

High-Speed Global Shutter CMOS Machine Vision Sensor with High Dynamic Range Image Acquisition and Embedded Intelligence

Francisco Jiménez-Garrido¹, José Fernández-Pérez¹, Cayetana Utrera¹, José Ma. Muñoz¹, Ma. Dolores Pardo¹, Alexander Giuliatti¹, Rafael Domínguez-Castro¹, Fernando Medeiro¹, and Angel Rodríguez-Vázquez^{2,1}

¹ AnaFocus (Innovaciones Microelectrónicas S.L.)

² IMSE-CNM/CSIC and Universidad de Sevilla

Avda Isaac Newton, Pabellón de Italia, Ático

Parque Tecnológico Isla de la Cartuja, 41092-Sevilla (SPAIN)

angel@imse-cnm.csic.es

angel.rodriguez-vazquez@anafocus.com

ABSTRACT

High-speed imagers are required for industrial applications, traffic monitoring, robotics and unmanned vehicles, movie-making, etc. Many of these applications call also for large spatial resolution, high sensitivity and the ability to detect images with large intra-frame dynamic range. This paper reports a CIS intelligent digital image sensor with 5.2Mpixels which delivers 12-bit fully-corrected images at 250Fps. The new sensor embeds on-chip digital processing circuitry for a large variety of functions including: windowing; pixel binning; sub-sampling; combined windowing-binning-sub-sampling modes; fixed-pattern noise correction; fine gain and offset control; color processing, etc. These and other CIS functions are programmable through a simple four-wire serial port interface.

Keywords: CMOS High-Speed Digital Image Sensors, Smart Image Sensors

1. INTRODUCTION

High-speed imagers are required for industrial applications, traffic monitoring, robotics and unmanned vehicles, movie-making, etc. Many of these applications call also for large spatial resolution, high sensitivity and the ability to detect images with large intra-frame dynamic range. These motivations have prompted proposals of different CMOS architectures and circuits for high speed downloading of images [1]. Also, new sensor devices employing CIS technologies and pinned photodiodes for large image quality are being devised [2], [3]. The sensor reported in this paper combines large resolution, high-speed and large image quality with large flexibility and programmability, on the one hand, and the ability of acquiring images with high intra-frame dynamic range, on the other. The sensor is conceived to deliver fully-corrected digital images, thus largely reducing the demands on the off-chip correction circuitry when the sensor is incorporated into a camera system.

The sensor employs 5-T pixels with 5 μ m pitch and delivers up to 12-bit fully-corrected digital images according to the QSXGA standard (2560 x 2048 pixels). Downloading speed is 250Fps at full resolution and increases to 1,845Fps at VGA resolution. The sensor can be digitally configured for either linear response with DR=60dB, or for high dynamic range response with intra-frame DR of 100dB. It incorporates on-chip digital circuitry for functions such as: windowing; pixel binning; sub-sampling; combined windowing-binning-sub-sampling modes; fixed-pattern noise correction; fine gain and offset control; and color processing. These and other CIS functions are programmable through a simple four-wire Serial Port Interface (SPI). It also includes 24 LVDS high-speed outputs allowing transferring 12-, 10-, or 8-bit image

data up to 16.6Gbit/sec, and two additional LVDS channels: one for clock recovery and one more for synchronization purpose. LVDS ports can be disabled when either the frame rate or the output word-length (selectable among 12, 10, and 8bit) is reduced, thus minimizing the complexity of the external components required. All the required timing and reference voltages are internally generated, thus minimizing the need for external components. It includes a power down capability for very low power dissipation.

2. SENSOR ARCHITECTURE

Figure 1 shows the block diagram of the sensor, designed in a 0.18 μ m CIS technology (TowerJazz). It comprises five main sections: i) the pixel-array; ii) the read-out and conversion channel; iii) the digital circuitry; iv) the communication interface; and v) the auxiliary blocks.

