

# Center-Surround Dynamic Vision Sensor

Brian McReynolds, Mert Polatkaya, Xi Chen, Hyunjung Hwang, Lioba Schürmann, Yassine Taoudi-Benchekroun

**Abstract**—Dynamic Vision Sensors (DVS), or event cameras, are an increasingly popular technology for many important applications in computer vision requiring high speed, low latency, and wide-dynamic range video. DVS are inspired by the design of biological retinas and achieve key benefits by responding only to dynamics in a scene, disregarding low-frequency or DC information in the time domain. This can drastically reduce data requirements, however uninformative low-spatial frequency information is still reported. Many early silicon retinas included an additional biological feature intended to reduce data by implementing an antagonistic center surround network to suppress low spatial frequency information. For practical reasons, this feature has not been implemented in any commercially available DVS thus far. A new center surround design utilizing unsalicyded polysilicon was recently proposed, and we explore the feasibility of the design through circuit simulations. Our results show that the proposed architecture is feasible, and we identify key design considerations for a possible future CSDVS.

**Index Terms**—DVS, event camera, center surround

## I. INTRODUCTION

“The notion of a “frame” of video data has become so embedded in machine vision that it is usually taken for granted”. This bold statement still holds true today, 14 years after its writing within the publication of the first Asynchronous Vision Sensor breakthrough [1]. Today, more than ever, with the advent of machine learning and the increasing amount of technologies crucially needing low power and low-latency function, it is of crucial importance to continue the efforts in developing asynchronous vision sensors. The main motivation behind asynchronous vision sensors, as opposed to frame based cameras, is that data is only generated when changes occur in the scene. This reduces data-redundancy and focuses attention on changes in the scene, which are of highest interest in many applications.

One shortcoming of this framework however, is that unwanted events are generated in many common situations, such as changes in lighting conditions (e.g. cloud suddenly obstructing the sun) or through artificial lighting systems (sodium, LED, and fluorescent) which flicker at some frequency. This can cause high peaks of uninformative events, making identification of important events more difficult. Multiple works in silicon vision sensors attempted to tackle this problem by implementing transistor based spatio-temporal filtering at the focal plane, but were noisy, suffered from mismatch, and were too complex (leading to large pixel size and lower resolution) [2]–[7]. Recent development has shown promising architectures improving on these works by proposing such a “Center Surround” (CS) based on a horizontal network of polysilicon resistors [8], [9]. This construct implements as a spatial high-pass filter, but conserves valuable pixel area.

In this work, we investigate the proposed circuit architecture and undergo design, simulation and layout work on Cadence Virtuoso Software.

## II. IMPLEMENTATION

We began with the basic DVS pixel described in [1]. Using ideas from [8] and [9], we then implemented a CS architecture. In addition, we explored two different methods of suppressing periodic “leak” events as described in [10] and [11].

### A. Circuit Design

The main circuit design decisions included how to implement the antagonistic output and transconductance element. We chose a unity gain inverter and connected the output of the standard DVS pixel source follower to the pFET so that it would remain subthreshold with a DC operating point of 1.3 V for our pixel design, selected bias configuration, and estimated photocurrent values. The pixel schematic is shown in Fig. 1.

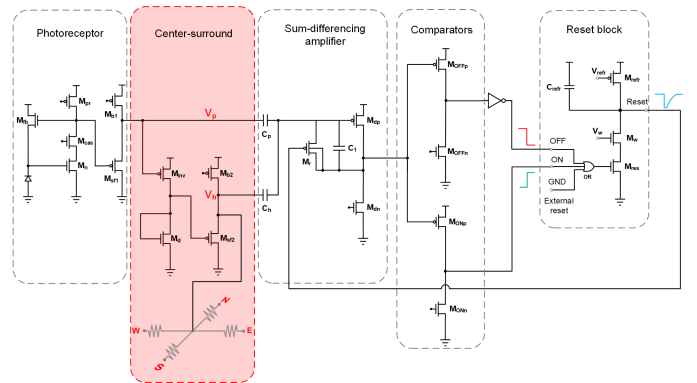


Fig. 1. CSDVS pixel schematic. The highlighted region shows the CS specific components

In order to tune the spatial extent of the antagonistic surround, a transconductance element ( $G$ ) is needed. Two different architectures were evaluated: a 2T source follower (SF) and 5T transconductance amplifier (TA). For the 2T SF configuration, there is a need to increase  $C_h$  to cancel out the  $\kappa$  gain of the SF. Comparing subthreshold transconductance of the two amplifiers designs (respectively  $G_{2T} = I_G/U_T$  and  $G_{5T} = I_G/2U_T$ ) the 5T amplifier has a disadvantage of increased power consumption for the same value of  $G$ . Hence, we concluded that 2T SF would be more efficient in terms of circuit area and power consumption.

