A 151dB High Dynamic Range CMOS Image Sensor Chip Architecture with Tone Mapping Compression Embedded in-Pixel

Sonia Vargas-Sierra, *Member, IEEE*, Gustavo Liñán-Cembrano and
Ángel Rodríguez-Vázquez, *Fellow, IEEE*

***Abstract*—This paper presents a high dynamic range CMOS image sensor that implements an in-pixel content-aware adaptive global tone mapping algorithm during image capture operation. The histogram of the previous frame of an auxiliary image, which contains time stamp information, is employed as an estimation of the probability of illuminations impinging pixels at the present frame. The compression function of illuminations, namely Tone Mapping Curve (TMC), is calculated using this histogram. A QCIF resolution proof-of-concept prototype has been fabricated using a 0.35µm opto-flavored standard technology. The sensor is capable of mapping scenes with a maximum intra-frame dynamic range of 151dB (25-bits/pixel in linear representation) by compressing them to only 7-bits/pixel, while keeping visual quality in details and contrast. The in-pixel on-the-fly fully-parallel tone mapping achieves high frame rate allowing real-time HDR video (120dB@30fps).**

***Index Terms*—CMOS Image Sensor, High Dynamic Range, Tone Mapping, Pulse Width Modulation, Real-Time HDR Video.**

I. Introduction

THE acquisition of High Dynamic Range (HDR) scenes is required in a variety of applications such as surveillance, automotive, vision-enabled wireless sensors, machine vision, etc.[1][2]. There are two basic alternatives for acquiring HDR scenes depending on how the pixel data are acquired and coded, namely, either *linearly* [3][4] or using some *compressive* function [5]. Linear coding can obtain representations with plenty of subtle details at the cost of using many bits per pixel. For instance, some 25-bits would be required for the 151dB range measured in this paper. Such large bit counts increase the demands on the communication, data storage and post-processing circuits subsequent to the sensor. Compressive coding requires much less number of bits to achieve the same DR−the herein reported chip employs 7-bits to capture scenes with up to 151 dB DR. Although using less bits may result into fine image details not being captured [6], compressed representations are the best suited for applications such as machine vision and robotics, where the relevant information relies in image features [7][8].

Compressive acquisition sensors can be classified into three groups depending on the compression law:

i. Fixed: employs a fixed compression law [9][10];
ii. Programmable: the compression law is tunable but remains fixed during operation [5][11];
iii. Content-aware: the compression law is adaptive and depends on the scene being captured [12][13].

Therefore, sensors in the first two groups may largely attenuate contrasts and details in certain HDR scenes. These problems may also appear when fixed compression is employed to acquire Low Dynamic Range (LDR) scenes as they are unnecessarily compressed. Alternatively, sensors using a content-aware compression law consider that light is not the central information itself, but just the vector for information. Hence, compression is focused on capturing the actual scene information rather that the total light power range. Recently reported sensors belonging to this class employ event-driven techniques that require extra steps to perform the frame-based output composition [12][13]. The sensor reported in this paper addresses a different architectural approach based on the use of *tone mapping* techniques [14].

Tone mapping techniques were devised to confront the challenge of displaying detailed, artifact-free HDR scenes on low dynamic range reproduction systems, such as PC monitors. This is quite an appealing asset for our purposes. The problem is that tone mapping algorithms call for intensive computation. Thus, using them with conventional sensor architectures creates bandwidth problems [15][16] that may result into ghost artifacts in the case of video images [17]. Therefore, tone mapping is mostly employed for still HDR images using *multiple exposures* [18]. Recent advances on stacked sensors [19] and dedicated processing architectures [20] are addressing the challenge of using tone mapping for video. However, the DR of nowadays real-time HDR videos using multiple exposure techniques is limited as it does not allow direct videos of light sources or high-speed moving objects (see for instance the video clip in [21]). The chip in this paper enables real-time HDR video by implementing tone mapping compression at the pixel level, so that the algorithm execution is fully parallel.

Embedding circuitry in-pixel obviously results in area

S. Vargas-Sierra, G. Liñán-Cembrano and A. Rodríguez-Vázquez are with the Instituto de Microelectrónica de Sevilla (IMSE-CNM), Consejo Superior de Investigaciones Científicas (CSIC) and Universidad de Sevilla, C/ Américo Vespucio, Parque Tecnológico Isla de la Cartuja, 41092, Sevilla (SPAIN), e-mails: sonia@imse-cnm.csic.es, linan@imse-cnm.csic.es, arodri-vazquez@us.es, angel@imse-cnm.csic.es.

penalties and has a negative impact on pixel *pitch* and *fill factor* − similar to what happens for imagers using the so-called *digital pixels sensors* [22]**Error! Reference source not found.** and for the general class of vision sensors with per-pixel *focal-plane processing* [24][25]. This kind of architectures is not aligned to the trends of minimum size pixels and maximum spatial resolutions, pushed basically by consumer applications [26]. However, it is not clear that the quest for "as many pixels as possible" is pertinent for all applications. Actually, there is no agreement regarding the minimum number of pixels needed to obtain meaningful images [27]. Thus, for applications where the absolute pixel count is secondary (once a minimum number is granted), architectures as the one reported in this paper are worth considering. Furthermore, the concepts underlying this architecture remain valid for vertically integrated technologies [28], thus paving the way to the future implementation of large resolution, vertically integrated HDR image sensors.

This paper relies on previous conference contributions by the authors [29][30]. As compared to these previous publications, this paper provides detailed descriptions and reports more accurate DR measurements, including motion capture for the first time. It is organized as follows. Section II explains the concept of content-aware compression through exemplary images. Section III outlines the proposed focal-plane aimed content-aware adaptive tone mapping algorithm. Section IV describes the proof-of-concept HDR CMOS image sensor prototype implementing this algorithm. Section V depicts the error cancellation mechanisms on the chip. Section VI shows experimental results. Section VII discusses the design and obtained results. Finally, Section VIII presents our conclusions.

II. Illustrating Content-Aware Compression

Illumination conditions and hence radiance maps may change vastly as video scenes flow on such as for instance when a car enters or leaves a tunnel. Using a fixed compression function in such cases would result into missing details in poorly illuminated areas as many output codes are reserved to represent bright areas that do not actually exist in the scenes. A sounder strategy is assigning the output codes depending on the actual scene contents, following the evaluation of some *scene descriptor*.

Differences between these two compression approaches are illustrated in Fig.1 and Fig.2. Images in Fig.1 correspond to HDR scenes while those in Fig.2 correspond to a LDR scene. All images depicted in the figures are obtained by applying compression algorithms, which are mathematically generated radiance maps. The Desk Lamp (at the top in Fig.1) input is generated by us through multiple exposures, while the two other inputs (Memorial − medium images in Fig.1, and Bristol Bridge − bottom images in Fig .1) have been downloaded from radiance maps databases [31][32] and are customarily used to compare the results of different computer graphics tone mapping algorithms.

Regarding the output images depicted in Fig.1 and Fig.2, they are all obtained by simulation; namely, by using

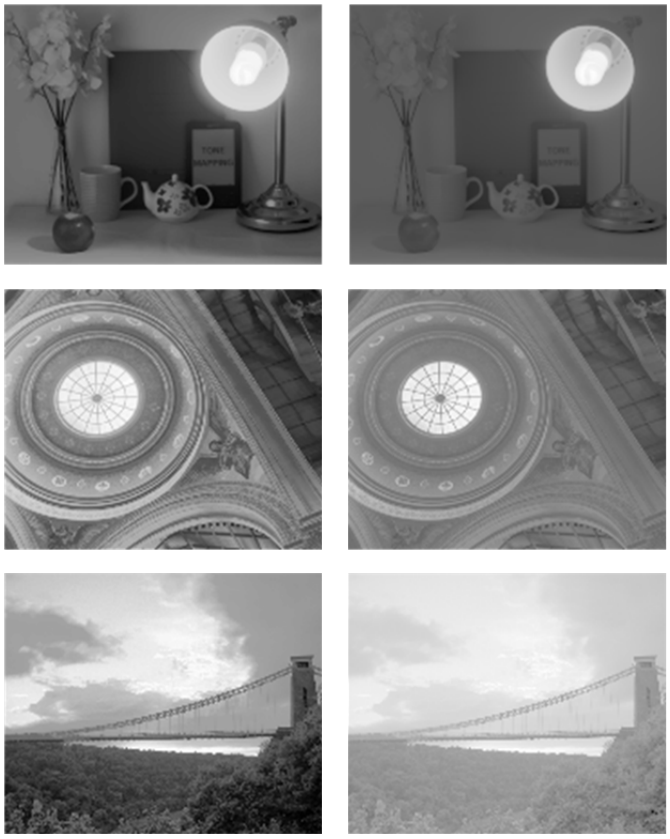


Fig. 1. Illustrating differences between logarithmic compression and content-aware compression for HDR scenes. Right-side images: logarithmic compression; Left-side images: content-aware compression..

logarithmic compression for the images in the right column, and content-aware compression for those in the left column. Specifically, the logarithmic images are obtained by applying a logarithmic compression and equalization (the lowest value corresponds to 0 and the largest to 255), which emulates a logarithmic pixel output [33] assuming the ideal case where the full output voltage swing is employed. The content-aware images are obtained by simulating the algorithm proposed in this paper through direct emulation of the circuit functionality. It means that only 128 digital codes are employed (half of those employed for the logarithmic compression) and that chip in-pixel hardware limitations are captured by the simulations.

A visual inspection of the figures reveals that the content aware compression function preserves more details and contrast despite using half the number of codes of logarithmic compression (128 vs. 256). However, the most remarkable feature of this compression mechanism is adaptability. This is illustrated by the images in Fig.2. They correspond to a LDR scene, which is actually a sub-scene of the desk lamp scene employed for the top images in Fig.1. The image at the right in Fig.2 is obtained through logarithmic compression while that at the left is obtained through content-aware compression. Letting aside the noticeable pixelation due to QCIF resolution, it is evident that Fig.2(a) contains much more information than Fig.2(b). Furthermore, the comparison of Fig.2(b) to the outcome of logarithmic compression for the corresponding whole scene, depicted at the right-top in Fig.1, shows that both

a) content-aware compression (b) logarithmic compression

Fig. 2. Illustrating differences between logarithmic compression and content-aware compression for a LDR scene.

images exhibit similar lack of details and contrast within the LDR sub-scene area that is magnified in Fig.2. It highlights the fact that a significant number of codes is reserved for bright pixels no matter if these pixels do either exist or not in the images.

The reason why content-aware compressed images preserve more details is that the compression law maps the light power containing actual information trying to cover the whole range of output codes.

Since the light power distribution of natural scenes is generally unknown, the chip reported in this paper first extracts the light distribution information from the incoming images (by evaluating a histogram) and then adapts the compression law to maximize the important captured details regardless the actual scene DR.

III. Tone Mapping Algorithm

Tone mapping classically belongs to computer graphics and it is usually performed by software in CPU, GPU, DSP, ISP, etc. This technique is used to compress HDR scenes into LDR output representations while preserving the details contained in the scene, usually as a human would perceive it.

