Early Vision on the Focal-Plane with High Dynamic Range Pixels

Marko Jaklin, D. García-Lesta, P. López, V.M. Brea
Centro Singular de Investigación en Tecnoloxías Intelixentes (CiTIUS)
Universidade de Santiago de Compostela
Santiago de Compostela, Spain
email: marko.jaklin@usc.es

***Abstract*—This paper introduces a high dynamic range pixel for early vision processing. Early vision is the first stage to subsequently extract semantic information for image processing or video analytics. This paper proposes to bring said processing to the focal plane, next to a high dynamic range image sensor working on the principle of lateral overflow capacitor. This brings the benefits of processing scenes with a wide dynamic range in a power efficient manner. Circuit simulations for edge detection, as an example of early vision processing conveyed in this paper, show that our proposal meets the accuracy typically found in applications like machine vision.**

***Index Terms*—Early vision, focal-plane processing, CMOS vision sensor, high dynamic range.**

I. Introduction

Early vision comprises the first processing stages of the pipeline of image processing or video analytics to tackle tasks like object recognition or visual tracking- to a name a few-, which are an integral part of applications like video surveillance. In fact, regardless of classical handcrafted feature-based solutions [1], or the nowadays ubiquitous deep learning approach [2], the first goal in computer vision is usually to find meaningful changes in the incoming image, like edges or corners [3], [4], before going deeper in the tool chain to draw out more semantic information [5].

Early vision processing builds on higher-resolution images or features than deeper stages do, which calls for more compute capacity. This opens the door to optimizing signal processing and the hardware next to the sensor.

This paper aims to edge detection on the focal-plane as an example of hardware optimization of early vision tasks next to the sensor [6]–[8], with the extra benefit of high dynamic range. This paper conveys electrical simulations of the per-pixel processing element (PE) and their physically close photo sites as the first proof-of-concept stage towards a future CMOS vision or smart image sensor chip.

This project has received funding from the European Union's Horizon 2020 research and innovation programme under the Marie Skłodowska-Curie grant agreement No 860370, from AEI through Grant PID2021-128009OB-C32, from the Xunta de Galicia-Consellería de Cultura, Educación e Ordenación Universitaria Accreditation 2019–2022 ED431G-2019/04 and Reference Competitive Group Accreditation 2021–2024, GRC2021/48 (FEDER). We also thank the **HISILICON** programme for funding for CMOS IC fabrication.

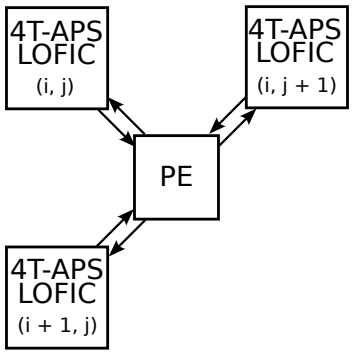

Fig. 1: Cell of a would-be 2D arrangement for focal plane processing, making up a CMOS vision sensor. Letter "i" represents row, letter "j" represents column.

II. Cell array element, dynamic range extension and edge detection algorithm

CMOS vision or smart image sensors implement the concept of focal-plane processing with arrangements of 2D cell arrays, with each cell comprising the photo sensor itself along with companion circuits physically close to them in order to run early vision tasks. In general, there are two ways to distribute processing circuitry, either, next to the photo site and in the periphery of the 2D cell array [9], the latter typically in a per-column approach, or to lay down all the processing circuitry next to photo sites, in a per-pixel approach [10]. The fully per-pixel approach increases parallelism with respect to the former one at the cost of a lower fill-factor. This obliges to share processing elements (PE) among several photo sites not to excessively worsen fill-factor. This is the option adopted in this paper.

Fig. 1 shows a cell of our array, where a PE is shared among 3 photo sites, and every photo site consists of a pinned photodiode that turns incoming photons into voltage, which, in turn, is read out by a four transistor active pixel sensor with the addition of a lateral overflow capacitor, what we call the LOFIC 4T-APS, to perform high dynamic range imaging [11]. The PE runs two types of operations, namely, the difference in intensity among neighboring pixels to provide edge detection,

Fig. 2: Numerical simulations of our edge detection algorithm applied to the original grayscale image shown with increasing threshold T_{TH} values.

and all the operations required to perform per-pixel high dynamic range.

