Squint: A Framework for Dynamic Voltage Scaling of Image Sensors Towards Low Power IoT Vision

Venkatesh Kodukula
Arizona State University
Tempe, AZ, USA
vkoduku1@asu.edu

Mason Manetta
Arizona State University
Tempe, AZ, USA
mmanetta@asu.edu

Robert LiKamWa
Arizona State University
Tempe, AZ, USA
likamwa@asu.edu

ABSTRACT

Energy-efficient visual sensing is of paramount importance to enable battery-backed low power IoT and mobile applications. Unfortunately, modern image sensors still consume hundreds of milliwatts of power, mainly due to analog readout. This is because current systems always supply a fixed voltage to the sensor's analog circuitry, leading to higher power profiles. In this work, we propose to aggressively scale the analog voltage supplied to the camera as a means to significantly reduce sensor power consumption. To that end, we characterize the power and fidelity implications of analog voltage scaling on three off-the-shelf image sensors. Our characterization reveals that analog voltage scaling reduces sensor power but also degrades image quality. Furthermore, the degradation in image quality situationally affects the task accuracy of vision applications.

We develop a visual streaming pipeline that flexibly allows application developers to dynamically adapt sensor voltage on a frame-by-frame basis. We develop a voltage controller that programmatically generates desired sensor voltage based on application request. We integrate our voltage controller into the existing RPi-based video streaming IoT pipeline. On top of this, we develop runtime support for flexible voltage specification from vision applications. Evaluating the system over a wide range of voltage scaling policies on popular vision tasks reveals that Squint imaging can deliver up to 73% sensor power savings, while maintaining reasonable task fidelity. Our artifacts are available at: https://gitlab.com/squint1/squint-ae-public

ACM MobiCom '23, October 2–6, 2023, Madrid, Spain

ACM ISBN 978-1-4503-9990-6/23/10…$15.00
https://doi.org/10.1145/3570361.3613303

CCS CONCEPTS

• **Computer systems organization → Sensors and actuators**.

KEYWORDS

Image Sensors, Voltage Scaling, IoT, AR/VR, Energy Efficiency

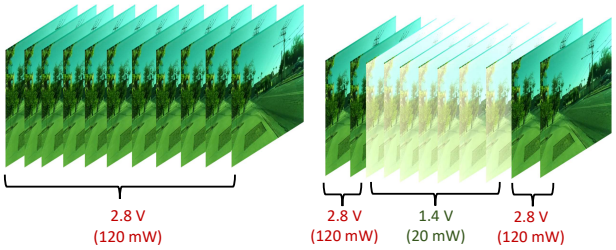
ACM Reference Format:
Venkatesh Kodukula, Mason Manetta, and Robert LiKamWa. 2023. Squint: A Framework for Dynamic Voltage Scaling of Image Sensors Towards Low Power IoT Vision. In *The 29th Annual International Conference on Mobile Computing and Networking (ACM MobiCom '23), October 2–6, 2023, Madrid, Spain.* ACM, New York, NY, USA, 15 pages. https://doi.org/10.1145/3570361.3613303

1 INTRODUCTION

Visual computing systems sense and perceive the world using cameras, enabling a variety of IoT and mobile applications. IoT systems use cameras to perform visual tasks such as predicting irrigation patterns in AI-enabled farms [1], detecting wildfires in forests [2], and monitoring animal movements in wildlife sanctuaries [3, 4]. Energy-efficient sensing is of utmost importance for such IoT systems, as they are deployed in the wild and typically operate on miniscule batteries. Energy-efficient sensing is also important for multi-camera mobile systems, such as augmented reality headsets, wherein several cameras, e.g., 18 cameras on the Magic Leap v2 headset [5], perform tasks such as world tracking, eye tracking, and body tracking, quickly draining the system's battery.

Unfortunately, modern image sensors are limited in their energy efficiency due to analog readout and consume hundreds of milliwatts of power while performing continuous visual tasks. This is because existing systems (Fig. 1a) always supply a fixed voltage to the sensor's analog circuitry regardless of the frame. This work aims to improve the capabilities of visual computing systems by allowing them to dynamically vary the analog voltage supplied to the sensor on a frame-by-frame basis (Fig. 1b) for significant efficiency gains.

While dynamic voltage scaling of image sensors leads to large power savings, it also degrades the sensor's imaging



(a) Status quo (b) Squint

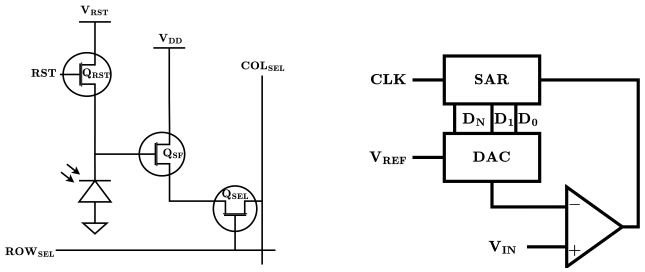
Figure 1: Existing systems supply a fixed voltage to the image sensor regardless of the frame. Squint adapts the camera supply voltage per frame for energy efficiency.

fidelity due to reduced ADC output swing and increased pixel noise (§2). To assess fidelity issues, we characterize the energy and fidelity implications of analog voltage scaling of three popular off-the-shelf cameras. In addition to confirming the large sensor power savings created by aggressive voltage scaling, our characterization reveals a useful insight: degradation in imaging fidelity due to aggressive voltage scaling does not significantly affect the task accuracy of modern neural network based vision workloads in most situations. That said, we find that high fidelity is still sporadically needed to precisely detect scene features for tracking tasks, especially under challenging situations such as low-light and crowded scenes.