HDR tone mapping compression can reduce the size of visual data by several orders of magnitude without introducing objectionable distortions. As an example, in this case, 151dB, which are approximately 7.5 orders of magnitude of visual data [$151=20\times\log_{10}(10^{7.55})$], can be reduced to only 128 levels with tone mapping. The tone mapping is not related with file image compressions (jpeg, etc.), but with illumination compression. This tone mapping compression has become an important part of many HDR image and video systems [14]. However, tone mapping techniques are usually intended for still images and little effort has been put towards HDR video tone mapping.

Tone mapping techniques are usually divided in four categories depending on the operators applied to the radiance maps of pixels:

- Global operators: The same function is applied to all the array of pixels. The function depends on global variables, such as the mean luminance or the histogram.
- Local operators: They apply a different function depending on the pixel. The function usually depends on local variables, such as the value of the pixel and its neighbours.
- Frequency Domain operators: The image is first transformed to the frequency domain. Then, the operators are executed on this domain. Finally, results are transformed back.
- Gradient Domain operators: The applied function depends on the derivative of the image; it is to say, the direction and value of the change between pixels.

Global operators are much faster than the others. Consequently, applications with limited resources and/or requiring high speed should preferably consider global operators. This is the case of the chip in this paper.

A. Image Coding Procedure

The chip captures images by using a dedicated tone mapping algorithm that expands the available output codes assigned to those illumination bands that are more largely populated, and reduces them in low populated bands.

Such expansions and compressions happen concurrently to image acquisition. This feature requires the on-line evaluation of some *image descriptor*; particularly, *histograms* are calculated and employed by our algorithm. Since we seek for employing a reduced number of output codes (128 for the herein reported chip) whereas illuminations may cover some 8 decades an involved trade-off arises between the number of captured illumination sub-bands and the number of output codes reserved to each band.

Pixels in the reported chip work by first, transforming light power into voltages according to the principle of integrating APS pixels [19]; then, transforming these voltages into temporal measurements; and, finally, coding these temporal measurements into tone-mapped digital codes.

Pixel operation is described with the help of Fig.3. The bottom part illustrates the transformation of voltages into temporal measurements. This part of the figure includes waveforms showing the temporal evolution of the pixel voltage following reset. Two cases are included, namely a

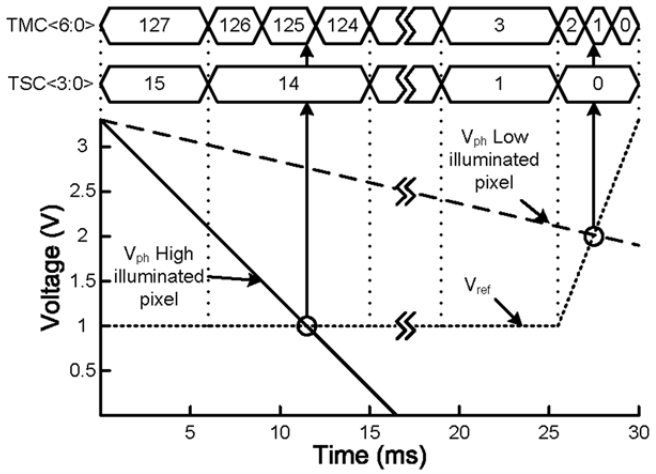


Fig. 3. Temporal evolution of the pixel voltage and illustration of the assignment of digital codes through digital sampling

bright pixel (continuous line) and a dark pixel (dashed line).

Into first order approach, the pixel voltage evolves during exposure according to:

$$V_{ph}(t) = V_{rst} - \frac{I_{pix}}{C_{pix}} t \quad (1)$$

where I_{pix} is the pixel photocurrent, C_{pix} is the pixel capacitance, and V_{rst} is the pixel voltage at the beginning of the exposure interval. This reset value is the outcome of an auto-zero, offset error compensation process which is explained in Section V-A. The equation above shows that the larger the photocurrent I_{pix}, the faster the pixel voltage drops. Voltage-to-time transformation is made by comparing the pixel voltage to a reference, V_{ref}. This reference signal is globally distributed to the whole array and hence shared by all pixels. Note that V_{ref} remains fixed during most of the exposure and ramps up near the end. This is so done to guarantee that voltages for low illuminated pixels do actually intersect V_{ref}.

When the pixel voltage crosses V_{ref} an event is created and detected by an in-pixel comparator as it is described in Section IV-B. The digital value codifying the pixel voltage, called Tone Mapping Code (TMC) is calculated at this event instant. The calculation consists of sampling a globally distributed digital signal (called TMC bus) at the time of the event. In the particular case of the chip reported in this paper, this bus has 7-bits (TMC<6:0>), but the algorithmic principle applies as well for other bit counts.

The TMC bus drives the digital codes sequentially during exposure, starting with the highest (127 for a 7-bits word) and ending with the lowest (0). However, the codes are lasting for different times in the bus, and those times depend on the illumination conditions; i.e., the time interval elapsed by each code is not a fixed parameter, but adapted to the nature of the stimuli. Actually, this interval is modified at each frame by using the outcome of the calculation of the light distribution in the previous frame. Thus, if the scene to be captured is dark, then a significant number of digital codes is assigned to the time interval during which V_{ref} ramps up - the last stage of the image capture process.

The V_{ref} waveform during this last stage of the image capture process is digitally-generated, driven by a DAC (see Fig.6 corresponding to the chip architecture), and digital coding happens in a way which is similar to that employed at imagers with single slope ADC converters [34]. However, in the present work, the digital codes stored at crossing events are the TMC values instead of the corresponding DAC input values, as in typical single-slope ADC.

The adaptation of the rate of change of the TMC word is made on the basis of the information provided by a second digital code called Time Stamp Code (TSC). As Fig.3 illustrates, the time stamp divides the exposure interval in different temporal windows. Similar to the TMC, the TSC is also obtained by sampling a digital bus at the time instant defined by the intersection of the pixel voltage with V_{ref}. However, this bus has fewer bits than the TMC bus, namely four (TSC<3:0>) in the case of the herein reported chip. Also not all pixels are equipped with the functionality of sampling the TSC bus (only one out of every four indeed). Thus, the image formed by the array of TSC codes has smaller dimensions than that formed by the TMC information sampled by the pixels. From now on, these two images will be respectively denoted as:

- Tone Mapping Image (TMI): It is the final tone-mapped representation of the scene, consisting of the TMC pixel sampled codes.
- Time Stamp Image (TSI): It is a radiance map descriptor consisting of the TSC pixel sampled codes.

The histogram of this latter auxiliary image is used to estimate, frame by frame, the statistics of impinging illumination and hence to adapt the rate of change of tone mapping codes accordingly. This is equivalent to adapt the number of TMC codes included within each temporal window defined by the time stamp signal.

In summary, the basic degrees of freedom for the tone mapping algorithm are: 1) The duration of the temporal windows for the TSCs, i.e. how long does every code last on the TSC bus.; 2) The number of TMC codes assigned to each TSC window, i.e. how many codes appear at the TMC bus within the temporal window of a single code in TSC.

Fig. 4 shows an example for 3-bits coding in TMC and 2-bits coding in TSC. The vertical thick lines indicate the temporal windows limits defined by the change of codes in TSC. Every temporal window will be a bin of the TSC histogram (radiance map descriptor). The vertical dotted lines indicate the evaluation windows limits, i.e. windows when we can sample the TMC bus (TMC code changes can only take place in the limits of evaluation windows). Since, a priori, all available output codes might be assigned during a temporal window, each temporal window must contain $2^{(\text{TMC-bits})}$ evaluation windows (2^3=8 in this case). Thus, there are $2^{(\text{TSC-bits})} \times 2^{(\text{TMC-bits})} = 2^2 \times 2^3 = 32$ evaluation windows. Obviously, since temporal windows have different time span, the evaluation windows are not equally distributed across time, although they are equally distributed within a given temporal window. When a temporal window was lowly populated in the previous frame, its relevance in the current frame will be largely attenuated by assigning very few TMC codes within its duration, or, in other words, by reducing the slope of the TMC curve within this temporal window. In the case of no new code assigned within this window, it will result in zero slope in this interval (as in this example between temporal windows TSC=3 and TSC=2, where TMC=7 does not change).

B. Temporal Window Estimation for TSC

The chip employs sixteen temporal windows. Thus, the TSC bus runs from 15 to 0 during exposure. The duration of each of the first 15 windows is selected by the user depending

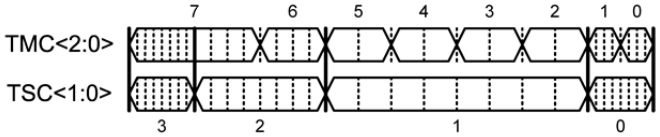


Fig. 4. TMC and TSC reference example with 3 and 2 bits, respectively.

on typical scenarios, maximum exposition time and hence required frame rate, etc. Regarding the last temporal window, when V_{ref} is ramped-up, its duration needs to be as short as possible. In this work, it has been set to 153.6μs, which allows to ensure that all the 128 TMC codes can be assigned within this last window meeting the accuracy constraints set by the circuitry employed to detect the temporal crossing events.

Regarding temporal windows during the V_{ref}-constant regime (first 15 temporal windows), a sound approach is defining them such that intervals corresponding to stronger illuminations (larger photocurrents) are more compressed than those with weaker illuminations (smaller photocurrents). The rationale for this is that dispersion in the crossing time estimation is inversely proportional to the photocurrent during the constant regime. By expressing the crossing time as:

$$T_{cross} = \frac{V_{rst} - V_{ref} \pm V_n}{I_{pix}} C_{pix} \qquad (2)$$

where V_n denotes an accumulated noise term at the pixel integration capacitance, the variance of this crossing time can be expressed as:

$$\frac{\sigma^2(T_{cross})}{T_{cross}^2} = \frac{\sigma^2(V_{rst}) + \sigma^2(V_n)}{(V_{rst} - V_{ref})^2} + \frac{\sigma^2(I_{pix})}{I_{pix}^2} + \frac{\sigma^2(C_{pix})}{C_{pix}^2} \qquad (3)$$

where $\sigma^2(V_{rst})$ represents dispersion of the applied reset voltage (including error correction), $\sigma^2(V_n)$ is the accumulated input referred noise, and $\sigma^2(I_{pix})$ accounts basically for the Photo Response Non-Uniformity (PRNU), and the dark current contribution.

Eq. (3) shows that the absolute required accuracy in the determination of the crossing time depends on the crossing time itself. This must be taken into account in the calculation of the duration of temporal windows. Specifically, the minimum duration of a temporal window must be such that all 128 codes can be assigned to a window and their corresponding crossing times estimated accurately. In order to achieve this, Fig.5 shows the results from MonteCarlo simulations for the evolution of the standard deviation of the crossing time vs. the average value of the crossing time for logarithmically spaced photocurrents. This figure confirms the inverse dependency between the error (standard deviation) and the photocurrent. The change in the trend observed for large photocurrents (above 1nA) has two reasons, which are related to the operation of the auto-zeroing reset circuit described in Section V-A, namely: 1) changes in the DC value of the reset voltage due to low power biasing of the pixels, and 2) errors in the settling time and dynamics of the reset process.