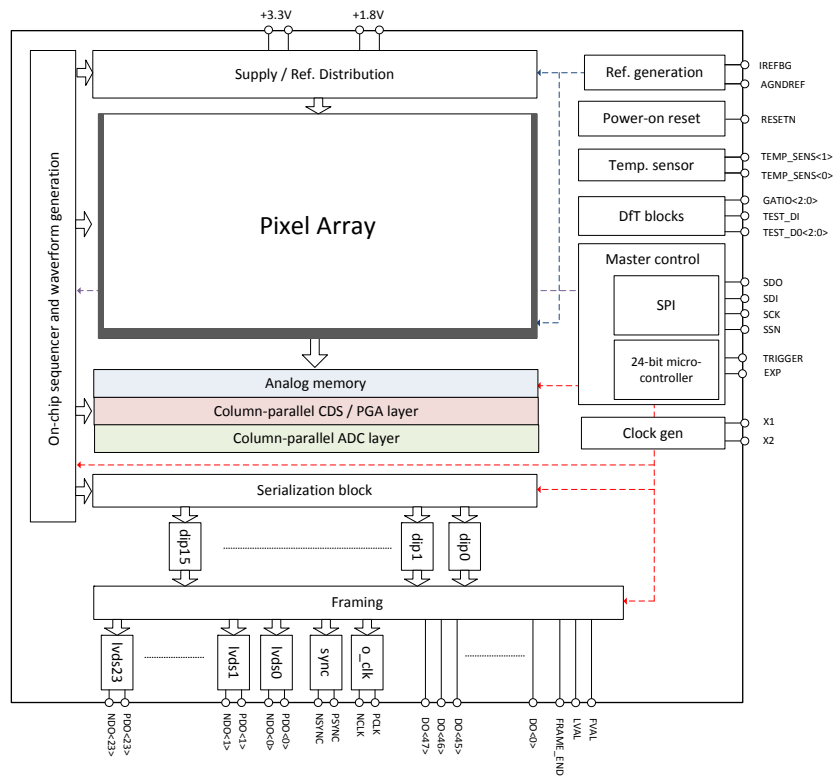


Figure 1. Block diagram of the sensor

The *Pixel-Array* consists of 2,640 x 2,073 pixels out of which 2,560 x 2,048 are actually used for image capture (QSXGA standard) while the remaining ones are used for fixed pattern noise sensing (horizontal and vertical), calibration and testing purposes. Some dummy extra rows are added to preclude photo-generated electrons reaching the active areas of the pixels employed for error sensing.

The *Readout Block* follows a column-parallel approach. It contains 2,560 readout elements (1 per pixel column) that operate in parallel, conditioning and digitizing the analog data of one entire row of pixels simultaneously. The control unit generates all the necessary signals to control the pixel-array and the readout circuitry.

The *On-Chip Digital Circuitry* consists of the following components: i) The master control, responsible for the overall control of the image sensor including the waveform generations for pixel array, the readout channel and the communication interface. The configuration of the sensor control can be performed with either an external host device via SPI-commands or by software running on the on-chip-embedded 16-bit microcontroller; ii) The serialization block; iii) The digital image processors consisting of a collection of interconnected image-processing engines able to perform a

variety of image-correction algorithms over the digitized pixel data stream; iv) The communication interface, comprising all blocks related to input/output signals, such as SPI, external control signals and data output.

The *Auxiliary Blocks* include circuits intended for: i) Power-on-reset; ii) Clock generation through three low-jitter low-power phase locked loops; iii) Reference generation with a high-accuracy band-gap; iv) Temperature sensing,; etc.

3. PIXEL ARRAY AND PHOTODIODE

Along with the 2,560 x 2,048 image sensing pixels, the array (Figure 2) includes 64 Optical-Black (OB) columns to compensate the effect of the systematic component of the dark-current and for horizontal line noise reduction. This group of 64 OB columns is surrounded by 4 (left) + 4 (right) OB dummy columns. The purpose of these OB dummy columns is to absorb any photo-generated charge flowing from the adjacent active pixels that otherwise could reach the OB pixels. The array also includes 8 OB rows that can be used for calibration of vertical fixed-pattern noise (VFPN). Like for OB columns, these OB rows are surrounded by 4 (top) + 4 (bottom) OB dummy rows. Finally, one row of test pixels is added for fine calibration processes and characterization of the readout channels. Test pixels are similar to conventional pixels but without a photodiode. The output voltage of test pixels is digitally controlled with internal D/A converters. These pixels are specifically included in order to enable a fine characterization of the DC and AC characteristics of the readout channel, and for calibration purposes. Surrounding the selectable active pixels, there is a layer of dummy pixels to avoid boundary effects.