The LOFIC technique extends the dynamic range by storing overflown electrons coming out of the pinned photodiode when this reaches saturation, which means that the number of impinging photons and their corresponding charges generated on the photodiode exceed its full well capacity, i.e., the sensor is capturing a scene with a higher intensity than what is physically feasible. If the incoming scene does not go beyond saturation, the LOFIC tecnhique does not act, and, thus, the LOFIC 4T-APS works as a conventional 4T-APS sensor.

The implementation of the LOFIC 4T-APS with per-photo site circuitry degrades fill-factor [12], which leads us to keep the edge detection algorithm simple. Said algorithm is the threshold of the difference between the image intensity of two neighboring pixels along horizontal and vertical directions. If the difference exceeds the threshold, the result is an edge.

The equation below describes said algorithm mathematically:

$$f(x_{i,j}, x_{i+1,j}, x_{i,j+1}) = \begin{cases} 1, & \text{if } |x_{i,j} - x_{i+1,j}| \geq T_{TH} \\ 1, & \text{else if } |x_{i,j} - x_{i,j+1}| \geq T_{TH} \\ 0, & \text{else} \end{cases} \tag{1}$$

$x_{i,j}$ represents the value of the pixel on the $i-th$ row and $j-th$ column. $x_{i+1,j}$ is its row neighbor, and $x_{i,j+1}$ is its column neighbor. If the difference among neighboring pixels along the x and y axes surpasses the user-defined threshold T_{TH}, either with a positive or a negative value, it is interpreted as an edge, returning a white pixel. This algorithm reduces the resolution of the image by one in each dimension, as each column or row contains one less PE than photodiodes, so that there is a one-pixel padding. Finally, numerical simulations to illustrate how our LOFIC 4T-APS would run edge detection are displayed on Fig. 2, showing the effect of different T_{TH} values. From the signal processing perspective, low T_{TH} levels make the system more sensitive to edges, which might come with the undesirable effect of bringing about not meaningful edges like artifacts from backgrounds with clutter. From the hardware perspective, there is a minimum T_{TH} value in hardware due to the non-idealities of the LOFIC 4T-APS and the PE.

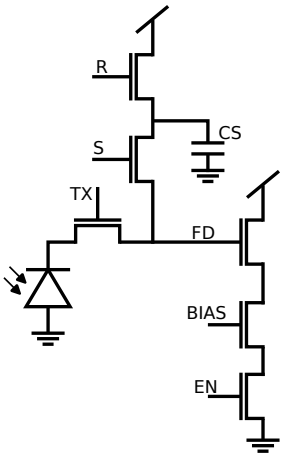

Fig. 3: LOFIC 4T-APS pixels.

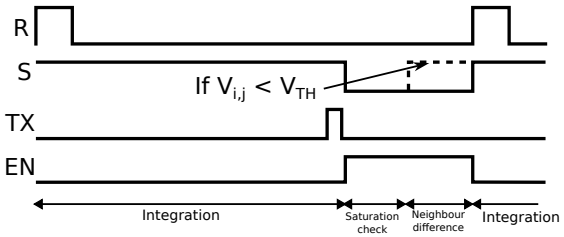

Fig. 4: Timing diagram of the LOFIC 4T-APS.

A. Active pixel sensor

Fig. 3 and Fig. 4 show the LOFIC 4T-APS and its corresponding timing diagram, respectively. The LOFIC structure was first introduced in [11] and is used in some other image sensors [13], [14].

The cycle of operations of our pixel starts with the reset of the FD node and the CS capacitor by pulling the signal R high while S is also high. During integration, S stays high to keep the way open for any overflown electrons. Signal S is generated by the combination of logic inside the PE and