In any case, both errors are accounted for in these simulations to obtain the accuracy in determining the crossing time as a function of the photogenerated current.

Based on these simulations, the estimation of the minimum temporal window durations proceed as follows:

1. Given the total exposure time, we first subtract the duration of the V_{ref} ramping-up regime, and obtain a time value $T_{end_{15}}$, which is the final time of the 15th temporal window.
2. The standard deviation of the error in determining this crossing time (obtained from Fig.5) is multiplied by $6\sigma(T_{cross}) \times 128$ to obtain the minimum duration of the last temporal window (128 is the number of evaluation windows of TMC within every temporal window → $2^{(\text{TMC-bits})}$=128).
3. Once we know the duration (either its minimum value or any other higher), we can compute the final time for 14th window, and repeat the same procedure.
4. This operation is performed repeatedly until there is either no more free exposition time to allocate additional temporal windows, or all temporal windows have been allocated.

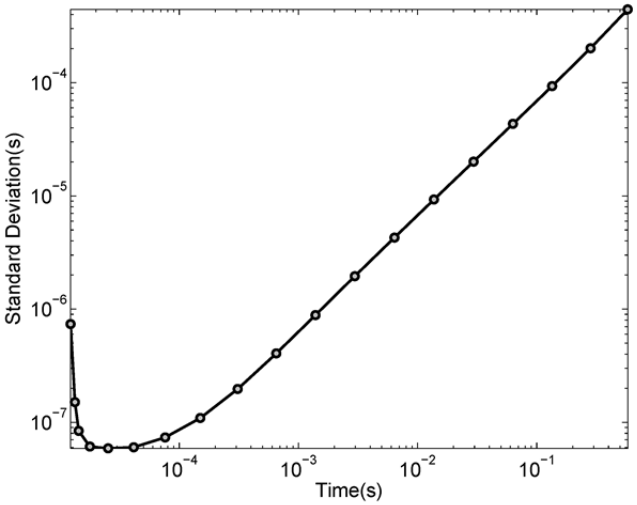


Fig. 5. Crossing Time Accuracy.

Clearly, for short exposure times, we would not be able to allocate the 15 available temporal windows meeting this 3σ accuracy strategy. In these cases, the user must choose whether to allocate lesser temporal windows (and reduce the number of different slopes in the Tone-Mapping Curve) or to relax the accuracy constraints in some temporal windows. Indeed, the minimum duration of temporal windows is defined such that, a priori, we could assign the 128 available output codes in a single window, something that is first very unlike, and second cannot occur for two windows in the same frame.

C. Assignment of TMC Levels per Window

The decision about how many TMC levels are allocated within every temporal window of TSC; in other words, the slope of the TMC curve in each temporal window, is calculated as a function of the percentage of pixels crossing V_{ref} within the different temporal windows in the previous frame. A simple way uses a direct correspondence between percentage and codes. Thus, if 50% of the pixels crossed during a certain temporal window, we will assign 50% of the available codes in TMC (64) to this window. Other assignment possibilities are contemplated in this work to enable the user to program the system such that certain light bands are prioritized. The currently implemented methods are:

1. *Histogram weighted*: The levels are distributed depending on the weight of the bins in the TSC

histogram (every temporal window –periods of constant value at the TSC bus– is also a bin). The number of TMC levels per TSC temporal window is obtained by directly scaling the contribution of a bin to the histogram by the number of TMC codes to assign (128). This simple calculation will result, in the most general case, in a non-integer number of codes to assign. The algorithm uses floor rounding and distributes all non-assigned codes (at the end of the process) among the bins with the higher remainders.

2. *Bin threshold*: A bin is active when the corresponding histogram bin value is higher than a fixed threshold. Levels are equally distributed between the all the active bins.
3. *Avoid concentration in one bin*: This mode operates as the weighted mode but it avoids assigning a high concentration of levels to a single bin. If the number of levels in a bin exceeds a configurable limit, saturation is applied and the exceeding levels are distributed between the other bins with already assigned levels.
4. *Weighted with bright priority*: If the most populated bin is a bright one, then this bin and the nearest lower neighbors are amplified by adding a fixed number of levels if they exceed a threshold. Remaining levels are assigned to TSC temporal windows proportionally to their weight in the TSC histogram.
5. *Low populated priority*: Low populated but non-empty bins are amplified by assigning them a fixed number of levels. The remaining levels are distributed as in the weighted mode among all the bins.
6. *Weighted with minimum threshold*: All non-empty bins are pre-assigned with a fixed number of levels. Remaining levels are distributed as in the weighted mode.
7. *Non-linear levels adjustment*: A non-linear function ($y = x^{1/1.75}$) is applied to the TSC histogram, and then the levels are distributed, as in the weighted case.
8. *Weighted with bright minimum low light priority*: A minimum amount of levels is assigned to all non-empty bins. Then, if the most populated bin corresponds to one bin receiving high illumination, the codes assigned to bins receiving low illumination are amplified by adding more codes. Finally, the remaining codes are distributed as in the weighted mode.

IV. Chip Architecture and Pixels

This system has been conceived as a complete Vision-System-on-Chip (VSoC). Therefore, all the digital functionality has been designed using a synthesis-ready Verilog code (does not employ any FPGA related macroblock), which could have been easily inserted on the chip due to the very reduced amount of computation. However, this chip being a proof of concept prototype, multiple redesigning of the tone mapping capabilities might be necessary to improve the system. Therefore, TSC Histogram Accumulation and Levels per Bin calculation (TMC codes per temporal window) and the subsequent timing control have been implemented off-chip in a FPGA for the sake of larger flexibility, redesign and reliability, as shown in Fig.6.

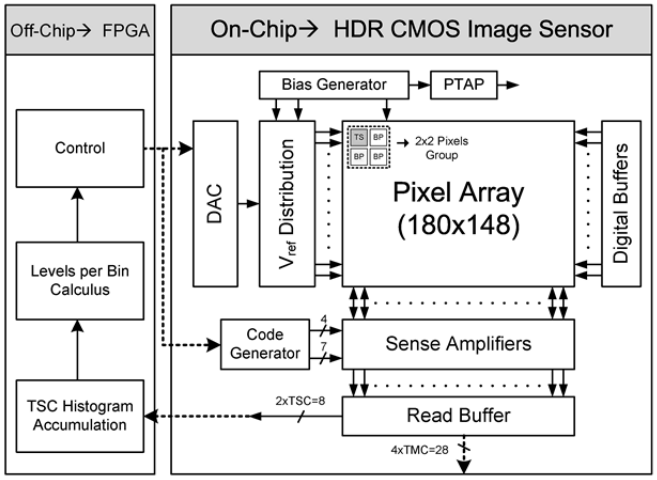


Fig. 6. Block Diagram of tone mapping system..

The proposed architecture follows the Single Instruction Multiple Data (SIMD) [35] concept as far as concurrency in the operation of the comparison and the storage of digital values is concerned. Fig.6 shows the architecture of the chip. Its core is a QCIF (176×144) resolution array of pixels, which are digital pixel sensors (DPS) with Pulse Width Modulation (PWM) [36], including a 2-pixel wide peripheral ring of dummy pixels. The functionality of the chip is supported by several blocks, which are responsible for biasing, generation of analog and digital references, digital control, and high-speed image I/O. The *global analog reference* for the pixels V_{ref} is generated by an internal resistor-ladder based DA converter, and it is supplied to the array by a row-wise distributed set of buffers. The *global digital control* signals are also supplied row-wise using a set of digital buffers including a clock-tree, which guarantee a maximum skew of 300ps. The *sense amplifiers* block is used to write and retrieve the digital codes TMC<6:0> and TSC<3:0> stored in static memories within the pixels. The *code generator* produces the digital signals to be stored in pixels using gray coding in order to diminish switching at the pixel level to only one bit per write cycle. The *read buffer* block is a 343 Mbit/s one-row buffer memory that uses a 36-bits output bus (4xTMC + 2xTSC).

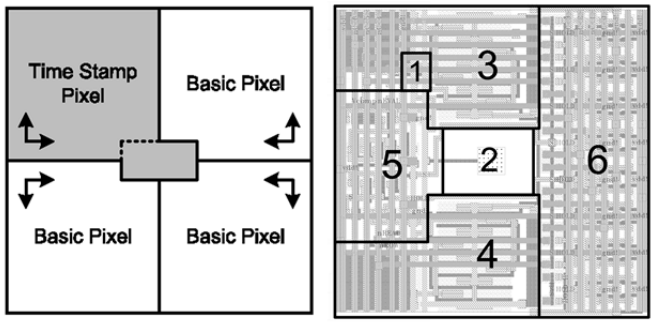
The chip contains all the necessary circuitry to acquire the TMI and the TSI images, as well as downloading them, already in 7-bits and 4-bits gray-coded digital format.

A. Pixel Arrangement in the Array

Two different pixel types are employed to extract both Tone Mapping Image (TMI) and Time Stamp Image (TSI), namely:

- Basic pixels (BP): Used to extract and store only tone-mapping codes (TMC).
- Time Stamp pixels (TSP): Used to extract and store both tone mapping codes and time stamp codes (TMC and TSC).

Hence, all pixels contribute to the image sensor output (TMI), while only TSP pixels contribute to the TSC image (TSI).



(a) Organization (including orientation marks)

(b) Base Layout.

Fig. 7. Pixel Grouping.

Pixels are physically arranged in groups of 2×2, each group including three BPs and one TSP −see Fig.7. It means that time stamp images (TSIs) have quarter the spatial resolution of TMIs. The reasons for doing so are: On the one hand, this simplifies the calculation of the histogram by reducing the number of pixels to account for; On the other hand, it improves the pixel pitch because hardware resources required to store time stamp codes can be shared among adjacent pixels −see central part of Fig.7 for illustration purposes; Last but not least, extensive simulations in the presence of circuit non-idealities show that such a sub-sampling has negligible impact on the resulting TMI, as this image is only employed for histogram calculation.

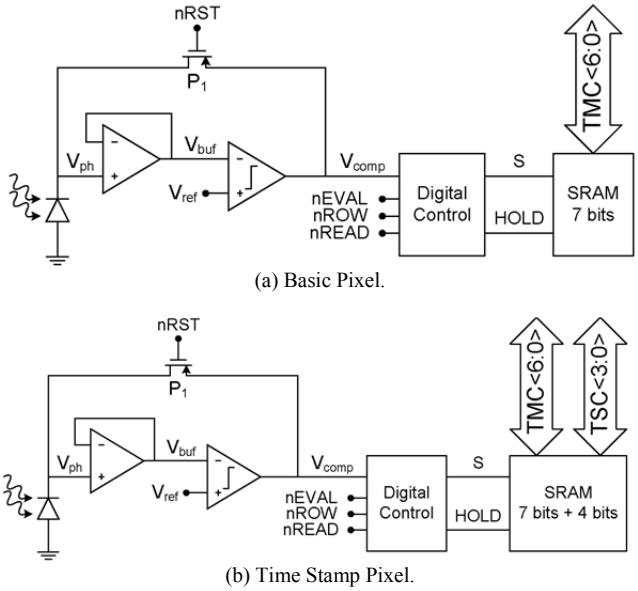
The TSP of each 2×2 group is placed, see Fig.7(a), at the upper-left corner of the group, and employs the group center area to allocate the 4 extra SRAM bit modules for TSC storage. Fig.7(b) shows the base layout of the pixels, which is replicated because they are virtually the same except for two wires. Table I shows the relative area occupation of the different building blocks.