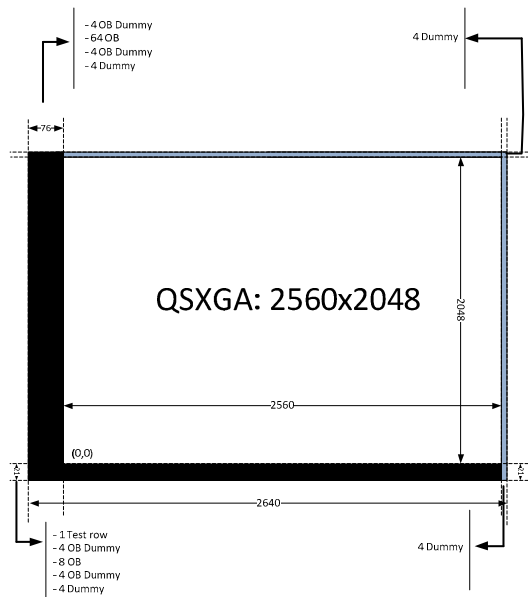


Figure 2. Pixel Array

Within the sensing area formed by the inner 2560 x 2048 rows, the selection granularity for rows and columns changes, depending on the binning/sub-sampling configuration. When neither binning nor sub-sampling is applied, row selection granularity is 2 while column selection granularity is 64. This means that the columns can only be selected in groups of 64, beginning with column #0, or #64, or #128, etc.

A maximum frame rate of 250fps is achieved when the selected number of columns is 2,560 and the selected number of rows is 2,048. Lowering the number of selected columns and rows (as a result of windowing, binning, or sub-sampling) yields a frame rate increase; namely, assuming 12-bit operation and downloading through the 24 LVDS outputs: i) 250fps @ 2,560 x 2,048; ii) 980fps @ 1,280 x 1,024; iii) 1845fps @ 640 x 480.

4. READOUT CHANNEL

Figure 3 shows the readout channel of the sensor – responsible for the acquisition, analog conditioning, and digitization of the image data from the pixel array. The resulting digital words are passed to the serialization block to be distributed to the embedded processors. Reading follows a column-parallel approach with 1 readout channel per sensor column. Each individual channel (dashed red line in Figure 3) consists of: i) 4 analog memories separated in two sets (with 2 analog memories each) for pipelining input data; ii) 1 double sampling circuit for Correlated Double Sampling (CDS); iii) 1 Programmable Gain Amplifier (PGA), and iv) 1 Analog-to-Digital Converter (ADC).

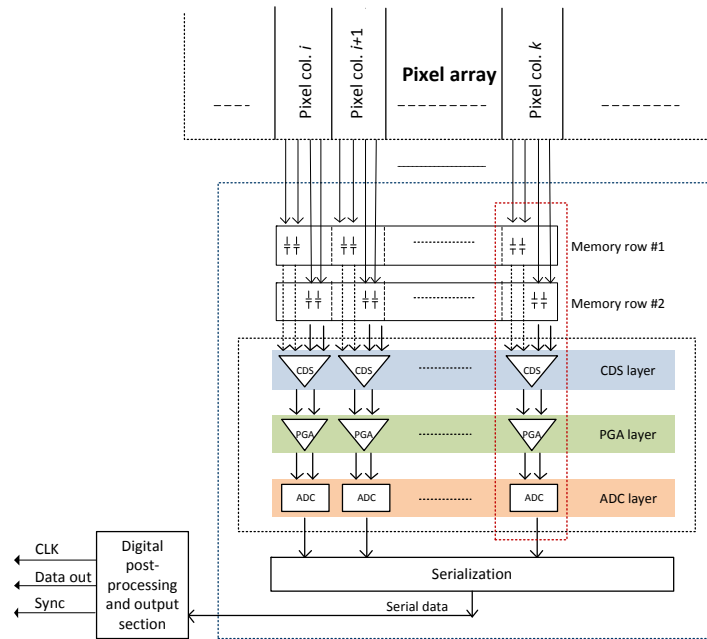


Figure 3. Functional diagram for the Readout Path

Figure 4 shows the data flow through the analog and digital paths. Note that the read-out channel incorporates also a Digital-to-Analog Converter (DAC) for addition/subtraction of an offset to/from the readout channel input.