Transistors  $M_{inv}$  and  $M_d$  form a unity gain inverter to provide antagonistic photoreceptor output, and  $M_{b2}$  and  $M_{sf2}$

act as the simple source follower (SF) transconductance element. Both transistors in the inverter remain sub-threshold for normal DC operating conditions. The SF provides an adjustable transconductance (G), enabling a tunable "space constant" (L) for the antagonistic center-surround. The output node of the SF is connected to the resistive network which links to four surrounding pixels and one input of the pixel's sum-differencing amplifier. Unsalicided polysilicon is used for the resistive elements, which has a sheet resistance of 6.7kohms/square for 1um wide 19um long resistor( $R = 140 \text{ k}\Omega$ ). Thanks to the high sheet resistance, these resistors occupy small area compatible with a reasonable pixel size.

### B. Feed forward path

We also made a minor modification to the pixel after observing an undesirable effect of feed forward pathways used in some DVS pixels. Both feed forward circuits are shown in Figure 2. The configuration increases the speed of OFF spikes by pulling down the input of the inverter which generates the spike when there is an OFF spike. However, for ON events, the change amplifier output is pulled low. Because this couples to the center surround node when the feed forward path is active, simulations showed this could generate false events in the surrounding pixels if the space constant is set too low. To alleviate this, we implemented the second circuit which includes a buffer for the ON spike path and pulls the intermediate node down when there is a spike. This way the change amplifier output is not pulled to ground and large voltage swings are not coupled to adjacent pixels via the CS network.

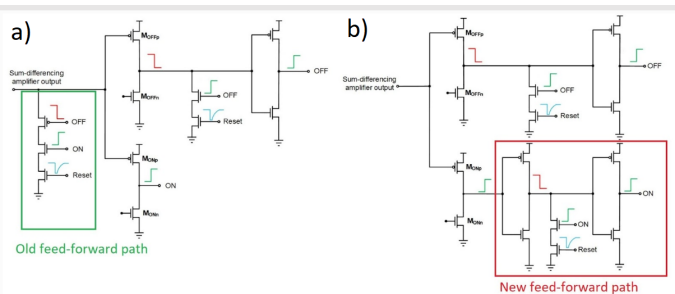


Fig. 2. a) Original feed forward circuit. b) Revised feed forward circuit.

### C. Reset Switch

We also explored methods to deal with the nagging issue in DVS cameras of so called "leak" events caused by the pFET reset switch that shorts the output and input of the change amplifier during event readout. When using a regular pFET switch that is bulked to Vdd, leakage current charges the capacitor of the amplifier and causes false ON events. Suh et al. identified GIDL as the most significant leakage current [10]. One approach to reduce the impact of leakage currents is bulking the body of the pFET to its drain [11]. This, however, requires a separate n-well, which consumes excessive pixel area. Suh et al. proposed an alternative switch consisting of

two pFET switches in parallel and a complementary source follower as shown in figure 3. The source follower forces the gate voltage of the pFET switches and the  $V_{net}$  voltage in-between to align and consequently reduces GIDL current. We explored the proposed switch to quantify its impact on leak event rate.

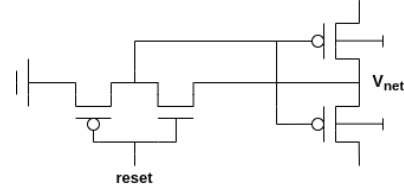


Fig. 3. Reset switch to minimize GIDL as proposed by [10].

## III. KEY RESULTS

### A. Leak Event Comparison for Reset Circuits

Performance of the reset circuit was evaluated by simulating the amplifier's behaviour for a constant DC current. If no preventative measurements for leakage current reduction are taken, our DVS pixel generated false ON events at a rate of 26 Hz. On the other hand, bulking the body of the switch to the drain of the reset transistor completely suppressed any false events. The proposed double switch significantly reduces the false event rate to 8 Hz. Nonetheless, the simulation results are based on the more reliable pFET switch that is body biased to its drain.