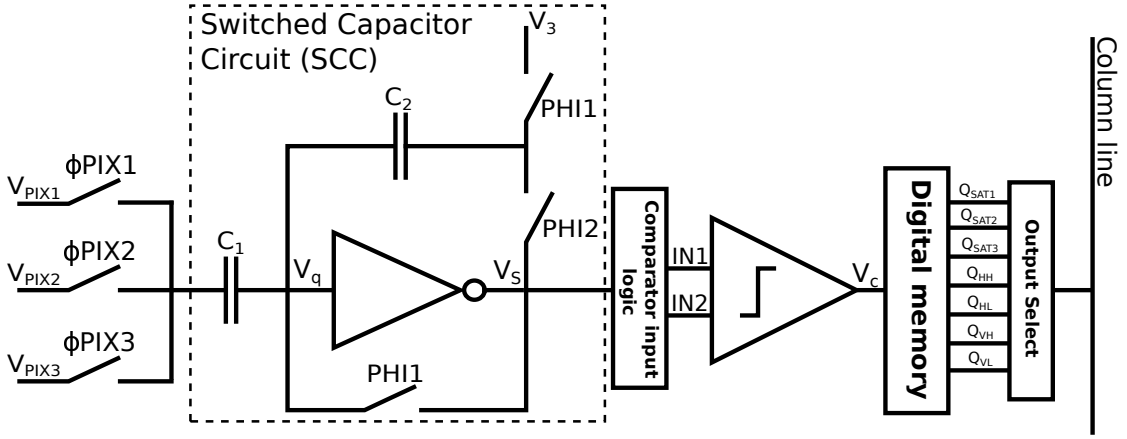

Fig. 5: PE schematics with its different components.

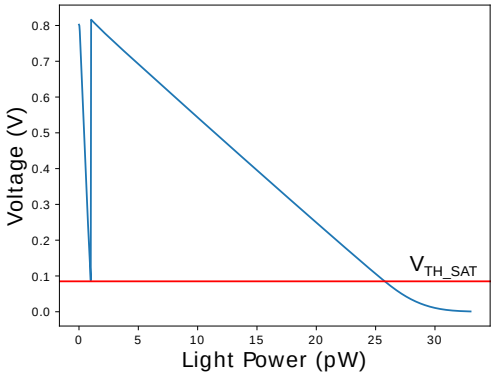

Fig. 6: Simulation showing the dynamic range extension. When signal generated by non-saturated electrons falls to V_{TH_SAT} the pixel adds the electrons stored on CS capacitor.

an external signal. S stays high until the end of integration, when it is pulled low to isolate FD from CS. At the end of integration TX, is pulsed high, and the photodiode transfers its electrons to the FD, generating a non-saturated signal.

The dynamic range extension is simulated in Fig. 6. This simulation has been done for a long integration time, i.e., the time along which the incoming photons are turned into voltage at node FD of our LOFIC 4T-APS (see Fig. 3). The voltage at node FD decreases with the incoming light and/or the integration time, approximately following a straight line with a negative slope. The simulation in Fig. 6 shows the output voltage of the source follower (SF), the output of the transistor driven by FD, against the power of the incoming light (x-axis) during a long enough time as to reach saturation with a very low input power, i.e. the full well capacity of the 4T-APS has been exceeded, marked in Fig. 6 as V_{TH_SAT}, and as a consequence, the pixel enters the second region of operation, producing a straight line response with a much smaller slope

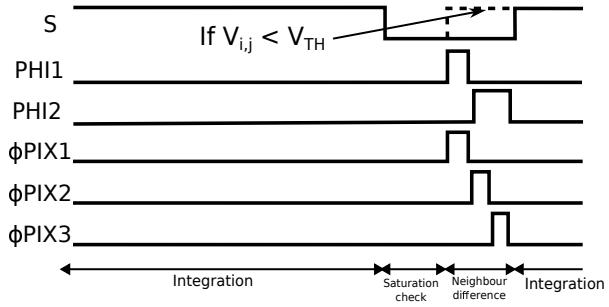

Fig. 7: Processing unit timing diagram.

than that of the first region of operation. This, in turn, extends the region of operation of our pixel. Also, as seen, when the threshold is surpassed, signal S goes high to connect FD with CS, extending the dynamic range at the cost of lowering the sensitivity due to the larger capacitance of the $FD+CS$ node.

The edge detection algorithm is run after checking the saturation of the 3 LOFIC 4T-APS pixels next to a PE within a cell (see Fig. 1), which in our circuitry is done by calculating the voltage difference between neighboring pixels and its corresponding comparison with user-defined thresholds. As shown in Fig. 4 this occurs when signal EN goes high after signal TX goes low, enabling the SF driven by the FD node. When the neighbor difference is completed, R is pulled high, and the cycle starts anew.

B. Processing Element (PE)

The schematics of the PE with its components is shown on Fig. 5.