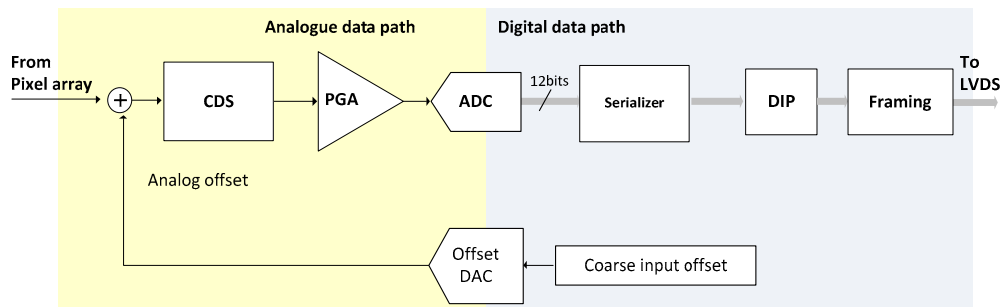


Figure 4. – Simplified functional diagram of the readout channel

At the analog side, the purpose of this offset is to make the input signal fit into the range of the ADC. However, it is not a fine offset adjustment, but a coarse one, as necessary to avoid loss of codes (due to having negative values at the input of the ADC) or significant loss of DR (in case a large dark signal from pixel exhausts a significant portion of the ADC full-scale range). Once the coarse analog input offset has been added, the signal is passed to the CDS block where the

difference between the pixel reset and signal level is computed, and then to the PGA where it is amplified before converting it to digital at the ADC. The digitized data is serialized out to a set of Digital Image Processors (DIPs) where additional arithmetic operations, as well as data correction are carried out. The digitally processed data are passed to a Framing Block and outputted through either LVDS or CMOS ports.

Figure 5 is an example of a timing diagram showing the exposure and the row-by-row readout sequence. Each time-slot corresponds to a row readout and conversion time. Readout and conversion time is the time taken by each readout channel to sample, amplify, and digitize the pixel signal. Data in line memory stays for two row times (row time is defined as the time required for a row of pixels readout) thus indicating that during one conversion time it is read from the pixel and, during the following one, data is fed to the rest of the signal path. This dual analog memory (or “ping-pong”) structure in the readout channel permits increasing the available time to read out a row; this significantly decreases the speed and the power consumption requirement of the readout channels.

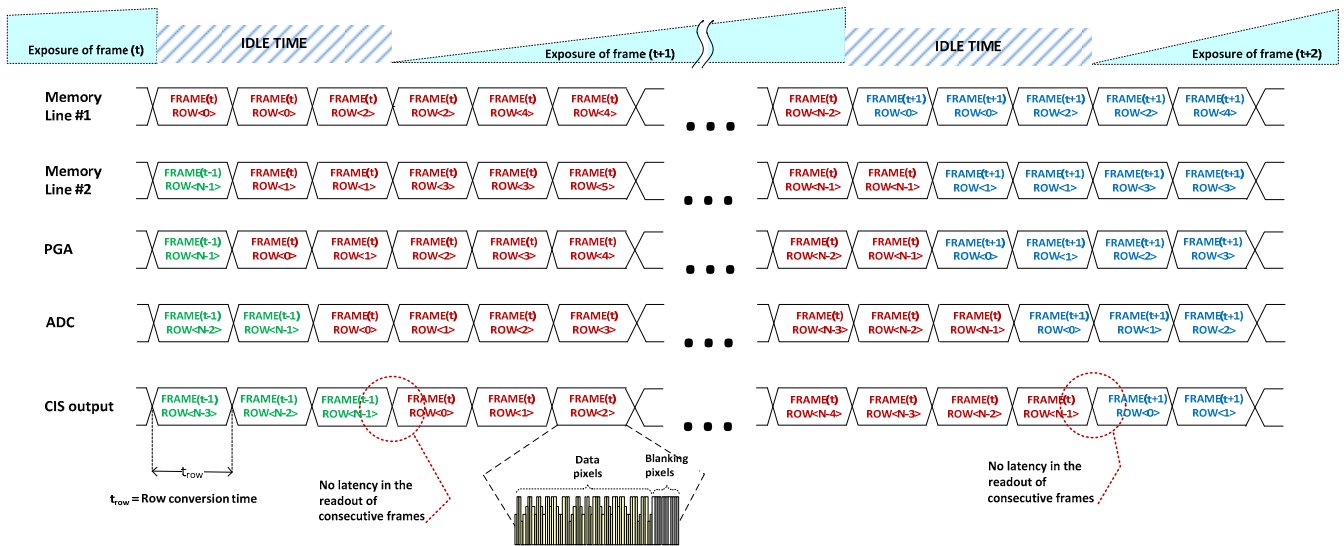


Figure 5. Readout Channel Timing Diagram

5. PIXEL

Figure 6 shows a simplified schematic of the 5T pixel. It comprises of photodiode, the MOS transistors and a Floating Diffusion (FD) node. The pixel is driven by the control signals AB, TRF, and RST. Once the photo-generated charge is converted to voltage at the FD capacitance, it is readout via the in-pixel source-follower and the selection switch (SEL) that connects it to the data column and current source.