### B. Sine Photocurrents

Before analyzing the CS, we first simulated a single pixel's response to so a sinusoidal input signal to verify functionality. After observing expected behavior, we tested a simple 5-pixel array (one central pixel with four surrounding) and simulated two case scenarios: 1) sinusoidal photocurrent *only* to the central pixel with constant DC photocurrent to surround, and 2) identical AC photocurrent to both the central pixel and the surround. We expected to observe spikes similarly to a single pixel simulation in 1) and no spikes at all in 2). We observed the expected behaviour in 2), as well as the expected behaviour from the central pixel in 1). However, in 1), we also observed unexpected false events on the surrounding pixels: ON events in the central pixel triggered OFF events in the surround, and vice-versa. We hypothesized that this might be due to the small size of the array, effectively limiting the spatial extent to which the signal from the central pixel could dissipate. To verify this hypothesis, we simulated a 5x5 pixel array and repeated set up 1). Here, the false antagonist events in the surrounding pixels did not occur anymore, verifying our hypothesis and further proving the reliability of the center surround mechanism.

### C. Simulated Photocurrent from Video Frames

To go further in our analysis, we decided to conduct simulation with more biologically plausible data. We used video

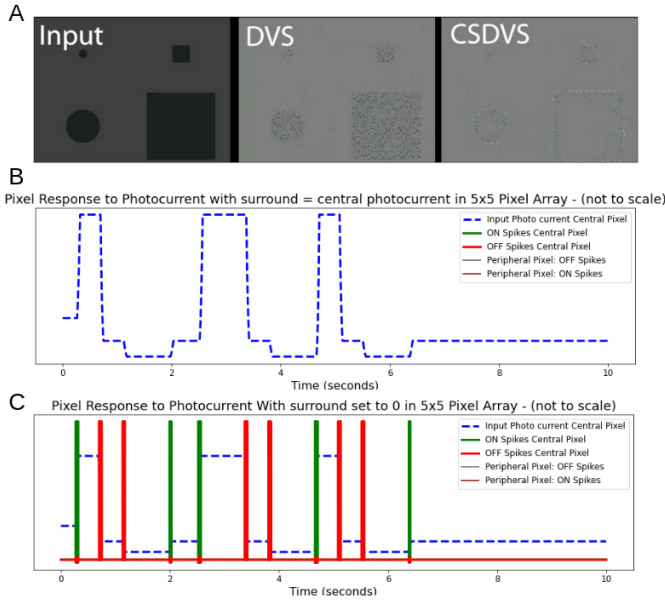


Fig. 4. A) Overview of the "Spots" video, presented in [8], which we used for real simulation. B) Simulation of pixel response in 5x5 pixel array where all pixels had matching photocurrent - no spikes were observed. C) Simulation of pixel response in 5x5 pixel array where only central pixel received changing photocurrent - spikes were observed as expected in the central pixel only.

data as presented in [8] to simulate photocurrent. Simply, we normalized pixel intensity (between 0 and 1), scaled it down by  $2.5 \times 10^{-9}$  in order to have light pixel intensity ranging between 0 and 25 nA - a range we know from the previous simulation works well to generate events. We also added a  $1 \times 10^{-9}$  constant bias, equivalent to some dark current, thus making our range of input photocurrent between 1 and 25 nA. Simulation results are illustrated in Fig 4.

#### D. Spatial Tunability

We also used the 5x5 array to analyze the ability to tune the spatial dimension of the antagonistic surround. In [12] Mead derived the space constant for a discrete 1-dimensional resistive network as

$$L = \frac{1}{\sqrt{RG}}$$

Here, L signifies the node at which the original signal has decayed to  $1/e$  of its original value, R is the horizontal resistance element, and G is the "vertical" transconductance of each node to ground. While the precise derivation is not directly applicable to a 2D resistive network, we assume the width of the spatial influence of a single pixel's transient response should be inversely proportional to both G and R. To verify the relationship, we stimulated a single pixel with a sinusoidal input and observed the comparator response at three neighboring pixels of varying distances (1, 4 and 8). R is constant, so higher bias current to the transconductor (higher G) results in a larger peak near the stimulus pixel, with a "faster" spatial decay. Results are shown in Fig. 5, validating the ability to effectively tune the space constant of the antagonistic CS.

