Special Research Session: A Quantitative Approach to Exploring Novel Image Sensor Architecture Towards Autonomous Edge Vision

Han Yan*
hyan20@ur.rochester.edu
University of Rochester
Rochester, New York, USA

Yifei Zou*
zou.yife@northeastern.edu
Northeastern University
Boston, Massachusetts, USA

Weikai Lin
wlin33@ur.rochester.edu
University of Rochester
Rochester, New York, USA

Yuhao Zhu
yzhu@rochester.edu
University of Rochester
Rochester, New York, USA

Xuan Zhang
xuan.zhang@northeastern.edu
Northeastern University
Boston, Massachusetts, USA

Abstract

Autonomous edge machine vision requires image sensors that simultaneously advance processing autonomy and energy autonomy. This paper introduces a quantitative modeling framework that spans conventional CMOS imagers and emerging designs—event-based sensors (DVS), coded-exposure sensors, and self-powered sensors with energy-harvesting pixels. The framework unifies optical, circuit, and algorithmic modeling to capture how architectural choices in pixel structures, readout pipelines, and sensing mechanisms propagate to power, latency, noise, and task-level performance. By enabling rapid, physically grounded exploration and comparative analyses, our results reveal key trade-offs that determine when unconventional sensor architectures meaningfully enhance edge intelligence. This framework offers a principled foundation for designing next-generation energy-aware, autonomous vision systems.

1 Introduction

Edge vision systems perform image sensing and processing directly on edge devices, enabling real-time perception while reducing the data required to be stored locally or transmitted to the cloud [3]. They are widely used in applications such as robotics, environmental monitoring, and smart infrastructure. However, many deployments require these systems to operate for long periods with minimal maintenance under strict energy and resource constraints.

These constraints call for autonomous edge vision — sensing and visual processing are sustained locally. In particular, such systems must achieve **energy autonomy**, where harvested energy is sufficient for system operation, and **processing autonomy**, where visual data can be processed locally under task constraints. However, the diversity of sensor architectures and task requirements makes it difficult to directly assess whether such autonomy can be achieved. This necessitates a first-principles quantitative framework to model system metrics from key design parameters.

Modeling Challenges. Unlike conventional systems with dedicated energy harvesters, vision systems use pixels that can act as both imagers and energy harvesters [23]. Sensor design therefore affects both energy supply and consumption, and becomes a new design space dimension. This is particularly relevant in emerging architectures such as coded exposure [10, 16], where some pixels remain idle during sensing and can harvest energy.

These effects propagate to downstream task performance. Sensor design determines both the available energy budget and the sensed data quality, which affects the task performance, e.g., accuracy and latency. This introduces a trade-off between sensor architecture design and task performance that must be captured in modeling.

Existing Work and Limitations. Existing work falls into two categories: autonomous computing systems model [11] and image sensor model [5, 7, 9, 15, 19, 20, 26]. However, neither account for the role of sensing in autonomous vision systems, where it determines both energy supply through pixel-level harvesting and data quality for downstream tasks.

Proposed Methods. To address this gap, we develop a unified quantitative modeling framework for autonomous edge vision systems. We treat energy autonomy as energy-balance problem and jointly model harvested energy and energy consumption. For processing autonomy, we model computation energy and sensing quality, e.g., FPS, resolution, and image noise, to evaluate downstream task performance accordingly.

Four representative sensor architectures are analyzed here: conventional CMOS image sensors (CIS) [13, 24] and three emerging designs: coded exposure (CE) sensors [17, 18, 25], dynamic vision sensor (DVS) [14, 21], and self-powered image sensors [12].

The paper is organized as follows. Sec. 2 presents our quantitative modeling methodology. Sec. 3 shows the model details. Sec. 4 examines novel sensor designs in representative application scenarios. Sec. 5 explores the design space of different sensors. We conclude the paper in Sec. 6.

2 System Abstraction

Our method directly links sensor design parameters to task performance by modeling their impact on energy consumption and signal quality.

*Both authors contributed equally to this research.