1) Switched Capacitor Circuit: The switched capacitor circuit (SCC) calculates the difference between neighbors. Its timing diagram is in Fig. 7. It operates in two phases, PHI1 and PHI2. During PHI1, the value of (i,j) pixels is stored on C1, and V_q is set. During PHI2, the voltages of the neighboring pixels (i+1,j) and (i,j+1) arrive at C1. Signals

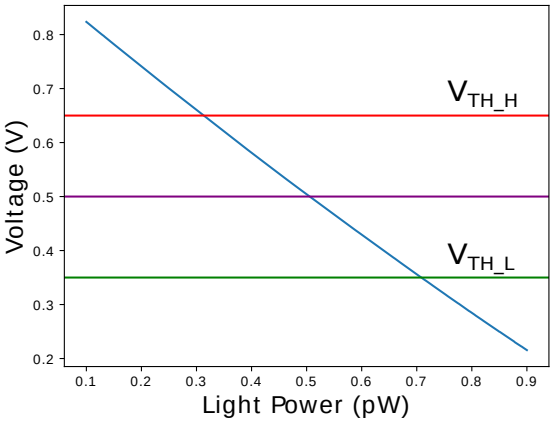

Fig. 8: Simulation of the neighbor difference. In this case, $V_3 = 0.5$V and is represented by a purple line. The comparator checks the difference between the V_{TH_H} and V_{TH_L} thresholds. If the difference is higher than V_{TH_H} or lower than V_{TH_L}, it is interpreted as an edge.

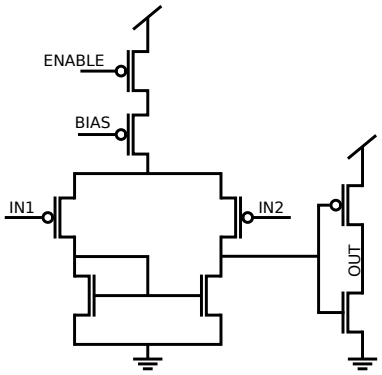

Fig. 9: Comparator schematic.

$\Phi PIX1$, $\Phi PIX2$, and $\Phi PIX3$ control the switches that connect C1 to the SF's of the pixels. $\Phi PIX1$ connects to the SF of (i,j) pixel, $\Phi PIX2$ to the (i, j+1) and $\Phi PIX3$ pixel and (i+1,j pixel). Using the conservation of charge law for the V_q node and by ignoring the parasitics, it can be shown that during PHI2, the output of the SCC for the vertical neighbor is:

$$\Delta \mathrm{V_{ver}} = \frac{\mathrm{C_1}}{\mathrm{C_2}}(\mathrm{V_{i,j}} - \mathrm{V_{i+1,j}}) + \mathrm{V_3} \tag{2}$$

where V_3 is an offset defined by the user to adapt voltage ranges of different circuits along the data path.

Similarly, the output along the horizontal axis is:

$$\Delta \mathrm{V_{hor}} = \frac{\mathrm{C_1}}{\mathrm{C_2}}(\mathrm{V_{i,j}} - \mathrm{V_{i,j+1}}) + \mathrm{V_3} \tag{3}$$

Electrical simulations of the neighbor difference are shown in Fig. 8. The simulations show the difference in intensity between two pixels, with one of them receiving a constant light power of 0.5 pW, while the other one changes linearly. The voltage along the y-axis is the voltage difference measured at the output node of SCC, i.e. V_s, (see Fig. 5), which, as seen, shows very linear behavior.

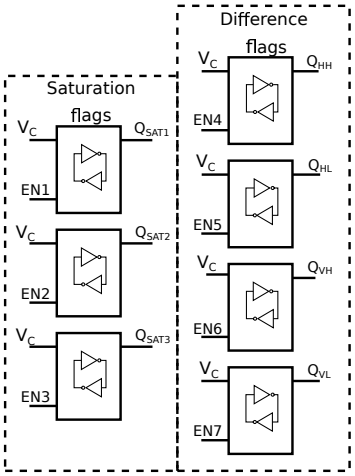

Fig. 10: Digital memory.

2) Comparator and its Input Logic: Comparator input logic is tasked with placing voltages from different sources on the correct inputs of the comparator at the right time. While checking for saturation, the threshold V_{TH} is on IN2 of the comparator, and on IN1 are the analog pixel values $V_{i,j}$, $V_{i+1,j}$, and $V_{i,j+1}$. During the neighbor difference, the input IN2 takes the voltage values of the threshold for edge detection, while IN1 is connected to the output of SCC node V_S.