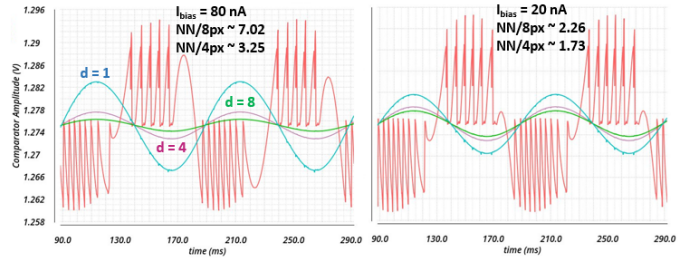


Fig. 5. Comparator amplitudes at three different pixels of varying distance from the stimulus ( $d=1, 4$  and  $8$ ) are shown for two different bias currents to transconductance element (G). With higher G, closer pixels are more strongly influenced by the stimulus, but the effect decays faster, consistent with a small space constant (L). Ratios of nearest neighbor ( $d=1$ ) comparator amplitudes to each of the other pixels are listed.

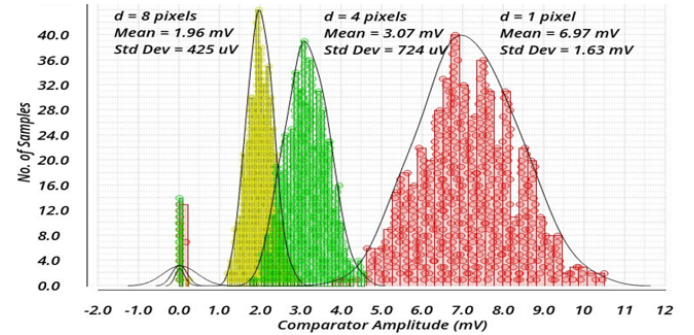


Fig. 6. A Monte Carlo analysis of 500 runs was conducted to analyze mismatch of the horizontal network.

#### E. Mismatch in Resistive Network

Another consideration for CSDVS design is mismatch in the resistive network. For the CS network to operate effectively, it is undesirable for some pixels to have a disproportionately large influence on their neighbors. In order to analyze this, we conducted a Monte Carlo analysis of 500 runs to examine how mismatch altered the resulting comparator amplitude for pixels in the vicinity of a stimulated pixel. Again, we used a sinusoidal stimulus with a DC surround, and probed the comparator nodes of surrounding pixels. Results are shown in Fig. 6. Surprisingly, in a non-trivial number of runs, we found the comparator amplitude to be nearly negligible for one of the surrounding pixels. This is evidenced by the three overlapping peaks at V 0 in Fig. 6. Upon further analysis, these outliers occurred on different runs (i.e. when the  $d=1$  pixel had 0V amplitude,  $d=4$  and  $8$  responded normally).

#### IV. LAYOUT

The layout of the CS-DVS pixel is shown in Figure 7. The pixel is  $58.5 \mu\text{m} \times 62 \mu\text{m}$ . The poly silicon lines at the right side of the pixel and at the bottom of the pixel are center surround polysilicon resistors. The pixel size could be reduced drastically by making the feedback capacitor of the change amplifier minimum size, making it minimum size will increase the number of unwanted leakage spikes as a lower capacitor will induce more voltage at the amplifier output for

TABLE I  
DEVICE DIMENSIONS AND BIASES

Transistors		W/L ratio	
$M_{fb}$		2u/2u	
$M_{pr}$		1.6u/5.6u	
$M_{cas}, M_n$		2u/1.2u	
$M_{b1}, M_{b2}, M_{refr}$		1.2u/1.2u	
$M_{sf1}, M_{inv}, M_{sf2}, M_w, M_{res}$		400n/600n	
$M_d$		600n/1.2u	
$M_r$		220n/180n	
$M_{dp}, M_{dn}, M_{OFFp}, M_{OFFn}, M_{ONp}, M_{ONn}$		1.5u/3.2u	
Biasing current	Value	Biasing voltage	Value
$I_{fb}$	50pA	$V_{DD}$	1.8V
$I_{sf}$	700pA	$V_{cas}$	1V
$I_{sf2}$	3nA	$V_w$	300mV
$I_{dn}$	1.35nA	$V_{refr}$	1.47V
$I_{OFFn}$	17nA		
$I_{ONn}$	100pA		

the same leakage current. The pixel size could also be reduced by changing the amplifying capacitors from unit size to single capacitors. This design choice is initially made to ensure the capacitor matching. But it nearly doubles the capacitor area of the pixel.