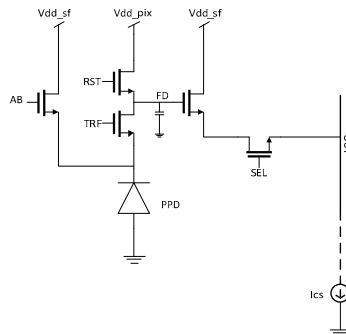


Figure 6. Simplified 5-T pixel schematic

The waveforms applied to the control signals above are generated via a flexible state machine that allows the user to program a large number of combinations involving the control sequence as well as the duration of each pulse. As illustrated in Figure 7, the exposure (or procedure for light sensing and photo-generated charge transfer and storage at the FD node), requires passing through a number of states that can be grouped as follows: i) State B, required to prepare for exposure by performing a global reset; ii) State E, light sensing begins with falling edge of AB; iii) State F, required to finishing exposure, transfer photo-generated charges to FD node (which requires an activation of the RST signal for removing the data previously stored at FD), activation of the TRF signal to transfer the charge from the photodiode to the FD node. The duration of the complete exposure is the sum of the duration of states B, E, and F, that can be independently programmed by the user.

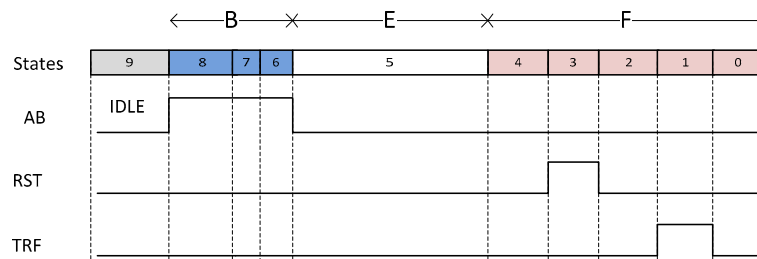


Figure 7. Example of pixel control waveforms

Once the photo-charge has been transferred to the FD, at the end of State F, it remains there till the readout of the pixel is activated. The way the exposure and readout are carried out depends, among others, on the selected exposure-readout sequence. However, as a general remark, it should be noticed that no readout is allowed for a pixel during State F, because at that time the FD of the pixel is being re-written with the new image data coming from the PPD.

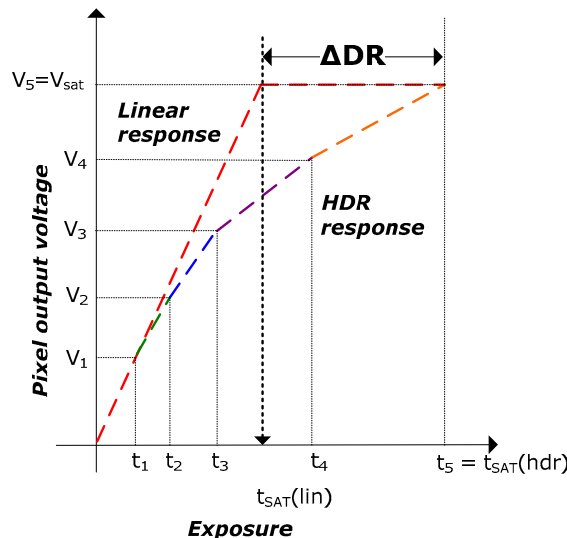


Figure 8. HDR mode integration characteristic

Sensing can be made either in linear mode or in HDR mode. In this latter mode, the image sensing process is controlled to achieve a piece-wise linear compression characteristic in the pixels response as illustrated in Figure 8. The result of such a compression characteristic is an increase in the effective intra-frame dynamic range as compared to linear pixel response. The enhancement of DR depends on the number of intermediate levels (also known as “knee points”) defined for the pixels response, and the time interval for which each particular level is maintained.

The increment of intra-frame dynamic range that we could expect from the HDR mode compared to linear integration is given by the equation:

$$(1) \quad \Delta DR = \frac{E_{MAX}(hdr)}{E_{MAX}(lin)} = \frac{k^N - 1}{N \cdot (k - 1)}$$

where $k (>1)$ is a constant determining the relation between the pixel exposure time (t_{exp}) and intermediate time instants (t_i) in which the pixel integration characteristic is varied, and N is the number of intermediate levels. For $k=4$ the enhancement of DR amounts to 37dB.

6. ON CHIP INTELLIGENCE

Figure 9 highlights the contents of the DIP within the analog and digital data paths. The DIPS perform the image correction and processing functions embedded on-chip.