Base layout is flipped left-to-right and upside-down, see orientation marks in Fig.7(a), seeking for: 1) improving the pitch by sharing well-areas, analog and digital power lines, etc., and 2) controlling the four center SRAM cells by the upper-left pixel (S and HOLD signals are extended).

Pixels contain an opening through all metal layers (taking advantage of the mandatory spacing around floating Nwells, such as the one used by our photodiode) of 9.75×7.3 μm^2. This opening enhances the sensitivity of the pixel due to the additional collection of photogenerated carriers, which reach the diode's depletion region by lateral diffusion [37]. Besides, the size of the opening increases as we move to higher metal layers thus reducing the vignetting effect.

TABLE I
PIXEL AREA OCCUPATION

Index	Devices	Area(%)
1	Reset Switch	1
2	Photodiode+Metal Opening	7
3	Comparator	21
4	Buffer	21
5	Digital Control	15
6	Memories (8-bits)	35



(a) Basic Pixel.

(b) Time Stamp Pixel.

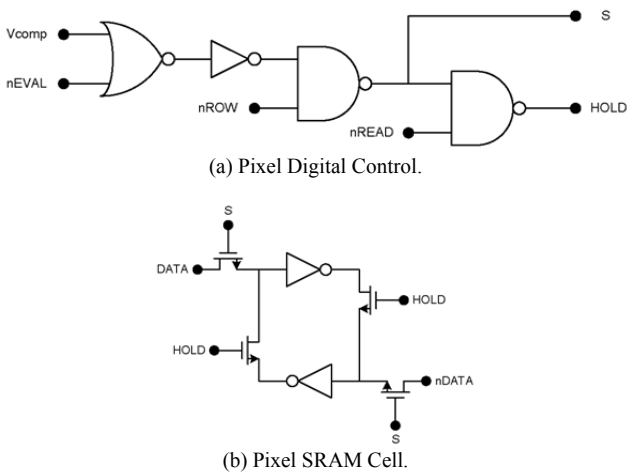
Fig. 8. Block Diagrams of Chip Pixels.

B. Pixel Circuitry

Fig.8 shows circuit diagrams for both pixel types. Note that the only difference is that TSPs include four more RAM bit cells to store the TSC<3:0> signal. In both cases, the integration node, V_{ph} is discharged by the photocurrent generated by a 3×3 μm^2 Nwell/Psub photodiode. This voltage is compared with the global analog reference, V_{ref}, by a comparator. A buffer is inserted between the integration node and the negative input of the comparator. This isolates the integration node (a mostly capacitive node discharged by a very weak photocurrent) from the kickback noise of the comparator, which is especially significant during the V_{ref} ramping up regime. MonteCarlo simulations of the kickback noise impact in case the buffer is not incorporated shows larger noise than allowed.

The photodiode is reset through a feedback loop established by the buffer, the comparator and the reset switch P_1. During this process a reset voltage, V_{rst}, is applied at the positive input terminal of the comparator and the resulting voltage at the V_{comp} node gets stored in the photodiode capacitance. This voltage includes the buffer+comparator offset. Hence, the offset gets subtracted from the signal path during integration, thus attenuating its impact. Such attenuation improves the operation during both the V_{ref} ramping up regime and the V_{ref} constant regime.

The buffer and the comparator employ the same 5T OTA topology. This way, the linear operation ranges of the buffer and the comparator are correlated and the combined linear range gets maximized thus maximizing the reset voltage V_{rst}. The PMOS feedback switch does not constraint the value of V_{rst}, and has its minimum size (0.4μm/0.35μm) in order to minimize the errors caused by channel charge injection. Despite the usage of an auto-zeroing loop to reduce the offset impact, the open loop offset must also be small enough to preclude the signal surpassing the linear range at the input of any of the amplifiers. Towards this end, the dimensions of the



(a) Pixel Digital Control.

(b) Pixel SRAM Cell.

Fig. 9. Pixel Digital Circuitry.

transistors in the OTA have been chosen by using mismatch models [38] to evaluate the trade-off between area and mismatch.

The digital control circuitry in the pixel, shown in Fig.9(a), has basically two functions: 1) To control the read and write operations of SRAM cells. 2) To amplify the output from the OTA acting as a comparator and to convert it to a rail-to-rail binary signal, i.e. to operate as the second stage of the comparator.

Proper tuning of the above mentioned amplification is critical for the final quality of the image since any mismatch among the quiescent point (input middle point between output high and low) of the pixel digital control in Fig.9(a) and the output of the comparator OTA for a zero differential input result in inaccuracies in the acquisition of either the TMC or TSC data. In order to minimize this effect, MonteCarlo simulations of every circuit in the signal path have been used to obtain the best design for the output stage of the comparator. Our target was the minimization of the variance of the time when the SRAM gets disconnected (through signal S) from the TMC or TSC bus.

The pixel SRAM cells employ an 8T SRAM structure, whose schematic is shown in Fig.9(b). Two transistors have been included to disengage the loop of the inverters during write operations (HOLD=0). This takes some more area but saves energy during the SRAM writing cycles, and also, more importantly, it reduces the disturbances in the local value of the power supply and ground by reducing the current consumption peaks. The external read and write operations to these memory cells are row-by-row controlled by the row-by-row signal nROW and the global signal nREAD. The signal nREAD is used to engage the inverters loop in the SRAM cells, in order to perform read operations without losing the data. The signal nROW controls the access to digital data row-by-row during image downloading operations. It is also employed to initialize the SRAM blocks to the maximum value of TMC and TSC. The global signal nEVAL controls the internal write operations. It acts as a clock that produces evaluation windows for the output of the comparator, diminishing uncertainties in determining the crossing time, and improving the quality of the final image.

V. Considerations For Error Cancellation

A. Auto-Zeroing Reset Technique

Offsets of the active circuits introduce errors within the signal path. For the pixels of Fig.8, the offsets are contributed by the buffer as well as by the comparator. Their impact can be addressed by resorting to cancellation schemes, such as *correlated double sampling* [39]. However, these techniques are useless in our case because they rely on subtracting images and hence would require linear coding. Instead, offset cancellation must be part of the capturing mechanism itself. Moreover, this cancellation must be stable in time without adding significant data distortion. Unfortunately, area constraints preclude embedding large analog memories in-pixel and restrain the options to use (and preferably reuse) simple circuitry.

The strategy adopted for offset attenuation consists of resetting the photodiode to a voltage that includes the buffer and comparator offsets− similar to a solution previously employed for time-to-first-spike pixels [40]. This strategy has already been mentioned in Section IV-B and is achieved by obtaining the pixel reset voltage through a feedback loop containing the buffer and the comparator

B. Dark Signal Impact Attenuation

The dark current [42] degrades the quality of the image. Being virtually independent of the signal level, their effects are clearly more noticeable at the dark parts of the image, where the signal is weaker. In our case, dark pixels are typically captured during the last temporal window, when the V_{ref} signal ramps up. Bearing this in mind, we have included a technique inherited from commercial reflex cameras, where a dark frame is subtracted in long time exposure shots [43]. Here, instead of subtracting the dark frame, we lower the top voltage V_{top} of the V_{ref} signal (i.e. we modify the range of the single slope DA converter at the end of the exposure), while keeping the bottom reference voltage V_{bot}. The idea is to measure dark current frames for a variety of exposure times and operating temperatures. This information can be used during normal operation, together with the readouts from a temperature sensor included in the chip and the exposure time, to control the value of V_{top}. Thus, V_{top} is lowered to the level (see Fig.10) produced by the average dark current ($\overline{I_{dark}}$) plus 3 standard deviations of it [$3 \cdot \sigma(I_{dark})$]. Conversely to what happens in typical dark frame subtraction techniques, where some output codes are unused , our schemes does not

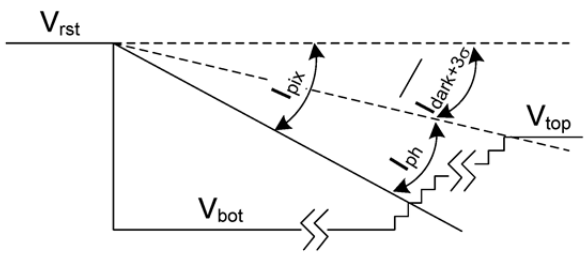


Fig. 10. Voltage adjustment for dark signal elimination.

sacrifice available output codes since it just modifies the operating range of the comparator. In comparison, our scheme uses all available output codes, while keeping accuracy due to the margin in voltage swing, which has been gained by the auto-zeroing technique.

C. Global Analog Reference Distribution Scheme

The precise and fast distribution of the global analog reference voltage, V_{ref}, is crucial to obtain uniform images with low noise − not easy since this signal must be distributed to a set of 26,640 (180×148) nodes in a pixel array of nearly 5mm by 6mm. Furthermore, the faster the evolution of V_{ref} from V_{rst} to V_{bot} (its fixed voltage during most of the exposition time), the higher the largest measurable light power, and hence, the wider the dynamic range.

Even using an ideal buffer with infinite power and speed, the evolution of the signal using a global buffer will be limited by the time constant of the whole array RC load. Therefore, despite the buffer used, a row-by-row approach is always faster, as speed is now only limited by the time constant of a single row, which can never be larger than that of an array made by adding a number of such rows. However, including a row-by-row buffer unavoidably leads to a row-distributed error in the stationary distributed value signals due to the different gains and offset voltages of each row buffer, and also to different temporal evolutions due to the different dynamic response of each buffer and different effective loads of every row. Being the slightly different temporal evolutions not too important (as long as the circuit remains stable, and the circuit does not become excessively slow) for our purposes, we concentrated efforts in devising a technique that on the one hand improves the overall dynamics (makes the circuit faster), and on the other hand improves the spatial uniformity of the distributed V_{ref} value.

This technique combines the use of row-wise distributed buffers inserted between the output of the DA converter, V_{dac}, and the V_{ref} horizontal metal lines (V_{ref_i}), during part of the dynamic evolution of the signal, and a direct connection to the DA converter at the final stages of this evolution.

The schematic of the inserted element (one per row) is shown in Fig.11. Complementary transfer gates provide two paths for the signal V_{dac} to V_{ref_i}. One path enables a direct connection whereas the other path inserts a buffer. The cell also includes the inverters to create the negate version of the control signal that enables either one path or the other, since it minimizes (as compared to a global generation of this signal, and its distribution cell by cell) the time when both paths are simultaneously engaged or disengaged. Besides, we have also included power-off control in the amplifiers in order to disconnect them from the power supply when they are not in use (during the constant V_{ref} regime), which is usually most of the exposure time.