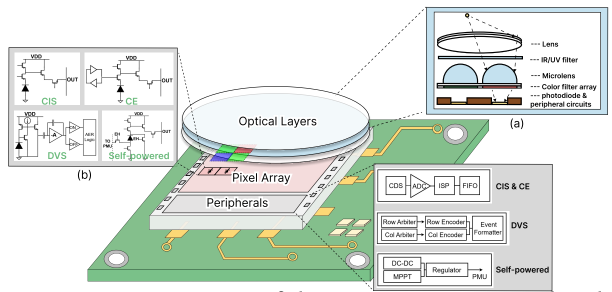

Fig. 1. System overview of the vision system considered in this work. (a) Optical layers. (b) Pixel array designs across sensor types. (c) Peripheral circuitry across sensor types.

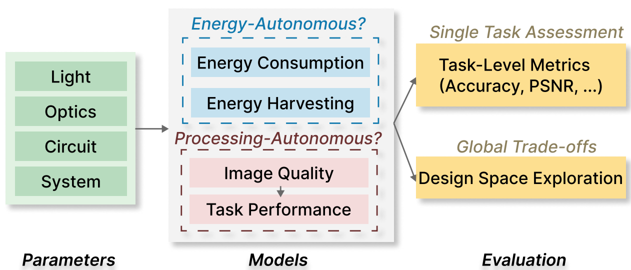

Fig. 2. Quantitative analysis framework for autonomous edge vision systems.

The system-level abstraction is illustrated in Fig. 1, consisting of sensor optical layer, pixel array, circuit peripherals, and circuit board. We instantiate this system with four sensor types, as shown in Fig. 1 (b)-(c): **(1) Conventional CIS**, a traditional frame-based image sensor with synchronous readout; **(2) CE CIS**, CIS with in-pixel storage and control for selective exposure; **(3) DVS**, where pixels asynchronously respond to intensity changes and output sparse event streams; **(4) Self-powered image sensors**, CIS with pixels that switch between imaging and energy harvesting across space and time.

Based on this abstraction, we can quantitatively analyze the system. As shown in Fig. 2, the design is parameterized across hierarchy levels, from component-level parameters to architecture and system settings. Each component uses a parameterized model to estimate circuit-level performance, and these outputs are composed into system-level energy and task models for autonomy assessment. These models support both single-design evaluation and design space exploration.

3 Analytical Modeling of Image Sensors

3.1 Modeling Energy Consumption

As more sensors integrate in-sensor/on-sensor processing, we model energy consumption by decomposing it into sensing and computing energy.

Sensing Energy. In our model, the energy consumption of conventional CIS, CE CIS, and DVS is derived from static bias energy and dynamic switching energy across the sensing and readout circuitry. The energy of CIS and CE CIS is modeled by integrating the power over a fixed frame time, while CE CIS additionally accounts for in-pixel control logic:

$$E_{\text{frame,CIS}} = T_{\text{frame}} \sum_{i=1}^{N_{\text{sub}}} \left(P_{\text{static},i} + P_{\text{dynamic},i}\right), \tag{1}$$

$$E_{\text{frame,CE}} = E_{\text{frame,CIS}} + E_{\text{CE\_overhead}}. \tag{2}$$

DVS power includes a constant static term and a dynamic term proportional to the pixel array event rate:

$$P_{\text{total,DVS}} = P_{\text{static,DVS}} + F_{\text{total}} E_{\text{event}}. \tag{3}$$

In-sensor/On-sensor Computation Energy. Energy consumption for computation can be broken down into two components: arithmetic operations and memory access at different memory hierarchy levels. The total energy is obtained by summing the product of the number of operations or accesses and their corresponding per-event energy cost:

$$\begin{aligned} E_{compute} &= E_{\text{comp}} + E_{\text{data}} \\ &= \sum_{\text{OP}\in\{\text{MAC,ADD,MULT},\ldots\}} N_{\text{OP}}\, E_{\text{OP}} + \sum_{\ell=1}^{L} N_{\text{acc},\ell}\, E_{\text{mem},\ell}, \end{aligned} \tag{4}$$

where E_{OP} and $E_{\text{mem},\ell}$ denote per-operation and per-access energy, and N_{OP} and $N_{\text{acc},\ell}$ denote their corresponding counts.