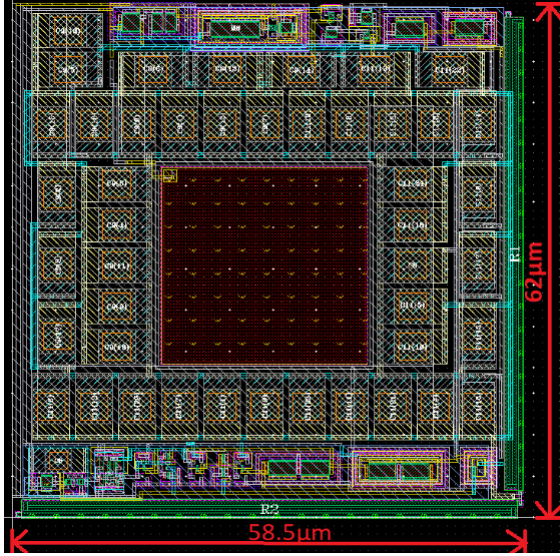


Fig. 7. Top level layout of the pixel.

### A. Post Layout Simulations

One important metric for the DVS pixel is event threshold - essentially percent change in photocurrent resulting in an event. ON mismatch before and after the parasitic extraction are shown in Fig. 8. The standard deviation did not change significantly after extraction though the mean dropped from 32.46% to 30.32%.

OFF spike threshold mismatch before and after the parasitic extraction was also examined. The standard deviation did not change significantly after extraction, though the mean increased from 26.46% to 29.52%.

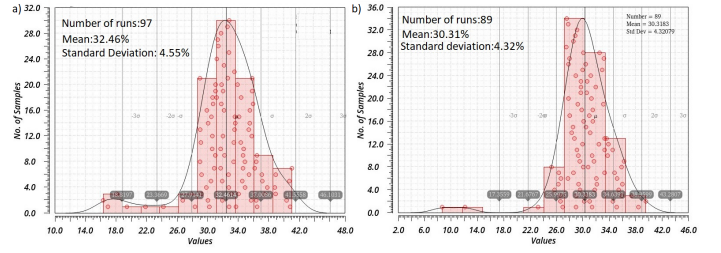


Fig. 8. a) On spike threshold mismatch before parasitic extraction. b) On spike threshold mismatch after parasitic extraction.

### B. Bias currents and power consumption

Bias current for the inverting amplifier that produces the antagonistic photoreceptor output highly depends on photocurrent. The -1 gain stage is interfaced by the pMOS transistor which means lower DC photocurrent results in higher static current and thus increased power consumption in the inverter. When the photodetector DC current is 50 pA the current in this stage is 3.71 nA, consuming 6.7 nW of power per pixel. However, if the DC photocurrent is 10 nA, the static current/power consumption is reduced to 168pA/302 pW per pixel. As a result, the proposed architecture would not be ideal under low photo current (dim lighting) conditions.

## V. CONCLUSION

In this project, we made important progress towards the realization of a practical CSDVS camera. Through circuit simulations, we validated the ability of the proposed antagonistic horizontal surround network to effectively mitigate spatially redundant information (i.e. event suppression when all pixels are stimulated with an identical or similar transient input). In addition, we showed that for realistic values of R, the space constant (L) can be tuned by proper selection of bias current to the trans-conductance element (G) without excessive current consumption. Finally, we explored some key metrics relevant to CS pixel design including threshold mismatch and surround non-uniformity.

We also identified key design considerations and challenges for the proposed center surround architecture. Static power consumption of the inverting amplifier preceding the  $V_{p-}$  node is problematic under low illumination conditions. Additionally, implementing the sum-differencing amplifier requires a design trade-off between pixel area and gain. Retaining the same gain for the positive ( $V_{p+}$  pathway) requires 2X the capacitor area in comparison to a standard DVS pixel, whereas maintaining the same pixel area results in a reduction in amplification in the change amplifier. In the DVS pixel, smaller gain has an adverse effect on sensitivity. To alleviate this, we recommend a unity gain 5T transconductance element in future designs, as opposed to the 2T source follower used in ours. This eliminates the need to offset the  $\kappa$  gain with additional capacitor area. An additional amplification stage as implemented in prior sensitive DVS cameras such as [13] and [14] could also be explored as a means to improve sensitivity without increasing the capacitor area which dominates the pixel size in our design.