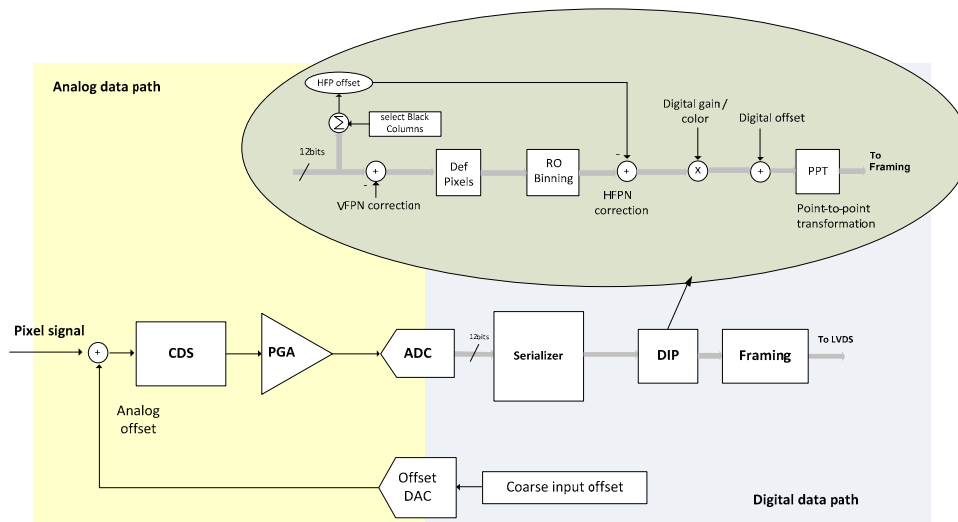


Figure 9. Data path scheme showing the location and contents of the DIP

On-chip image processing functions

Analog offset and analog gain. As already mentioned, the sensor allows addition/subtraction of a coarse analog offset to/from the pixel output. The analog gain applied after CDS is selectable among $\times 1$, $\times 2$, $\times 4$, $\times 8$, and $\times 16$. In fact, a single readout channel can work with two different analog gains so that these are exchanged in alternating pixel rows. This makes it possible to configure analog gain per color in devices with Color Filter Array (CFA). Changing analog gain implies modification of the waveforms controlling the readout blocks. Thus, there is no register for directly specifying the required analog gain, but a set of registers configuring such waveforms.

Digital offset and digital gain. After digitization, global digital offset and gain can also be adjusted through registers programmed by SPI. For color sensing, global gain can be defined separately for green, red, and blue pixels, so that it can be used to perform on-chip white balance. For monochrome devices, the same gain is applied to all the pixels. Global offset has a range of ± 2048 . Global gain can be expressed as:

$$(2) \quad G_{GLOB} = 2^{G_{COARSE}} \times G_{FINE}$$

where $G_{COARSE} = 0, 1, 2 \dots 7$ and G_{FINE} goes from 0 to 2.

Auto-black level control. Black level can be self-adjusted on-chip by using the OB pixels present in all sensor rows. The black level of a row is measured and subtracted from all pixels in that row before data is outputted.

Point-to-point transformation The Point-to-Point Transformation (PPT) block allows the user to program a pixel value-based transformation law. The sensor DIP implement a piece-wise linear approximation of the desired PPT, based on a Look-Up-Table (LUT). The on-chip PPT LUT can store 33 points defining 32 intervals in the overall response curve. The PPT procedure implements the following interpolation:

$$(3) \quad y = \frac{Y_{j+1} - Y_j}{X_{j+1} - X_j} \cdot (x - X_j) + Y_j$$

where x represents the incoming pixel code and X_j, Y_j are the coordinates of the interval containing such a code. X_j values are prefixed to 0, 128, 256, ... 4096; whereas Y_j 's can be programmed by the user, except the last one which is pre-fixed to $Y_{33}=4096$. This enables implementation of any non-linear characteristic, such as the ones required for gamma-correction or histogram equalization.

For color devices, the 32-point programmable values can be split into two sets with 16-point each, so that 2 different LUTs can be programmed. Since on-chip digital processing is implemented so that each DIP processes pixels of two colors only, this is useful to perform PPT per color. Figure 10 illustrates this approach.

Readout channel binning. When enabled via configuration register, CMOS pixel binning with enhanced image quality is performed. It is accomplished through averaging of the response of the binned readout channels. This function is also performed at the DIP level.