3.2 Modeling Harvested Energy

We consider a self-powered vision system with two energy sources: a solar panel (SP) and energy-harvesting pixels (EHPs) within the image sensor. Pixels harvest energy when not used for imaging, and the harvested energy powers the sensing and computation pipeline.

The energy harvested by EHPs and SP are modeled as the harvesting power multiplied by the harvesting time:

$$Q_{\text{EH\_EHP}} = R \cdot \left(\frac{1}{\text{FPS}} - SR \cdot T_{\text{exp}}\right) \cdot P_{\text{EH,CIS}}, \tag{5}$$

$$Q_{EH\_SP} = \frac{P_{EH,SP}}{\text{FPS}}. \tag{6}$$

The effective EH time in EHPs depends on resolution R, frame rate FPS, exposure time T_{exp}, and the fraction of imaging pixels (spatial sampling rate, SR). In contrast, SP can harvest energy continuously over the entire frame.

The harvesting power of each EHP and SP is given by:

$$P_{\text{EH},x} = \eta_x \cdot E_{\text{light}} \cdot \text{Area}_x, \quad x \in \{\text{CIS, SP}\}, \tag{7}$$

where η_x is light-to-electricity conversion efficiency, E_{light} is incident light irradiance, and $Area_x$ is light-sensitive area.

3.3 Modeling Task Performance

In vision systems, task performance depends on the quality of the sensed input, which is largely determined by sensor noise. Therefore, we focus on modeling noise characteristics at the sensor output.

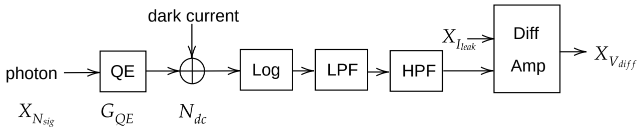

Fig. 3. DVS signal propagation pipeline.

CIS and CE CIS noise. For conventional and CE CIS, we consider four primary noise sources: photodiode shot noise, dark-current noise, read noise from the analog readout chain, and ADC quantization noise. These sources are grouped into photon-related shot noise and circuit-originated readout noise, and combined in root-sum-square (RSS) form:

$$\sigma_{\text{shot}} = \sqrt{\sigma^2_{\text{shot,PD}} + \sigma^2_{\text{dark}}}, \tag{8}$$

$$\sigma_{\text{read}} = \sqrt{\sum_{i=1}^{N} \sigma^2_{\text{read},i} + \sigma^2_{\text{ADC}}}, \tag{9}$$

$$\sigma_{\text{total}} = \sqrt{\sigma^2_{\text{shot}} + \sigma^2_{\text{read}}}, \tag{10}$$

where N denotes the number of readout stages that contribute to thermal noise.

DVS noise. We model DVS noise by decomposing the sensing pipeline into stages, identifying noise sources, and propagating them to the output. Fig. 3 illustrates this sensing pipeline. Photons N_{sig} are converted to electrical signal through photodiode quantum efficiency (QE), with dark current introduced. The signal then passes through a logarithmic photoreceptor acting as a low-pass filter (LPF), followed by a capacitive differencing circuit introducing a high-pass filter (HPF). The HPF output is affected by leakage current I_{leak}, which induces voltage drift before producing the differencing voltage V_{diff}. The voltage is then compared with a threshold θ to trigger events, followed by a refractory period that limits the event rate.

To analyze noise, we treat the input signal and intermediate quantities as stochastic processes (denoted with X). Photon arrivals and leakage current propagate through the pipeline, yielding a stochastic differencing voltage:

$$X_{V_{diff}}(t) = f_{DVS}(X_{N_{sig}}(t), G_{QE}, N_{dc}, X_{I_{leak}}), \tag{11}$$

where f_{DVS} represents the DVS sensing pipeline. Event generation occurs when the differencing voltage exceeds a pixel-specific threshold θ_i:

$$X_{\text{event},i}(t) = \mathbf{1}\left(X_{V_{\text{diff}},i}(t) > \theta_i\right). \tag{12}$$

This physics-based formulation enables noise prediction from circuit parameters and input signals.