## REFERENCES

- [1] Lichtsteiner, P., Posch, C. and Delbruck, T., 2008. A  $128 \times 128$  120 dB  $15 \mu$ s latency asynchronous temporal contrast vision sensor. *IEEE journal of solid-state circuits*, 43(2), pp.566-576.
- [2] Boahen, Kwabena Adu. *Retinomorph vision systems: Reverse engineering the vertebrate retina*. California Institute of Technology, 1997.
- [3] Liu, Shih-Chi. "A neuromorphic aVLSI model of global motion processing in the fly." *IEEE Transactions on circuits and systems II: analog and digital signal processing* 47, no. 12 (2000): 1458-1467.
- [4] Harrison, Reid R., and Christof Koch. "A robust analog VLSI Reichardt motion sensor." *Analog integrated circuits and signal processing* 24, no. 3 (2000): 213-229.
- [5] Delbrück, Tobi, and Shih-Chii Liu. "A silicon early visual system as a model animal." *Vision Research* 44, no. 17 (2004): 2083-2089.
- [6] Zaghoul, Kareem A., and Kwabena Boahen. "A silicon retina that reproduces signals in the optic nerve." *Journal of neural engineering* 3, no. 4 (2006): 257.
- [7] Costas-Santos, Jess, Teresa Serrano-Gotarredona, Rafael Serrano-Gotarredona, and Bernab Linares-Barranco. "A spatial contrast retina with on-chip calibration for neuromorphic spike-based AER vision systems." *IEEE Transactions on Circuits and Systems I: Regular Papers* 54, no. 7 (2007): 1444-1458.
- [8] Delbruck, Tobi, Chenghan Li, Rui Graca, and Brian McReynolds. "Utility and feasibility of a center surround event camera." *arXiv preprint arXiv:2202.13076* (2022).
- [9] Li, Chenghan. "Two-stream vision sensors." PhD diss., ETH Zurich, 2017.
- [10] Y. Suh, S. Choi, M. Ito, J. Kim, Y. Lee, J. Seo, H. Jung, D.-H. Yeo, S. Namsung, J. Bong, S. Yoo, S.-H. Shin, D. Kwon, P. Kang, S. Kim, H. Na, K. Hwang, C. Shin, J.-S. Kim, P. Park, J. Kim, H. Ryu and Y. Park, "A  $1280 \times 960$  dynamic vision sensor with a  $4.95\text{-}\mu$ pixel pitch and motion artifact minimization", *IEEE International Symposium on Circuits and Systems (ISCAS)*, October 2020
- [11] B. Son et al., "4.1 A  $640 \times 480$  dynamic vision sensor with a  $9 \mu$ m pixel and 300Meps address-event representation," *2017 IEEE International Solid-State Circuits Conference (ISSCC)*, 2017, pp. 66-67, doi: 10.1109/ISSCC.2017.7870263.
- [12] C. A. Mead, *Analog VLSI and Neural Systems*. Reading, MA: Addison Wesley, Jan. 1989, ISBN: 9780201059922. [Online]. Available: <https://www.amazon.com/Analog-VLSI-Neural-Systems-Carver/dp/0201059924>.
- [13] M. Yang, S.-C. Liu and T. Delbruck, "A Dynamic Vision Sensor With 1% Temporal Contrast Sensitivity and In-Pixel Asynchronous Delta Modulator for Event Encoding," in *IEEE Journal of Solid-State Circuits*, vol. 50, no. 9, pp. 2149-2160, Sept. 2015, doi: 10.1109/JSSC.2015.2425886.
- [14] D. P. Moeys et al., "A Sensitive Dynamic and Active Pixel Vision Sensor for Color or Neural Imaging Applications," in *IEEE Transactions on Biomedical Circuits and Systems*, vol. 12, no. 1, pp. 123-136, Feb. 2018, doi: 10.1109/TBCAS.2017.2759783.