On-chip image corrections

Vertical FPN (VFPN). The readout path includes a digital calibration mechanism to remove column-to-column VFPN associated to the use of multiple readout channels. The DIP incorporates a memory to store correction data for each readout column. In order to perform column-to-column VFPN correction, it is necessary to update one correction register per internal readout channel. The calibration can be done using either the test rows or the OB rows of the pixel array. Both options are available to calculate the zero signal image level. The process of VFPN calibration is not performed during normal acquisition. This is a calibration step and must be performed during the camera startup phase and occasionally in the field, especially if operational conditions change significantly. Acquiring calibration data is performed several times in each calibration cycle. This is done by reading several times the test row or by reading several OB lines. Multiple readings reduce the effect of the temporal noise in the signal path (by averaging). If the selected VFPN calibration mode is external, data will be outputted through the LVDSs like a regular image. In this case, the user must process this data externally to compute the correction coefficients and put them in at the internal memories via SPI.

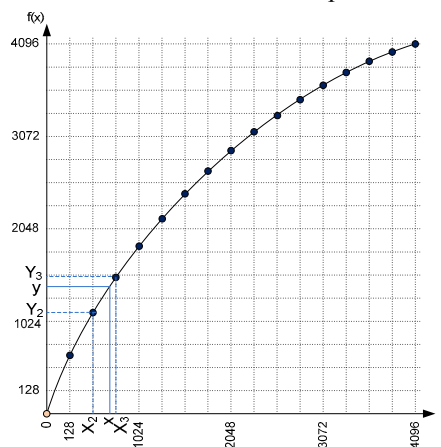


Figure 10. Point-to-point transformation by LUT interpolation

Horizontal FPN (HFPN). The sensor can calibrate and correct row-by-row (horizontal) fixed-pattern noise (HFPN). Such noise is normally due to a random, row-to-row variation in the readout channel response, which generates a horizontal pattern superimposed to the signal. In comparison to VFPN, this error varies randomly from frame to frame as it is not associated to the readout channel mismatch, but to random temporal noise (from supplies, reference voltages, and bias currents) that is “frozen” (sampled and held) when a pixel row is read out. Consequently, the error is not the same for all pixels in a row, but changes from row to row and from frame to frame. In order to remove the HFPN, the sensor uses the information from OB columns. When a row is read out, the first pixels that are digitized correspond to the OB columns. Via a configuration register, some of these OB columns can be selected to enter an averaging block to compute a per-row value of the error. It is important to note that since the sampled OB columns belong to the same row, the row-to-row HFPN sensed will be the same as in the remaining (active) pixels in the row. Such correlation allows cancelling this noise out by simply subtracting the averaged value from the row regular pixels. Note that this function inherently provides auto-black level control. In fact, row-by-row black level compensation is more effective than performing a global calibration because with the latter large dark currents (when the temperature is high) may produce wrong black level correction for pixels that are processed at the beginning or at the end of the readout time.

Defective pixels correction. The sensor can correct on-chip up to 512 defective pixels. A dedicated internal memory stores the position of defective pixels. Changes in sub-sampling or binning will require a reconfiguration of the defective pixels memory. Defective pixels in monochrome and color devices are handled differently. A defective pixel is tagged and replaced by one of its closest active neighbors in the same row and within the same group of 16 pixels. Figure 11 and Figure 12 show examples of defective pixel replacements. Note that the neighboring location changes when either sub-sampling or binning are applied or the device is a color sensor.

7. CONCLUSIONS

CMOS imagers currently dominate the market of area imagers, with more than 90% share. Although consumer applications clearly dominate the arena, the volume for other applications (such as machine vision, surveillance, military applications, X-ray imagers, medical, etc.) is forecasted to reach some 0.7billion units in 2015. For many of these applications image resolution must be complemented with other features such as speed and smartness. For instance, sensors intended for surveillance applications should be capable to analyze complex spatial-temporal scenes and combine high-quality image recording of significant events with high-speed decision making. Just to mention another example, scientific applications call for the smart selection of salient points and region-of-interests and for the ultra-high-speed downloading of the so selected areas. Also, machine vision sensors require image content analysis and decision making to be made with largest possible throughput. All these features require the incorporation of processing circuitry together with the photo-sensing and readout circuitry. The sensor reported in this paper features high image quality, high-speed and large embedded intelligence and is very well suited for industrial vision applications.

8. ACKNOWLEDGEMENTS

Partially funded by the spanish Project INNPACTO IPT-2011-1625-430000.