4 Application-Driven Case Studies

4.1 3-D Stacked Perovskite Color Sensor

Recently, 3D-stacked image sensors have gained attention as a promising alternative to conventional color imaging architectures due to their vertical color-sensitive stacking and high photon utilization. To analyze the benefits of the 3-D stacked design, we compared a 3D-stacked perovskite [27] CIS with a standard Bayer RGB CIS [2], shown in Fig. 4.

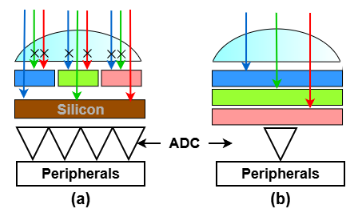

Fig. 4. (a) 2D Bayer RGB sensor optical front-end; (b) 3D-stacked perovskite sensor optical front-end.

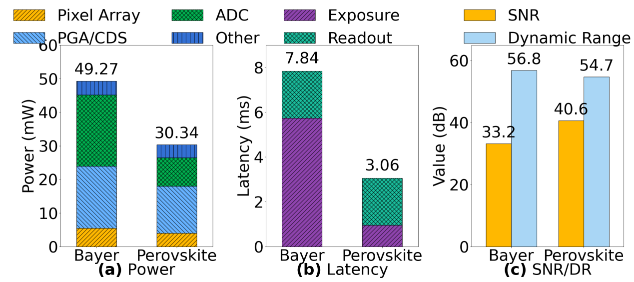

Fig. 5. Comparison of the perovskite and Bayer designs in terms of (a) power, (b) latency, and (c) SNR/DR.

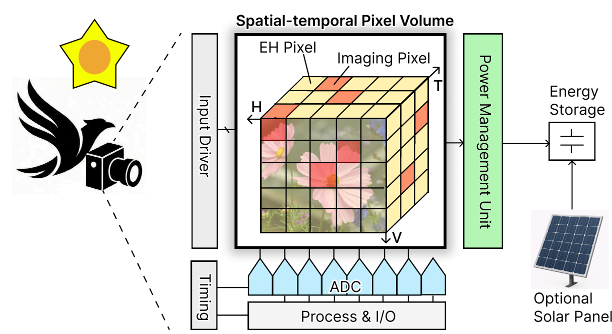

Fig. 6. Self-powered bird-borne camera with compressive sensing capability and EHPs.

As shown in Fig. 5(a), the perovskite architecture reduces total power by 38% by reducing number of effective readout paths. Its higher external quantum efficiency (EQE) and CFA-free structure also shorten the maximum exposure time to reach saturation (Fig. 5(b)), benefiting high-speed imaging.

For visual performance, the simulated outputs were evaluated for Signal-to-Noise Ratio (SNR) and Dynamic Range (DR). At a fixed 0.9 ms exposure, the perovskite sensor achieves 40.6 dB SNR, a 22% improvement over the Bayer design, indicating better low-light sensitivity (Fig. 5(c)). Although its lower capacitance density slightly reduces dynamic range, it remains a strong alternative overall.

4.2 Self-powered Bird-borne Compressive Camera

Bird-borne cameras capture visual data inaccessible to conventional systems, but their design is constrained by stringent weight, power, and storage limits [6]. These constraints make full-resolution, high-frame-rate imaging impractical.

A natural approach is to sample only a subset of pixels and reconstruct images off-sensor [4], leaving remaining pixels available for energy harvesting. As shown in Fig. 6, this leads to a sparsely sampled imaging system where pixels can be configured as either EHPs or imaging pixels across space and time. However, the benefit of such a design remains unclear; our framework evaluates it. Additional energy harvested by EHPs enables higher sampling rates, which may improve reconstruction quality.

We consider a design with 3.5 g total system weight as an example. Reconstruction quality is measured using Peak Signal-to-Noise Ratio (PSNR) between the reconstructed and ground-truth images. The results are shown in Fig. 7. With EHPs, the sampling rate doubles (from 4% to 8%), resulting in a PSNR increase from 20.03 dB to 21.78 dB (+1.75 dB).