Our comparator is a 5T-OTA followed by an inverter with PMOS as input transistors, as shown in Fig. 9. It checks the saturation and determines if the difference between neighbors is big enough to be interpreted as an edge.

3) Digital Memory: Digital memory stores the results of saturation checks, and difference checks: Digital memory shown in Fig. 10 stores the results.

Regarding the saturation checks, if the $V_{i,j}, V_{i+1,j}, and V_{i,j+1}$ fall below V_{THSAT}, the Q_{SAT1}, Q_{SAT2}, and Q_{SAT3} store 1. Otherwise, they store 0.

The Q_{HH}, Q_{HL}, Q_{VH} and Q_{VL} are determined during the difference phase of the operation. If $\Delta V_{VER} > V_{TH_H}$ Q_{HH} stores 1, and if $\Delta V_{HOR} > V_{TH_H}$, Q_{HL} stores 1. Otherwise, Q_{HH} and Q_{VH} store zero. If $\Delta V_{VER} < V_{TH\_L}$ Q_{HH} Q_{HL} stores 1, and if $\Delta V_{VER} < V_{TH\_L}$ Q_{VL} stores 1. Otherwise, Q_{VL} and Q_{HL} store zero. This logic is shown in Table I.

As already mentioned, the dynamic range extension introduces non-linearity due to the different sensitivities of the FD and $FD + CS$ nodes. If the signals of neighboring pixels are of different sensitivities, the difference between them will not represent a meaningful result. Due to this, we use another logic with the saturation flags presented in Table II. If all three pixels do not have all equal saturation flags, Q_{SAT1}, Q_{SAT2}, and

TABLE I: Edge detection logic.

Flags				Edges		
Q_{HH}	Q_{HL}	Q_{VH}	D_{VL}	HOR EDGE	VER EDGE	EDGE
0	0	0	0	0	0	0
1	0	0	0	1	0	1
1	0	1	0	1	1	1
0	0	0	1	0	1	1
0	1	0	0	1	0	1

TABLE II: Edge detection logic when saturation flags are not equal. In case of "?" logic from Table I is used

Saturation flags			Edges		
Q_{SAT1}	Q_{SAT2}	Q_{SAT3}	HOR EDGE	VER EDGE	EDGE
0	0	0	?	?	?
1	1	1	?	?	?
0	1	0	1	?	1
1	0	0	1	0	1
0	0	1	?	1	1
1	0	1	1	?	1
1	1	0	?	1	1
0	1	1	1	1	1

Q_{SAT3}, the edge is automatically declared regardless of the difference result. Thus, in total, seven bits of information are required for edge detection.

C. Mismatch analysis

Numerical simulations do not provide edges if neighboring pixels feature the same intensity. In contrast, circuit realizations do it due to mismatch, i.e., two supposedly identical neighboring circuits do not yield the same outputs, bringing about additional edges not present in the scene. This obliges us to set a minimum voltage difference of $V_{TH\_H} - V_{TH\_L}$.

We have simulated mismatch among two neighboring circuits in Fig. 11, obtained from 1,000 Monte Carlo (MC) simulations. This figure plots the percentage of false edges from said 1,000 MC due to mismatch vs $\Delta V_{TH} = V_{TH\_H} - V_{TH\_L}$ (thresholds can be seen in Fig. 8), for a given input light power. In our case, 80 mV would lead to zero artificial edges coming from physical imperfections. This means a minimum intensity difference level of 0.008 pW for the nonsaturated signal and 0.232 pW for the saturated signal. Also, 80 mV with a voltage range of 700 mV is around 30 levels of gray scale, which is acceptable in typical applications of machine vision.

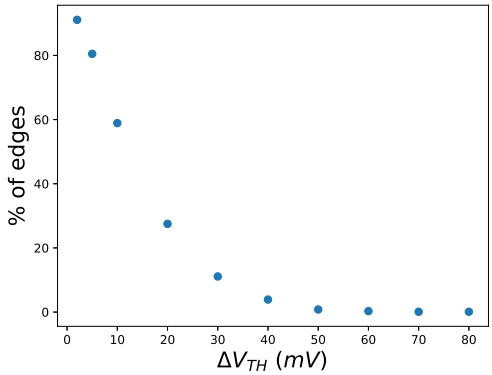

Fig. 11: Mismatch analysis when two neighbouring pixels receive same light intensity.