9. REFERENCES

- [1] Steven Huang et al., "Design of a PTC-Inspired Segmented ADC for High-Speed Column-Parallel CMOS Image Sensor". 2011 IISW, pp. 328-331, June 2011.
- [2] X. Wang et al., "A 2.2M CMOS Image Sensor for High-Speed Machine Vision Applications". Proc. SPIE, vol. 7536, January 2010.
- [3] Jan Bogaerts et al., "High Speed 36Gbps 12Mpixel Global Pipelined Shutter CMOS Image Sensor with CDS". 2011 IISW, pp. 335-338, June 2011.
- [4] A. Rodríguez-Vázquez, F. Medeiro and E. Janssens, “*CMOS Telecom Data Converters*”. ISBN 978-1-4020-7546-9, Kluwer Academic Publishers 2003 .

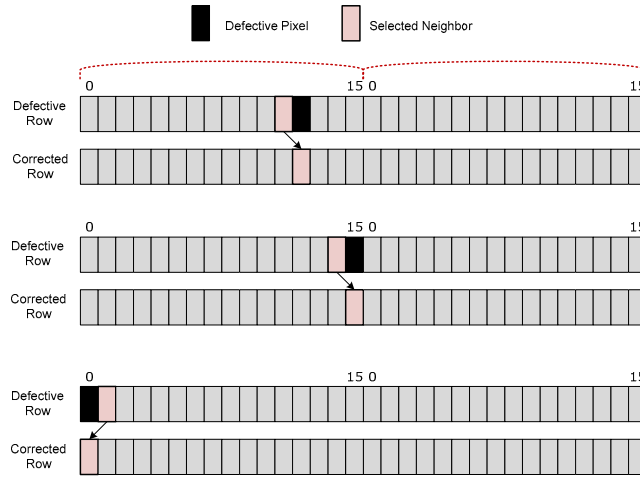


Figure 11. Defective pixel correction by closest neighbor replacement

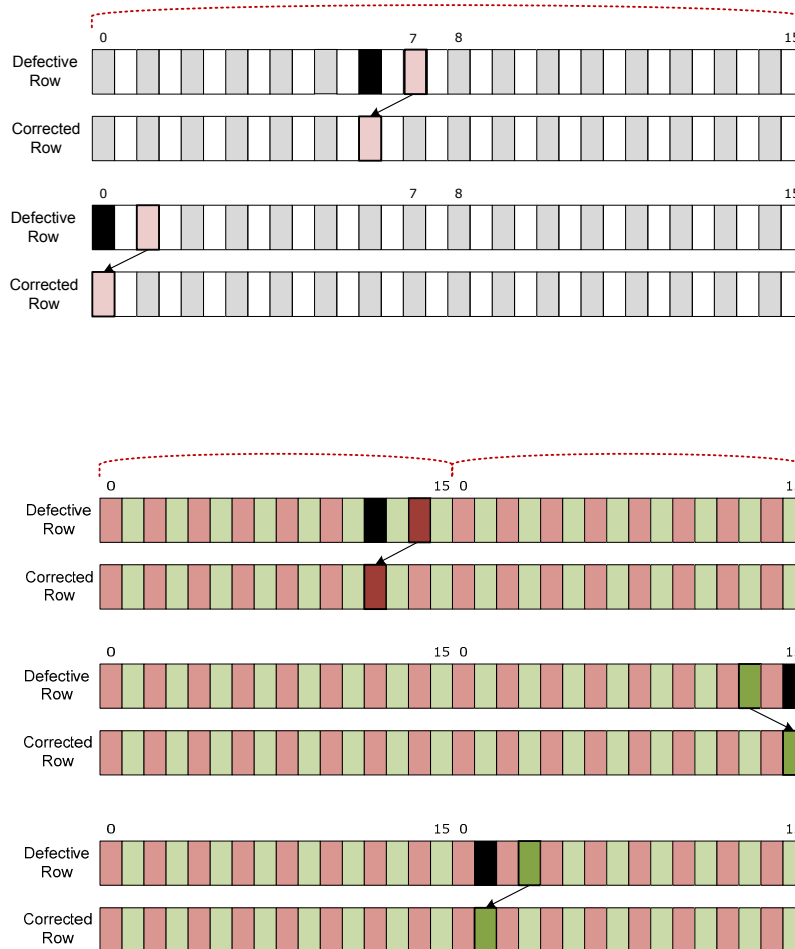


Figure 12. Defective pixel correction in monochrome 2x2 sub-sampling case (top) and in colour (bottom)