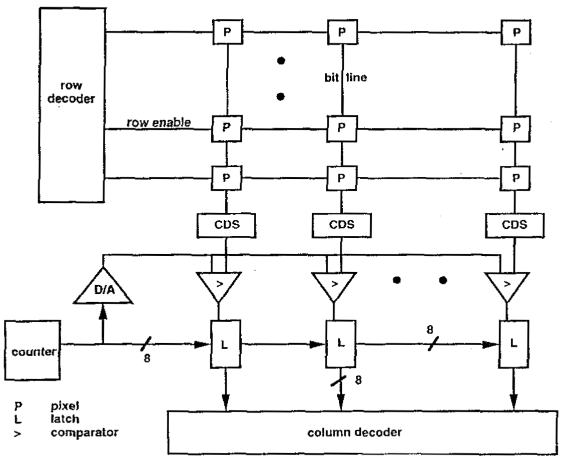


Figure 8: Per-column, single-slope A/D converter.

R	G	R	G	R	G
G	B	G	B	G	B
R	G	R	G	R	G
G	B	G	B	G	B
R	G	R	G	R	G
G	B	G	B	G	B

Figure 9: Color mosaic pattern.

TP 2.1: A 3.3V 128Mb Multi-Level NAND Flash Memory for Mass Storage Applications
(Continued from page 33)

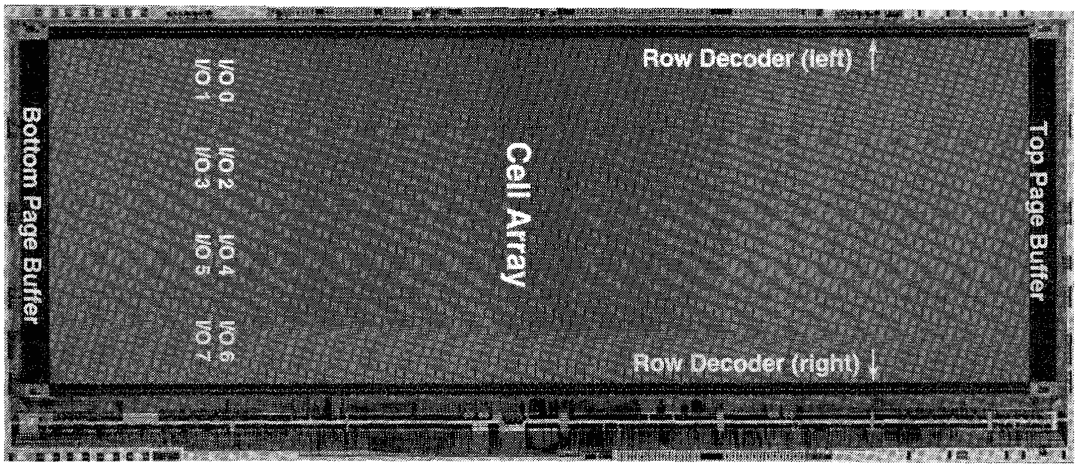


Figure 6: 128Mb NAND flash memory chip micrograph.