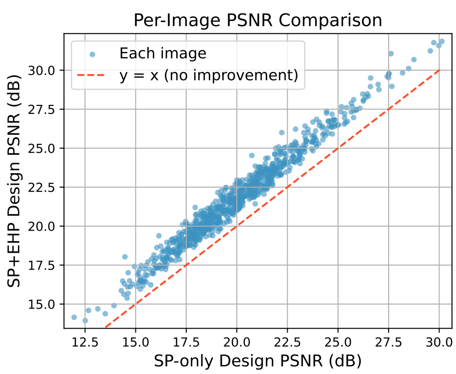

Fig. 7. PSNR comparison on images from DIV2K dataset [1].

5 System-Level Exploration

5.1 Benefits of EHPs in Self-Powered Image Sensors

To evaluate the benefits of EHPs systematically, we perform a design space exploration (DSE) to characterize how much benefit EHPs can provide across different system configurations, and how this benefit varies with design parameters.

In DSE, we compare a conventional solar-panel-powered system (SP-only) with a system that additionally harvests energy from pixels (SP+EHP). As shown in Fig. 8, the design space spans spatial resolution (R), sampling rate, and system weight, where each point represents a design configuration and the color indicates the relative FPS gain.

The results indicate that adopting EHPs are not universally beneficial; their impact depends on system weight allocation, with gains increasing when more weight is assigned to the image sensor. Such configurations may arise in vision platforms with strict weight budget but moderate imaging requirements. Notably, the bird-borne camera evaluated earlier is one such instance.

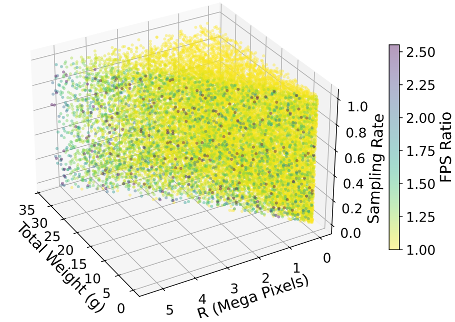

Fig. 8. FPS gain from SP-only system to SP+EHP system.

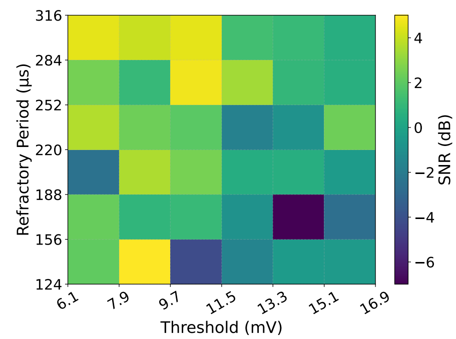

Fig. 9. SNR across different threshold and refractory settings.

5.2 Noise Sensitivity in Event Camera

Event cameras have two key tunable parameters: threshold and refractory period. Increasing either parameter suppresses noise but also attenuates signal events, creating a trade-off in SNR that is not directly observable. Using our quantitative framework, we can systematically explore this design space and identify noise-optimal parameter choice.

We explore this design space by sweeping these two parameters and evaluating each configuration using SNR. SNR is computed by comparing the noisy event trace generated by the noise model with corresponding ideal event trace. Fig. 9 shows the result. The noise-optimal design lies at the maximum SNR in the design space.

Additionally, two trends emerge: increasing the refractory period generally improves SNR, while increasing the threshold tends to reduce SNR. This is consistent with prior observations: a longer refractory period suppresses temporally clustered noise events [22], whereas a higher threshold suppresses both noise and true events [8].

6 Conclusion

In this paper, we proposed a quantitative framework for analyzing autonomous edge vision systems. The framework jointly models energy consumption, energy harvesting, and task performance, spanning optical, circuit, and algorithmic levels. Using this approach, we evaluate emerging image sensor architectures under application-specific scenarios and across the global design space.

References

[1] Eirikur Agustsson and Radu Timofte. 2017. Ntire 2017 challenge on single image super-resolution: Dataset and study. In *Proceedings of the IEEE conference on computer vision and pattern recognition workshops*. 126–135.