III. Conclusion

This paper has introduced a high dynamic range pixel for early vision processing on the focal plane, i.e., next to the sensor, showing it to be accurate enough for edge detection. This the first step towards a smart image sensor, where good figures of merit in terms of area and power consumption are expected thanks to the focal plane processing principle.

References

[1] D. Lowe, "The Silicon Retina," *International Journal of Computer Vision*, vol. 60, pp. 91–110, 2004.

[2] G. H. Y. LeCun, Y. Bengio, "Deep Learning," *Nature*, vol. 521, pp. 436–444, 2015.

[3] P. Ganesan, V. Rajini, and R. I. Rajkumar, "Segmentation and Edge Detection of Color Images Using CIELAB Color Space and Edge Detectors," in *INTERACT-2010*, 2010, pp. 393–397.

[4] T. T. Nguyen and J. W. Jeon, "Camera Auto-Exposing and Auto-Focusing for Edge-Related Applications Using a Particle Filter," in *2010 IEEE/RSJ International Conference on Intelligent Robots and Systems*, 2010, pp. 1177–1182.

[5] S. Ren, K. He, R. Girshick, and J. Sun, "Faster r-cnn: Towards real-time object detection with region proposal networks," *Advances in neural information processing systems*, vol. 28, 2015.

[6] C. Soell, L. Shi, J. Roeber, M. Reichenbach, R. Weigel, and A. Hagelauer, "Low-Power Analog Smart Camera Sensor for Edge Detection," in *2016 IEEE International Conference on Image Processing (ICIP)*, 2016, pp. 4408–4412.

[7] M.-J. Park and H.-J. Kim, "A Real-Time Edge-Detection CMOS Image Sensor for Machine Vision Applications," *IEEE Sensors Journal*, vol. 23, no. 9, pp. 9254–9261, 2023.

[8] C. Lee, W. Chao, S. Lee, J. Hone, A. Molnar, and S. H. Hong, "A Low-Power Edge Detection Image Sensor Based on Parallel Digital Pulse Computation," *IEEE Transactions on Circuits and Systems II: Express Briefs*, vol. 62, no. 11, pp. 1043–1047, 2015.

[9] X. Zhong *et al.*, "A Fully Dynamic Multi-Mode CMOS Vision Sensor With Mixed-Signal Cooperative Motion Sensing and Object Segmentation for Adaptive Edge Computing," *IEEE Journal of Solid-State Circuits*, vol. 55, no. 6, pp. 1684–1697, 2020.

[10] D. García-Lesta, D. Cabello, P. López, and V. M. Brea, "HOPBAS10K: A 98 × 98 Pixels CMOS Vision Sensor for Background Subtraction," *IEEE Sensors Journal*, vol. 24, no. 7, pp. 11 927–11 935, 2024.

[11] N. Akahane *et al.*, "A Sensitivity and Linearity Improvement of a 100 dB Dynamic Range CMOS Image Sensor using a Lateral Overflow Integration Capacitor," in *Digest of Technical Papers. 2005 Symposium on VLSI Circuits, 2005.*, 2005, pp. 62–65.

[12] M. Jaklin, D. García-Lesta, P. López, and V. M. Brea, "Global shutter CMOS vision sensors and event cameras for on-chip dynamic information," *International Journal of Circuit Theory and Applications*, vol. n/a, no. n/a. [Online]. Available: https://onlinelibrary.wiley.com/doi/abs/10.1002/cta.3925

[13] M. Jaklin, D. García-Lesta, P. López, and V. Brea, "CDS Free Frame Differencing Event Vision Pixel with Lateral Overflow Capacitor for Dynamic Range Extension," in *2023 30th IEEE International Conference on Electronics, Circuits and Systems (ICECS)*, 2023, pp. 1–4.

[14] M. Jaklin, D. García-Lesta, P. López, and V. M. Brea, "Global shutter CMOS vision sensors and event cameras for on-chip dynamic information," *International Journal of Circuit Theory and Applications*, vol. n/a, no. n/a. [Online]. Available: https://onlinelibrary.wiley.com/doi/abs/10.1002/cta.3925