By means of the signal NOVOFF, buffers are activated only during changes in V_{dac} (low NOVOFF). This inserts an error (a priori, one can assume that this error has a Gaussian distribution) in the V_{ref_i} lines due to the offsets of the buffers around its average value (V_{dac} output voltage). Once the V_{ref_i} lines are partially established, the direct path is engaged to eliminate the offset error. This produces a very fast redistribution of charges, and the lines evolve with the time constant of the whole array towards V_{dac}. When all V_{ref_i} are sufficiently established near V_{dac}, the evaluation signal (nEVAL=0) is transmitted to the pixels so that SRAM registers can sample the TMC or TSC buses depending on the output of the in-pixel comparator.

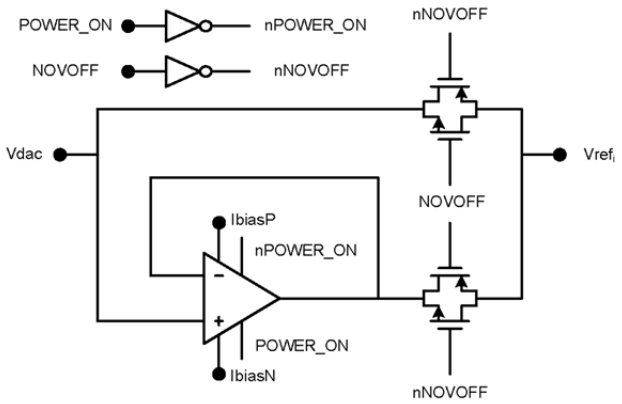


Fig. 11. V_{ref} distribution row cell schematic.

VI. Experimental Results

A prototype has been fabricated using 0.35µm 2P4M opto-flavored CMOS technology [44]. This technology incorporates an EPI substrate to decrease dark current and an anti-reflection-coating on top of the passivation layers for a smoother spectral response. Fig.12 shows a microphotograph of the chip whose size is 7.3×6.8mm^2 with the pixel array occupying 60% of the total area and the sense amplifiers and the read buffer placed at the bottom of the chip. In order to perform image capture and measurements, the chip has been mounted on a control and acquisition system, which for further details can be consulted in [41].

A. Performance Data

Table II summarizes the most relevant data measured from the chip. The Photon Transfer Curve (PTC) method [45] has

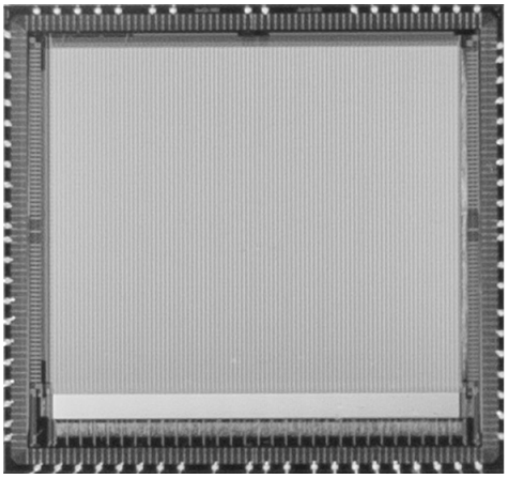


Fig. 12. Chip Microphotograph.

TABLE II
SUMMARY OF THE IMAGE SENSOR PERFORMANCE

Parameter	Value
Fabrication Process	AMS 0.35 µm 2P4M OPTO Technology
Supply Voltage	3.3V
Chip Size	7330x6780 μm^2
Array Size	180x148 (QCIF + 4 dummies)
Pitch	33x33 μm^2
Photodiode	3x3 μm^2 Nw-Psub
Pixels Metal Opening	9.75x7.3 μm^2
Fill Factor	0.8% (Photodiode), 7% (including metal opening)
Exposition time range	2.34 µs to 8 s
Image Coding	7-bits gray code
Sensitivity	5.79V/lux-s
Dark Signal	10.8mV/s (1.4DN/s)
Full Well Capacity	12.2 ke^-
Conversion Factor	129 (e^-/DN)
PRNU (PTC FPN Quality Factor)	0.018 (1.8 %)
Read Noise	25 e^- (0.2 DN)
Linear SNR (PTC DR)	53.8 dB
Linear DR (7-bit)	42.1 dB
Tone Mapping DR	151.45 dB @0.125fps 123.3dB@25fps 121.7dB@30fps
Min. Image Download Time	666 µs
Maximum Frame Rate	1205.4 fps (42.1dB)
Minimum Power Consumption	4.5mW
Maximum Power Consumption	111.2 mW@1205.4 fps

been used to characterize some of its main parameters. It requires the chip to be configured in linear mode (with all the TMC codes assigned to the last temporal window of TSC). Measurements reports a *read noise* of 25e^-, a *conversion factor* of 129e^-/DN (DN means data number in the photon transfer curve method), a *fixed pattern noise* (FPN) of 1.8%, a full well capacitance of 12.2ke^- and a maximum SNR (in linear mode) of 53.8dB. The read noise defines the denominator of the expression employed to calculate the SNR and contains the reset noise, the dark current noise, the quantization noise, the FPN (caused mostly by mismatch in the residual error of the auto-zeroing technique), and the input-referred noise of the signal processing chain (buffer + comparator + digital circuitry).

The dynamic range is calculated as follows:

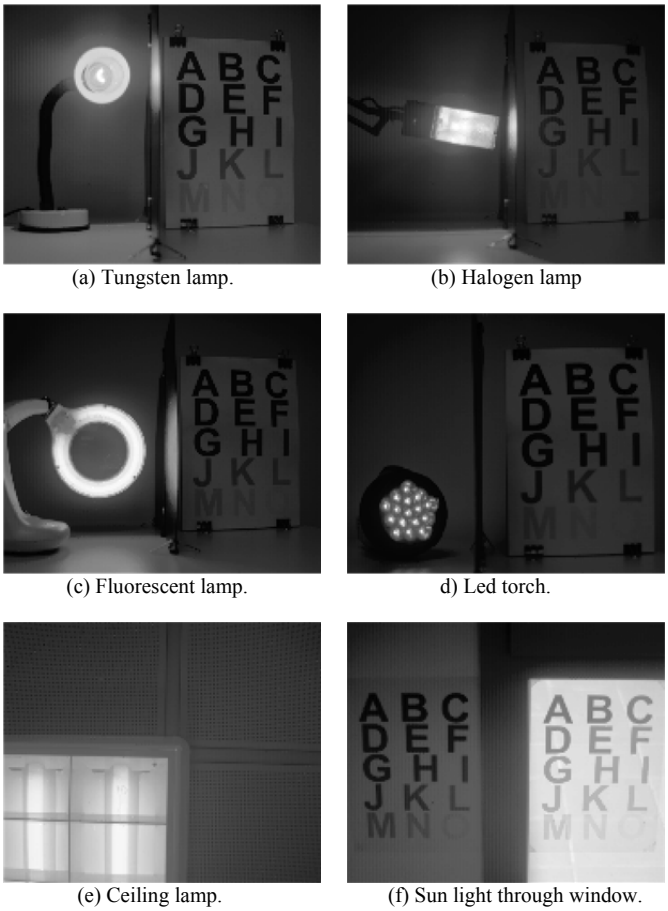
- *Maximum illumination*. The maximum measurable light power is determined from the shortest integration time measurements, which depends on the maximum measurable photocurrent and the sense node capacitance. Due to the auto-zeroing, the maximum photocurrent that allows a correct reset operation is limited by the bias current of the comparator OTA (50 nA). Moreover, any photogenerated current discharging the photodiode capacitance faster than the time it takes to change V_{ref} from V_{rst} to V_{bot} (i.e. how fast can we start the comparisons to V_{ref}) will not be detected either. Measurements reports a maximum measurable illumination of 55329lux captured within 2.34µs.
- *Minimum measurable illumination*. It is determined by the maximum integration time (dark signal limited) that produces an image where SNR=1. Therefore, the stable source of light output is dimmed through a combination of neutral density filters, which produce a light power decrement of several decades. Measurements results in a minimum measurable light of 1.48mlux, which is captured after an exposure lasting for 8s.

From these data, and taking into account that the chip can provide, in the same image, pixels that are captured just after the first 2.34µs, with pixels captured 8s later, the resulting maximum intra-frame DR is 151.45dB, which represents an enhancement of 109.35dB versus linear acquisition. This enhancement is based in skipping saturation basically by doing temporal instead of voltage measurements It means that there is a trade-off between the DR, on the one hand, and the exposition time and frame rate (speed), on the other hand. The maximum attainable frame rate is 1,205.4fps (for which a DR of only 42.1dB is obtained).DR values for video frame rates are above 120dB, particularly 123.3dB@25fps and 121.7dB@30fps.

B. Captured Tone Mapped Images

A typical approach to artificially create HDR images is making a direct photograph of a scene including a source of light and a set of objects not directly illuminated by it. Standard DR cameras are unable to simultaneously capture either the filament or other details of the light source and the surrounding objects in shadows. We have created some of this type of scenes by using different kind of light sources in order to test our system under several spectral emissivity values, emitting surfaces and continuous (tungsten, LED, halogen) or pulsating operation (fluorescent). In the four scenes of Fig.13(a)-(d), the light from sources is blocked to the printed characters (decreasing in contrast) by interposing a black cardboard. Regarding Fig.13(e), it contains a ceiling lamp containing 4 fluorescent tubes inside a sort of focusing light structure (high intensity source of light), while the ceiling is receiving only reflections from the room. The last image in Fig.13 corresponds to another typical HDR scene arising when one looks through a window, which is also the illumination source. Usually, the details observed outside the window are either clouds or buildings, while inside are either people or objects. However, these scenarios have smaller DR than previous ones. Hence, in Fig.13(f) the outside scenario is a courtyard where the year-highest intensity sun light is vertically hitting the floor near the window, while the room is only illuminated through the window.

The images depicted in Fig.13 shows that the proposed architecture and the prototype chip works properly with different kinds of light sources, while capturing the details of both very bright (even tungsten filament) and very dark areas with moderate-low noise (neither postprocessing adjustment nor de-noising have being applied). This is performed while creating very natural-like photograph, which means that tone-mapping implemented without causing compression artifacts. It is worth mentioning that poor tone mapping makes photographs to look very unrealistic with plenty of artifacts



(a) Tungsten lamp. (b) Halogen lamp

(c) Fluorescent lamp. d) Led torch.