[2] Bryce E. Bayer. 1976. Color Imaging Array. https://patents.google.com/patent/US3971065A

[3] Pietro Bonazzi, Thomas Rüegg, Sizhen Bian, Yawei Li, and Michele Magno. 2023. Tinytracker: Ultra-fast and ultra-low-power edge vision in-sensor for gaze estimation. In *2023 IEEE SENSORS*. IEEE, 1–4.

[4] Steven L Brunton and J Nathan Kutz. 2022. *Data-driven science and engineering: Machine learning, dynamical systems, and control*. Cambridge University Press.

[5] Nicolas Callens and Georges G. E. Gielen. 2021. Analysis and Comparison of Readout Architectures and Analog-to-Digital Converters for 3D-Stacked CMOS Image Sensors. *IEEE Transactions on Circuits and Systems I: Regular Papers* 68, 8 (2021), 3117–3130. https://doi.org/10.1109/TCSI.2021.3085027

[6] Ruth M Casper. 2009. Guidelines for the instrumentation of wild birds and mammals. (2009).

[7] Daniel G. Costa, Luiz Affonso Guedes, Francisco Vasques, and Paulo Portugal. 2012. Effect of frame size on energy consumption in wireless image sensor networks. In *2012 IEEE International Conference on Imaging Systems and Techniques Proceedings*. 239–244. https://doi.org/10.1109/IST.2012.6295487

[8] Tobi Delbruck, Rui Graca, and Marcin Paluch. 2021. Feedback control of event cameras. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*. 1324–1332.

[9] Ryan D. Gow, David Renshaw, Keith Findlater, Lindsay Grant, Stuart J. McLeod, John Hart, and Robert L. Nicol. 2007. A Comprehensive Tool for Modeling CMOS Image-Sensor-Noise Performance. *IEEE Transactions on Electron Devices* 54, 6 (2007), 1321–1329. https://doi.org/10.1109/TED.2007.896718

[10] Yasunobu Hitomi, Jinwei Gu, Mohit Gupta, Tomoo Mitsunaga, and Shree K Nayar. 2011. Video from a single coded exposure photograph using a learned over-complete dictionary. In *2011 International Conference on Computer Vision*. IEEE, 287–294.

[11] Xiaofeng Hou, Tongqiao Xu, Chao Li, Cheng Xu, Jiacheng Liu, Yang Hu, Jieru Zhao, Jingwen Leng, Kwang-Ting Cheng, and Minyi Guo. 2024. A tale of two domains: Exploring efficient architecture design for truly autonomous things. In *2024 ACM/IEEE 51st Annual International Symposium on Computer Architecture (ISCA)*. IEEE, 167–181.

[12] Jong Hwan Ko, Mohammad Faisal Amir, Khondker Zakir Ahmed, Taesik Na, and Saibal Mukhopadhyay. 2017. A single-chip image sensor node with energy harvesting from a CMOS pixel array. *IEEE transactions on circuits and systems I: regular papers* 64, 9 (2017), 2295–2307.

[13] Paul P. Lee, Robert C. Gee, Ronald M. Guidash, Tien-Hsiung Lee, and Eric R. Fossum. 2000. Active Pixel Sensor Integrated With a Pinned Photodiode. Originally filed April 13, 1995, issued Aug. 8, 2000.

[14] Patrick Lichtsteiner, Christoph Posch, and Tobi Delbruck. 2008. A 128 × 128 120 dB 15 μs latency asynchronous temporal contrast vision sensor. *IEEE journal of solid-state circuits* 43, 2 (2008), 566–576.

[15] Robert LiKamWa, Bodhi Priyantha, Matthai Philipose, Lin Zhong, and Paramvir Bahl. 2013. Energy characterization and optimization of image sensing toward continuous mobile vision. In *Proceeding of the 11th Annual International Conference on Mobile Systems, Applications, and Services* (Taipei, Taiwan) *(MobiSys '13)*. Association for Computing Machinery, New York, NY, USA, 69–82. https://doi.org/10.1145/2462456.2464448

[16] Weikai Lin, Tianrui Ma, Adith Boloor, Yu Feng, Ruofan Xing, Xuan Zhang, and Yuhao Zhu. 2025. SnapPix: efficient-coding–inspired in-sensor compression for edge vision. In *2025 62nd ACM/IEEE Design Automation Conference (DAC)*. IEEE, 1–7.