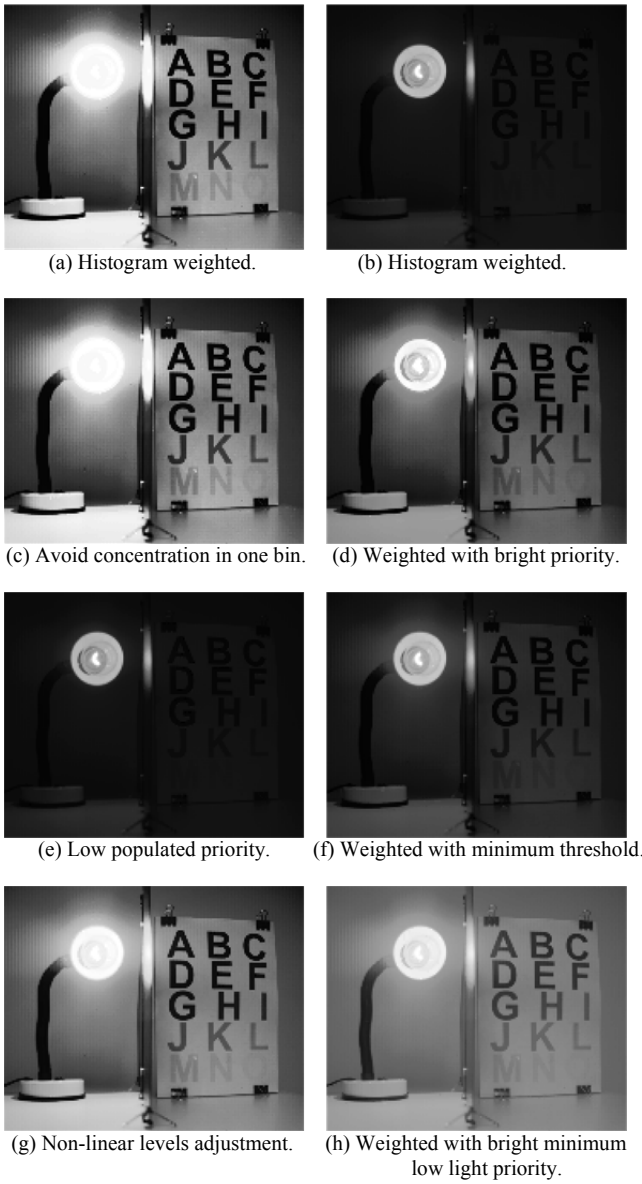
(e) Ceiling lamp. (f) Sun light through window.

Fig. 13. HDR scene captures using different sources of light.

created by the compression mechanism. It is also remarkable that the pulsating light does not create the typical gradients in rolling shutter image sensors. This is caused by the pixel architecture working as including a pseudo global shutter (global shutter for same illuminations).

Fig.14 shows the results of applying the different HDR operation modes featured by the system to the scene in Fig.13(a), while a summary of HDR modes performance has been included in Table III. Note that a little modification in this computation over only the sixteen values from the TSI histogram, has a great impact in the amplification or reduction of details of the scene, which can be very useful depending on applications. The following observations can be drafted from Fig.14:

- Fig.14(a)-*Histogram weighted*: The levels are distributed according to the histogram. Therefore, it shows a very high contrast in low light areas, but at the expense of losing details in the surrounding of the source of light because it occupies a small percentage of area in the whole image. It is not observable in the printed paper due to the low DR of the reproduction system, but a better reproduction system would reveal that the filament has been captured although with low contrast.
- Fig.14(b)-*Bin threshold*: It gives a good performance around the source of light, but it does not work so good in other areas because the light source intensity spreads over many temporal windows activating them.
- Fig.14(c)-*Avoid concentration in one bin*: Avoiding concentration has little effect in comparison with the mode used for Fig.14(a), because in this scene the concentration is not strong.
- Fig.14(d)-*Weighted with bright priority*: It is clear that it amplifies bright pixels (filament is more noticeable) taking some levels of detail from the background.
- Fig.14(e)-*Low populated priority*: It is meant to make clean borders (more codes assigned to low populated temporal windows) between significant illumination zones (highly populated temporal windows) at the expense of details in the these zones.
- Fig.14(f)-*Weighted with minimum threshold*: This mode generally gives a good overall result as it retains levels in any active bin, which creates clean borders, and retain good contrast as the rest of levels are weighted depending on area.



(a) Histogram weighted. (b) Histogram weighted.

(c) Avoid concentration in one bin. (d) Weighted with bright priority.

(e) Low populated priority. (f) Weighted with minimum threshold.

(g) Non-linear levels adjustment. (h) Weighted with bright minimum low light priority.

Fig. 14. HDR automatic tone mapping modes.

TABLE III
HDR MODES PERFORMANCE

HDR Mode	Contrast	Source of Light	Avoid Details Concentration	Clean Borders	Dark Areas
Histogram weighted.	Very high	Poor	Very poor	Poor	Good
Bin threshold	Very low	Very good	Very poor	Good	Very poor
Avoid concentration in one bin	High	Poor	Very good	Poor	Good
Weighted with bright priority	High	Good	Good	Medium	Good
Low populated priority	Very low	Very good	Very poor	Very good	Very poor
Weighted with minimum threshold	High	Very good	Good	Good	Good
Weighted over non-linearly transformed histogram	High	Medium	Very poor	Poor	Very good
Weighted with minimum threshold and low light priority	Medium	Good	Good	Medium	Very good

- Fig.14(g)-*Weighted over non-linearly transformed histogram*: It tries to mimic the application of the tone mapping over a non-linear acquisition, e.g. logarithmic. However, the applied function $y=x^{(1/1.75)}$ has better effect in light sources of lower intensities than the present in this scene
- Fig.14(h)-*Weighted with minimum threshold and low light priority*: Its effect is similar to the mode used for Fig.14(f), but it features lower contrast in the source of light because it amplifies dark areas.

Closer observation of Fig.14 highlights a column-wise pattern in the surroundings of the sources of light. This is actually an optical effect caused by the heterogeneous pixel array−meaning that most of the observable image disturbance is optical, rather than electrical. It happens that, at certain light incident angles, the shadows produced by the vertical metal lines of TMC and TSC are different in one pixel from those in its left and right neighbours. This effect can be worked out by either using a light shield surrounding the photodiode opening (using an additional metal layer, which was not available in this process), or, even better, by using a back side illuminated process in future evolutions of this chip.

It must be noticed that TSC can be extracted from TMC only in minimum threshold HDR operating modes. It is to say, if at least one level is always reserved for every temporal window. However, in the lack of experimental information about the behavior of this kind of HDR modes in combination with real world HDR scenes, this was not included in the present proof of concept prototype for a better understanding of the whole concept. After many test in real world HDR scenarios where the general good behavior of reserving levels is observed [see Fig. 14 (f)], future works will not include TSC generation, but it will be extracted from TMC.

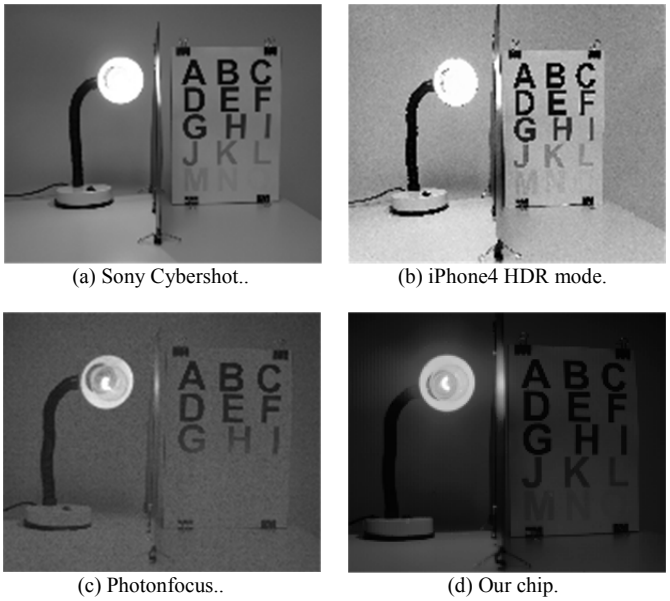
This section on static scene measurements is closed by showing comparisons to three commercial cameras, namely:

- Sony Cybershot DSC-W80, which uses a CCD array of sensors with enhanced sensitivity (Super HAD^{TM} CCD) [46].
- iPhone 4, that allows HDR mode by using a combination of 3 pictures of different exposure (since iOS 4.2) [47].
- Photonfocus Linlog MV-D752E-40-U2-12, which embeds the LINLOG^{TM} technology that combines linear response at low illumination levels with logarithmic compression at high intensities [48].

Images for comparison purposes are included in Fig.15. Noticeably, despite using only half of the codes (128 vs. 256) for image representation, our approach produces an image whose visible quality does not seem poorer than the others. The DSC-W80 is a little less noisy, but it shows both over and under exposed areas because it is a standard DR camera. The HDR mode in the iPhone 4 shows some similar performance in the darker areas, despite producing tone mapping artifacts (white pixels surrounding dark characters), but fails to produce details in the brightest areas because the composition of only 3 frames is not enough to cover such a high DR. Moreover, the algorithm of HDR composition based on the combination of noisy pixels (usual in mobile phone CIS) amplifies the noise in the final image. Finally, the Linlog sensor shows more details around the light bulb, but at the expense of a higher noise and a complete loss of details in darkest areas (J K L M N O characters are not distinguishable).

C. Moving Targets

Conventional tone mapping techniques use multiple exposures to synthesize the final HDR images and hence require the scene to be relatively stationary; otherwise,



(a) Sony Cybershot.. (b) iPhone4 HDR mode.

(c) Photonfocus.. (d) Our chip.

Fig. 15. Comparison with commercial cameras captures.

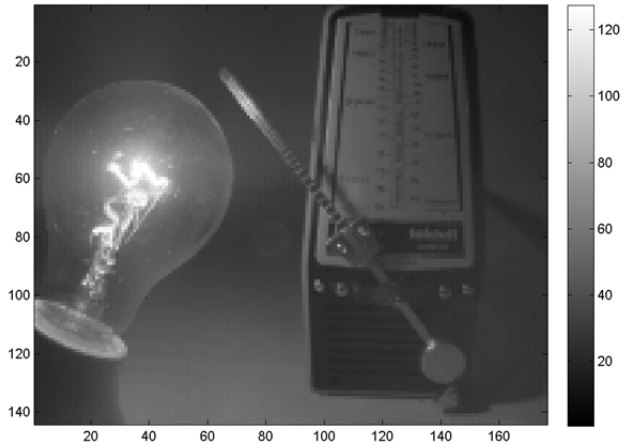


Fig. 16 HDR scene capture with fast moving objects.

ghosting artifacts (undefined blurred details) appear in the synthesized image [17].

A typical test scene for motion artifacts is a photograph with a controllable moving object, such as metronome. In our case, the chip is able to capture in the same frame both the filament of a light bulb and a fast moving object such as the metronome pendulum, as shown in Fig.16. It contains an image with time per frame of video rate of 30ms, while the metronome is set at its highest speed, observable by the bottom most position of the pendulum marker (208 beats per minute=288.5 ms per beat). The lack of blurred details in the pendulum proves the lack of ghost artifacts in the final image. The image contains two circles that softly highlight two circular zones, caused by an optical distortion. They are a typical photograph effect named *lens flare*, caused by internal reflection and scattering from material inhomogeneities in the lenses, and so independent of our image sensor.

D. *Illumination changes*

The algorithm compression function is not affected by moving objects (as far as histogram is kept basically unaltered). However, when illumination conditions change greatly, the compression function will need to change as well in order to adapt to the new situation to improve contrast and avoid saturation.

A typical scenario of heavily changing illumination conditions is that of entering and leaving a tunnel. In such cases, the video frames show a continuously increasing image area for the aperture of the tunnel (which turns to be an increasing dark area when entering and the opposite when leaving) for which the algorithm would be progressively adjusting its operation. Therefore, just at the moment of entering (or leaving) the tunnel, a set of TMC codes will be already reserved for that illumination bands and consequently the result is a smooth change in the compression function. A much more demanding scenario occurs when the system is imaging a light source surrounded by objects of interest, and this light source is switched on (or off). In this case, the compression function needs to change abruptly to adapt to the new illumination conditions.

Fig. 17 shows a set of consecutive frames from a 30fps video sequence, where the scene contains a switching LED lamp and a fast moving metronome pendulum. The scene also has a window in the background to show adaptability to multiple concurrent illuminations (indoors, outdoors and directly facing a light source). It is noticeable that when the LED light is switched on, TMC codes are taken from the rest of light bands (the contrast is slightly lower) to allocate them in the LED lamp zone to avoid saturation and so details inside the lamp are observable. In this abrupt change of illumination, the algorithm takes 1-2 frames to adapt to the new conditions, demonstrating the correct operation and stability of the algorithm when working under these very demanding conditions.

VII. Discussion of Results

Historically, there have been many excellent contributions to the field of HDR image sensor chips [6]. A detailed overview of all these contributions and the comparison of pros and counters of our technique versus all alternative approaches is out-of-the-scope of this paper. However, to the best of our knowledge no previous contribution had addressed the porting of tone mapping algorithms [14] into a custom designed chip. Such porting is relevant because tone mapping has uncontroversial advantages for HDR scene representation in LDR reproduction systems. Furthermore, the architectural solution reported in this paper, relying on in-pixel circuitry, yield parallel implementation of tone mapping algorithms, thus enabling tone mapping for video rates.

An appealing functional feature of our proposed architecture is its ability for content-aware, adaptive acquisition. Table IV compares our solution to other content-aware solutions reported in literature and yielding DR figures above 100dB [12][13]. Besides featuring large DR figures and allowing the automatic content-aware compression of illumination our proposal has the following advantages: 1) the output is ready to be used for frame-based post-processing, as

Fig. 17. Video sequence of switching lamp.

TABLE IV
CONTENT-AWARE COMPRESSION HDR IMAGE SENSORS COMPARISON.

Image Sensor	Dynamic Range	Compression Artifacts	Output Format	Image Output	Output Need Composition[1]	Focal Plane Storage[2]	Pixel Pitch	Photodiode Type	Technology Node	Power Consumption	Image Resolution
[12]	>100 dB	Yes	15-bits	Event-based	Yes	No	17 μm	PN	0.35 μm	540μW	128×128
[13]	143 dB	No	18-bits	Event-based	Yes	No	30 μm	PN	0.18 μm	50mW	304×240
This work	151 dB	No	7-bits	Frame-based	No	Yes	33 μm	PN	0.35 μm	4.5mW	180×148

[1]for typical image post-processing (frame-based).
[2]for 3D integration next tier direct post-processing.

it does neither contains artifacts nor require composition, 2) it exhibits noise levels low-enough as to preclude the usage of de-noising algorithms, and 3) the image is stored at the focal plane with low bit-count representation.

This latter feature is very appealing when considering a migration to future 3-D vertically integrated system [28]. Specifically, the digital focal plane storage permits a straightforward use of the image sensor as an image frame buffer when incorporated as part of a vertically-integrated system. In this potential system, our image sensor might be massively connected to the image processor by using, for instance, Through Silicon Vias (TSV) [49]. Thus, the die (or tier, as they are typically known on 3-D technologies) on top might contain our design, whereas the die underneath might contain a digital processor, standardized I/Os, third-party DSP blocks, etc. Moreover, using hybrid 3-D integration technologies, one might also figure out the possibility of designing the HDR image sensor front-end using a specialized CMOS image sensor (CIS) technology, and designing the digital part using an ultra-deep sub-micron purely digital CMOS technology. In this case, it will be only one signal (the output of the comparator) going from top-tier to second tier to inform the digital circuitry when to sample the TMC and TSC buses.

Another important aspect of our solution is the reduction in computation requirements, power and area occupation, which appears when comparing the present work with the concept of frame-based off-chip real-time HDR video integrated solutions. The most important point to figure this out is to keep in mind that our system is able to provide a frame where pixels may have been captured as fast as only 2.34μs after reset and as slow as 8s after, with a granularity of 20ns (while V_{ref} is constant in the first 15 temporal windows). Moreover, as the histogram is calculated concurrently to image download, the calculations for algorithm adaptation (levels per bin calculus, etc.) takes only 320ns, which is negligible when compared to frame time (only 0.04% of frame time for the fastest frame rate). In order to obtain such granularity and processing speed using multiple frame composition (which always requires downloading the whole image), one would need to capture, download, and compose images at 50,000,000 fps. Moreover, it should be able to combine as much images as possible evaluations windows → 16×128=448 frames in 1/30fps=33ms in order to allow video rates. Clearly, the finally composed image will feature much more bits-per pixel that our solution, but, would it result in a more useful image? Probably not, in particular in applications such as those our chip is targeted to, where what it is pursued is to detect situations (like surveillance, automotive safety) and not to produce ultra-realistic scenes. In addition to that, working at this frame-rate would require: 1) very high-speed pixels, which penalize area and power due to the need to strengthen the in-pixel source follower, 2) very high-speed column-parallel ADC, also penalizing area and power, 3) a very fast Image Sensor Processor (ISP) to combine an impressive amount of frames at video rates.

VIII. Conclusions

We have presented a 151dB (SNR1) image sensor that compresses the HDR scene in a 7-bits/pixel format using a self-adaptable Tone Mapping algorithm, which is capable of working at video frame rates without producing ghosting artifacts in the presence of fast moving objects, neither producing gradients under pulsating light illumination. Pixels include auto-zeroing to diminish FPN, and SRAM cells to allow for very long exposure shots with focal-plane digital storage. A dark signal contribution mitigation scheme has been implemented to enhance the visual quality in dark areas. Global analog reference to the pixels is dynamically distributed to provide fast and precise operation.

The main conclusion is that it is possible to execute an on-the-fly in-pixel adaptive tone mapping (resulting in good quality images) at video frame rates, using an in-pixel architecture that creates an arbitrary illumination compression function depending on external references.

Acknowledgment

This research has been supported by ONR through Project N000141110312 and the Spanish Ministry of Science and Innovation through Project IPT-2011-1625-430000.

References

[1] D. Hertel, A. Betts, R. Hicks, and M. ten Brinke, "An adaptive multiple-reset CMOS wide dynamic range imager for automotive vision applications," in Intelligent Vehicles Symposium, 2008 IEEE, June 2008, pp. 614 -619.

[2] N. Akahane, S. Sugawa, S. Adachi, and K. Mizobuchi, "Wide dynamic range CMOS image sensors for high quality digital camera, security, automotive and medical applications," in Sensors, 2006. 5th IEEE Conference on, Oct. 2006, pp. 396 -399.

[3] S. Kawada, S. Sakai, N. Akahane, K. Mizobuchi, and S. Sugawa, "A color-independent saturation, linear response, wide dynamic range CMOS image sensor with retinal rod- and cone-like color pixels," in VLSI Circuits, 2009 Symposium on, June 2009, pp. 180 -181.

[4] M.W. Seo, S.H. Suh, T. Iida, T Takasawa, K. Isobe, T Watanabe, S. Itoh, K. Yasutomi, and S. Kawahito, "A low-noise high intrascene dynamic range CMOS image sensor with a 13 to 19b variable-resolution column-parallel folding-integration/cyclic ADC," Solid-State Circuits, IEEE Journal of, vol. 47, no. I, pp. 272 -283, Jan. 2012.

[5] D. Stoppa, M. Vatteroni, D. Covi, A. Baschirotto, A. Sartori, and A. Simoni, "A l20-dB dynamic range CMOS image sensor with

programmable power responsivity," Solid-State Circuits, IEEE Journal of, vol. 42, no. 7, pp. 1555 -1563, July 2007.

[6] A. Spivak, A. Belenky, A. Fish, and O. Yadid-Pecht, "Wide dynamic range CMOS image sensors - comparative performance analysis," Electron Devices, IEEE Transactions on, vol. 56, no. II , pp. 2446 - 2461, Nov. 2009.

[7] C. Posch, M. Hofstatter, M. Litzenberger, D. Matolin, N. Donath, P. Schon, and H. Gam, "Wide dynamic range, high-speed machine vision with a 2x256 pixel temporal contrast vision sensor," in Circuits and Systems, 2007. ISCAS 2007. IEEE International Symposium on, May 2007, pp. 1196 -1199.

[8] P. Lichtsteiner, T Delbruck, and C. Posch, "A 100 dB dynamic range high-speed dual-line optical transient sensor with asynchronous readout," in Circuits and Systems, 2006. ISCAS 2006. Proceedings. 2006 IEEE International Symposium on, May 2006, pp. 4 pp. - 1662.

[9] H. Amhaz and G. Sicard, "A high output voltage swing logarithmic image sensor designed with on chip FPN reduction," in Ph.D. Research in Microelectronics and Electronics (PRIME), 2010 Conference on, July 2010, pp. I -4.

[10] D. Das and S. Collins, "Fixed-pattern-noise correction for an integrating wide-dynamic-range CMOS image sensor," Electron Devices, IEEE Transactions on, vol. 60, no. I, pp. 314 -319, Jan. 2013.

[11] M. Vatteroni, P. Valdastri, A. Sartori, A. Menciassi, and P. Dario, "Linear logarithmic CMOS pixel with tunable dynamic range," Electron Devices, IEEE Transactions on, vol. 58, no. 4, pp. 1108 - 1115,, April 2011.

[12] C. Shoushun and A. Bermak, "Arbitrated time-to-first spike CMOS image sensor with on-chip histogram equalization," Very Large Scale Integration (VLSI) Systems, IEEE Transactions on, vol. 15, no. 3, pp. 346-357, March 2007.

[13] C. Posch, D. Matolin, and R. Wohlgenannt, "A QVGA 143 dB dynamic range frame-free PWM image sensor with lossless pixel-level video compression and time-domain CDS," Solid-State Circuits, IEEE Journal of, vol. 46, no. I, pp. 259 -275, Jan. 2011.

[14] E. Reinhard, G. Ward, S. Pattanaik, and P. Debevec, High Dynamic Range Imaging: Acquisition, Display, and Image-Based Lighting. Elsevier / Morgan Kaufmann, 2006.