[17] Yi Luo, Derek Ho, and Shahriar Mirabbasi. 2018. Exposure-Programmable CMOS Pixel With Selective Charge Storage and Code Memory for Computational Imaging. *IEEE Transactions on Circuits and Systems I: Regular Papers* 65, 5 (2018), 1555–1566. https://doi.org/10.1109/TCSI.2017.2763822

[18] Yi Luo and Shahriar Mirabbasi. 2022. A 30-fps 192 × 192 CMOS Image Sensor With Per-Frame Spatial-Temporal Coded Exposure for Compressive Focal-Stack Depth Sensing. *IEEE Journal of Solid-State Circuits* 57, 6 (2022), 1661–1672. https://doi.org/10.1109/JSSC.2022.3149916

[19] Tianrui Ma, Yu Feng, Xuan Zhang, and Yuhao Zhu. 2023. CamJ: Enabling System-Level Energy Modeling and Architectural Exploration for In-Sensor Visual Computing. arXiv:2304.03320 [cs.AR] https://arxiv.org/abs/2304.03320

[20] Tianrui Ma, Zhe Gao, Zhe Chen, Ramakrishna Kakarala, Charles Shan, Weidong Cao, and Xuan Zhang. 2025. Systematic Methodology of Modeling and Design Space Exploration for CMOS Image Sensors. *IEEE Transactions on Computer-Aided Design of Integrated Circuits and Systems* (2025), 1–1. https://doi.org/10.1109/TCAD.2025.3585753

[21] Misha Mahowald. 1994. The silicon retina. In *An Analog VLSI System for Stereoscopic Vision*. Springer, 4–65.

[22] Gokul B Nair, Michael Milford, and Tobias Fischer. 2024. Enhancing visual place recognition via fast and slow adaptive biasing in event cameras. In *2024 IEEE/RSJ International Conference on Intelligent Robots and Systems (IROS)*. IEEE, 3356–3363.

[23] Shree K Nayar, Daniel C Sims, and Mikhail Fridberg. 2015. Towards self-powered cameras. In *2015 IEEE international conference on computational photography (ICCP)*. IEEE, 1–10.

[24] Robert H. Nixon, Sabrina E. Kemeny, Craig O. Staller, and Eric R. Fossum. 1995. 128 x 128 CMOS photodiode-type active pixel sensor with on-chip timing, control, and signal chain electronics. In *Charge-Coupled Devices and Solid State Optical Sensors V*, Morley M. Blouke (Ed.), Vol. 2415. International Society for Optics and Photonics, SPIE, 117 – 123. https://doi.org/10.1117/12.206529

[25] Navid Sarhangnejad, Nikola Katic, Zhengfan Xia, Mian Wei, Nikita Gusev, Gairik Dutta, Rahul Gulve, Peter Zhi Xuan Li, Hui Feng Ke, Harel Haim, Manuel Moreno-García, David Stoppa, Kiriakos N. Kutulakos, and Roman Genov. 2019. Dual-Tap Computational Photography Image Sensor With Per-Pixel Pipelined Digital Memory for Intra-Frame Coded Multi-Exposure. *IEEE Journal of Solid-State Circuits* 54, 11 (2019), 3191–3202. https://doi.org/10.1109/JSSC.2019.2932623

[26] Zhao TAO, Wenbin CHEN, Xiaoyang LI, and Rui KANG. 2024. Reliability modelling and assessment of CMOS image sensor under radiation environment. *Chinese Journal of Aeronautics* 37, 9 (2024), 297–311. https://doi.org/10.1016/j.cja.2024.05.008

[27] S. Tsarev, D. Proniakova, X. Liu, et al. 2025. Vertically Stacked Monolithic Perovskite Colour Photodetectors. *Nature* 642 (2025), 592–598. https://doi.org/10.1038/s41586-025-09062-3