[15] C.-T Chiu, T-H. Wang, W.-M. Ke, C.-Y. Chuang, J.-R. Chen, R. Yang, and R.-S. Tsay, "Design optimization of a global/local tone mapping processor on arm SOC platform for real-time high dynamic range video," in Image Processing, 2008. ICIP 2008. 15th IEEE International Conference on, Oct. 2008, pp. 1400 -1403.

[16] Aptina. AP0I0IAT dedicated automotive image co-processor. [Online]. Available: http://www.aptina.com/products/co-processors/ap0101at/

[17] E. Khan, A. Akyuz, and E. Reinhard, "Ghost removal in high dynamic range images," in Image Processing, 2006 IEEE International Conference on, Oct. 2006, pp. 2005 -2008.

[18] T. Jinno and M. Okuda, "Multiple exposure fusion for high dynamic range image acquisition," Image Processing, IEEE Transactions on, vol.21, no. 1, pp. 358 -365, Jan. 2012.

[19] Sony Develops "Exmor RS," the World's First*1 Stacked CMOS Image Sensor. [Online] Available: http://www.sony.net/SonyInfo/News/Press/201208/12-107E/.

[20] Chimera™: The NVIDIA Computational Photography Architecture.[Online]Available: http://www.nvidia.com/docs/IO//116757/Chimera_whitepaper_FINAL.pdf.

[21] Xperia™ Z - Exmor RS™ for mobile, the world's first image sensor with HDR video for smartphones [Online]Available: https://www.youtube.com/watch?feature=player_embedded&v=viE6LcrlwiQ

[22] A. El Gamal and H. Eltoukhy, "CMOS image sensors," Circuits and Devices Magazine, IEEE, vol. 21, no. 3, pp. 6 - 20, May-June 2005.

[23] Cheng-Hsiao Lai; Ya-Chin King; Shi-Yu Huang, "A 1.2-V 0.25-µm clock output pixel architecture with wide dynamic range and self-offset cancellation," Sensors Journal, IEEE , vol.6, no.2, pp.398,405, April 2006

[24] A. Zarandy, Ed., Focal-Plane Sensor-Processor Chips. Springer New York, 2011.

[25] T. Roska and A. Rodriguez-Vazquez. Towards the Analogic Visual Microprocessor. John Wiley & Sons, Chichester 2001.

[26] Kitamura, K.; Watabe, T.; Sawamoto, T.; Kosugi, T.; Akahori, T.; Iida, T.; Isobe, K.; Watanabe, T.; Shimamoto, H.; Ohtake, H.; Aoyama, S.; Kawahito, S.; Egami, N., "A 33-Megapixel 120-Frames-Per-Second 2.5-Watt CMOS Image Sensor With Column-Parallel Two-Stage Cyclic Analog-to-Digital Converters," Electron Devices, IEEE Transactions on, vol.59, no.12, pp.3426,3433, Dec. 2012.

[27] A. Torralba, "How many pixels make an image?", Visual Neuroscience, vol. 26, pp. 123–131, 2009.

[28] K. Chen and C. Tan, "Integration schemes and enabling technologies for three-dimensional integrated circuits," Computers Digital Techniques, IET, vol. 5, no. 3, pp. 160 –168, May 2011.

[29] S. Vargas-Sierra, G. Linan-Cembrano, and A. Rodriguez-Vazquez, "A 148dB focal-plane tone-mapping QCIF imager," in Circuits and Systems (ISCAS), 2012 IEEE International Symposium on, 2012, pp. 1616-1619.

[30] S. Vargas-Sierra, G. Linan-Cembrano, and A. Rodriguez-Vazquez, "A 176xl44 148dB adaptive tone-mapping imager," in Sensors, Cameras, and Systems for Industrial and Scientific Applications XIII, 82980A (February 9, 2012), vol. SPIE 8298, 2012.

[31] P. Debevec. Radiance maps .hdr files data base. [Online]. Available: http://www.pauldebevec.com/Research/HDR/#radiancemaps

[32] E. Reinhard. Radiance maps .hdr files data base (book). [Online]. Available:http://www.mpi-inf.mpg.de/~reinhard/hdr_book/html/hdr_images_5.html

[33] B. Choubey, S. Aoyoma, S. Otim, D. Joseph, and S. Collins, "An electronic-calibration scheme for logarithmic cmos pixels," Sensors Journal, IEEE, vol. 6, no. 4, pp. 950–956, 2006.

[34] M. Snoeij, A. Theuwissen, K. Makinwa, and J. Huijsing, "Multiplea-ramp column-parallel ADC architectures for CMOS image sensors," Solid State Circuits, IEEE Journal of, vol. 42, no. 12, pp. 2968 -2977, Dec.2007.

[35] W.-Y. Lo, D. Lun, W.-C. Siu, W. Wang, and J. Song, "Improved SIMD architecture for high performance video processors," Circuits and Systems for Video Technology, IEEE Transactions on, vol. 21, no. 12, pp. 1769 –1783, Dec. 2011.

[36] D. Chen, D. Matolin, A. Bermak, and C. Posch, "Pulse modulation imaging - review and performance analysis," Biomedical Circuits and Systems, IEEE Transactions on, vol. 5, no. I, pp. 64 -82, Feb. 20 II.

[37] I. Shcherback and O. Yadid-Pecht, "Photoresponse analysis and pixel shape optimization for CMOS active pixel sensors," Electron Devices, IEEE Transactions on, vol. 50, no. I, pp. 12 - 18, Jan 2003.

[38] Lovett, S.J.; Welten, M.; Mathewson, A.; Mason, B., "Optimizing MOS transistor mismatch," Solid-State Circuits, IEEE Journal of , vol.33, no.1, pp.147,150, Jan 1998.

[39] M. Hafiane, W. Wagner, Z. Dibi, and O. Manck, "Depth resolution enhancement technique for cmos time-of-flight 3-d image sensors," Sensors Journal, vol. 12, no. 6, pp. 2320–2327, 2012.

[40] X. Guo, X. Qi, and J. Harris, "A time-to-first-spike CMOS image sensor," Sensors Journal, IEEE, vol. 7, no. 8, pp. 1165 –1175, Aug. 2007.

[41] Vargas-Sierra, S.; Linan-Cembrano, G.; Rodriguez-Vazquez, A., "Control and acquisition system for a High Dynamic Range CMOS Image Sensor," Electronics, Circuits and Systems (ICECS), 2012 19th IEEE International Conference on , vol., no., pp.586,589, 9-12 Dec. 2012. doi: 10.1109/ICECS.2012.6463679

[42] X. Wang, Noise in sub-micron CMOS image sensors. Ph.D. dissertation, Technische Universiteit Delft, 2008.

[43] Nikon. D90 user manual (long exposure noise reduction). [Online]. Available:http://www.nikonusa.com/pdf/manuals/kie88335f7869dfuejdl=-cww2/D90_en.pdf

[44] AMS CMOS processes [Online] Available: https://www.ams.com/eng/Products/Full-Service-Foundry/Process-Technology/CMOS

[45] J. R. Janesick, Photon Transfer. Bellingham, WA: SPIE Press, 2007. [Online]. Available: http://dx.doi.org/10.1117/3.725073

[46] Dpreview. Sony cyber-shot W80 review.. [Online]. Available: http://www.dpreview.com/reviews/sonyw80/

[47] Apple. Iphone 4 user guide. Apple. [Online]. Available: http://manuals.info.apple.com/MANUALS/1000/MA1565/en_US/iphone_user_guide.pdf

[48] Photonfocus. Photonfocus linlog MV-D752E-40-U2-12. Photonfocus. [Online]. Available: http://www.photonfocus.com/html/eng/products/products.php?prodId=55

[49] D. Henry, J. Charbonnier, P. Chausse, F. Jacquet, B. Aventurier, C. Brunet-Manquat, V. Lapras, R. Anciant, N. Sillon, B. Dunne, N. Hotellier, and J. Michailos, "Through silicon vias technology for CMOS image sensors packaging: Presentation of technology and electrical results," in Electronics Packaging Technology Conference, 2008. EPTC 2008. 10th, Dec. 2008, pp. 35 –44.

Sonia Vargas-Sierra received the 3-year degree in industrial engineering speciality in industrial electronics from the University of Cádiz, Spain, in July 1998, B.S. degree in electronic engineering from the University of Málaga, Spain, in July 2001 and Ph.D. degree in Microelectronics from the University of Seville, Spain, in September 2012. From 2005 to 2009, she got a doctoral Grant at the Institute of Microelectronics of Seville CNM-CSIC funded by the Spanish Ministry of Education. She has performed stays in both the cognitive computing group of the Fraunhofer Institute of Darmstadt (Germany) and the vision group of the Fraunhofer Institute of Duisburg (Germany). She has received ICECS 2012 Ph.D. Demo Competition Award. Her main research areas of interest are the High Dynamic Range Image Sensors, Vision Systems on Chip (VSoC), image processing and 3D integrated circuits.

Gustavo Liñán-Cembrano received his PhD) in Physics, from the University of Seville, Spain. He was an Assistant Professor in the University of Seville until 2004 when he joined the Institute of Microelectronics of Seville as a Tenured Scientist of the Spanish National Research Council. His main areas of interest are the design and VLSI implementation of massively parallel analog/mixed-signal image sensors and processors. He has received the Best Paper Award 1999 and Best Paper Award 2002 from the International Journal of Circuit Theory and Applications. He is also corecipient of the Most Original Project Award, of the Salvà i Campillo Awards 2002, conceded by the Catalonian Association of Telecommunication Engineers.

Ángel Rodríguez-Vázquez (PhD, IEEE Fellow) is a Full Professor of Electronics at the University of Seville. He is also the President and the responsible for long term R&D at AnaFocus Ltd.

His research is on the design of analog and mixed-signal front-ends for sensing and communication, including smart imagers, vision chips and low-power sensory-processing microsystems. He has authored 11 books, some 43 chapters in contributed books, including original tutorials on chaotic integrated circuits, design of data converters and design of chips for vision, and some 150 journal articles in peer-review specialized publications. He has presented many invited plenary lectures at different international conferences and has received a number of international awards for his research (the IEEE Guillemin-Cauer best paper award, two Wiley´s IJCTA best paper awards, two IEEE ECCTD best paper award and the IEEE ISCAS best demo-paper award). He was elected Fellow of the IEEE for his contributions to the design of chaos-based communication chips and neuro-fuzzy chips. His research work got some 5,700 quotes; he has an h-index of 39 and a i10-index of 117.

He has always been looking for the balance between long term research and innovative industrial developments. He founded AnaFocus Ltd. in 2001 and served as CEO, on leave from the University, until June 2009, when the company reached maturity as a worldwide provider of smart CMOS imagers and vision systems-on-chip.

He has served as Editor, Associate Editor and Guest Editor for different IEEE and non-IEEE journals, is in the committee of many international journals and conferences, and has chaired several international IEEE and SPIE conferences. He served also as VP Region 8 of the IEEE Circuits and Systems Society (2009-2012) and as Chair of the IEEE CASS Fellow Evaluation Committee (2010, 2012 and 2013).