The idea to dynamically scale a component's voltage for reduced energy usage is inspired by the CPU power savings technique of dynamic voltage scaling (DVS) [6]. However, unlike CPU DVS, analog voltage scaling here introduces unique noise considerations, due to the direct reliance on the analog circuitry's fidelity for image readout. Sensor manufacturers have been moving towards designing sensors [7–10] with lower pixel supply voltage, for lower energy consumption. Such sensors typically include additional circuitry to mitigate fidelity issues caused due to lower pixel voltage.

To design the system support (§3) to allow sensor dynamic voltage scaling, we introduce a simple voltage controller hardware interface that can programmatically generate desired sensor voltage requested by applications. On top of that, we introduce the Squint runtime that includes a software API allowing application developers to seamlessly specify voltage schedules directly from the vision applications on a frame-by-frame basis. These interfaces integrate well with existing and future visual computing systems, allowing for an easier upgrade.

To evaluate our system, we design our implementation (§4) around a RPi based streaming pipeline. We support various visual workloads, including neural network-based people detection and OpenCV-based camera pose estimation. Through our evaluation (§5) across a range of voltage scheduling policies, we demonstrate the opportunity of Squint



(a) Pixel design (b) ADC design

Figure 2: Analog voltage variation affects pixel and ADC circuits. A lower voltage (V_{DD}) for pixels would affect transistor settling times, resulting in noise. A lower reference voltage (V_{REF}) for ADC would decrease the dynamic range, resulting in reduced contrast. Note: V_{RST} helps clear pixel charge on soft reset (RST) and is unaffected by voltage scaling.

based imaging to decrease sensor power by up to 73%, while only minimally degrading the visual task accuracy. Through reduction in sensor power, Squint imaging techniques can reduce the power of popular IoT and XR systems by 10%.

In summary, we make the following contributions:

- We explore the idea of dynamic voltage scaling for image sensors, where different frames are captured at different fidelities for overall system energy efficiency, while respecting task needs.
- We characterize the energy and fidelity implications of analog voltage scaling on off-the-shelf commercial image sensors.
- We develop a lightweight and fully programmable voltage controller to generate desired analog supply voltage for the camera. On top of that, we develop a library and runtime to coordinate vision applications with voltage controller operation.
- We augment our hardware and software support on top of an existing commercial IoT streaming pipeline built around the RPi platform. We evaluate the augmented system on a variety of vision tasks to demonstrate significant reduction in sensor power with controllable accuracy loss.

2 BACKGROUND AND MOTIVATION

2.1 Analog readout is the bottleneck for sensor energy efficiency

Modern IoT and mobile systems employ CMOS image sensors for video analytics and photography. CMOS image sensors exhibit higher efficiency compared to CCD sensors, but they still consume hundreds of milliwatts, mainly due to their power-hungry analog readout [11].

Table 1: Characterized image sensors. Sensors can be undervolted far below the minimum voltage limit specified in datasheets.

Camera system	Sensor	Max. res	Voltage range (datasheet)	#cams char	Actual lower limit
RPi cam	Sony IMX219 [17]	4K	2.8 +/- 10%	5	1.2 - 1.5
Pixycam	OmniVision OV5647 [18]	1080p	2.8 +/- 10%	1	1.8
Python1300	OnSemi Python1300 [19]	SXGA	2.8 +/- 10%	1	1.7

Modern image sensors comprise three major elements: a pixel array, an analog readout, and digital logic. The pixel array converts light into voltage, consuming minimal energy [12]. Analog readout amplifies analog signals generated by the pixel array and converts them into digital values. Finally, the digital logic oversees sensor operation, managing timing generation and data preparation for external transmission.

Notably, the analog readout consumes 50%-75% of overall energy in recent sensor designs [12–14]. This is due to two factors: (i) Analog circuits need to operate all the time, making them power hungry compared to their digital counterparts which operate only on the active clock edge. (ii) Technology scaling improvements for energy efficiency have been much slower for analog components compared to digital.

2.2 Effects of analog voltage variations on image sensor circuitry

Existing systems supply higher analog voltage, typically 2.8 V, for high fidelity imaging [15]. Higher analog voltage (VDD) is needed to maintain sufficient drive strength of the photodiode, source follower, and bitline components in the pixel design, shown in Fig. 2a, for smoother charge transfer. Lowering analog voltage would weaken the drive strength of those components affecting transistor settling times [15, 16], resulting in image noise. Lowering analog voltage would also slow down [8, 9] the charge transfer process. This would not significantly affect the imaging process if the photodiode accumulates enough charge, e.g., well-lit scenes. However, this would make the images dimmer in low-lit scenes where photodiodes do not accumulate enough charge.

Higher analog voltage (VREF) is also needed for high fidelity ADC operation for maintaining high dynamic range. Image sensors typically have a successive-approximation (SAR) ADC (Fig. 2b) that performs a binary search for generating digital output by successively comparing input voltage (VIN) against reference voltage (VREF) [20]. A higher VREF would enable higher precision output supporting a wider range of pixel intensity values leading to high dynamic range images. A lower VREF, on the other hand, would decrease the dynamic range forcing the pixels to saturate early and decreasing the overall image contrast.

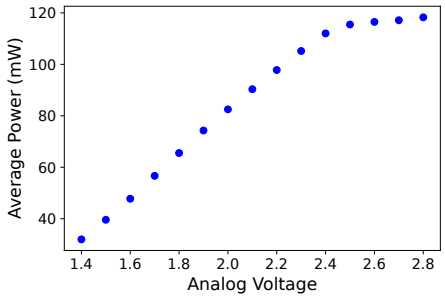


Figure 3: Lowering the supply voltage significantly reduces the sensor power. The flat region above 2.4 V is due to strong voltage regulation inside the sensor.

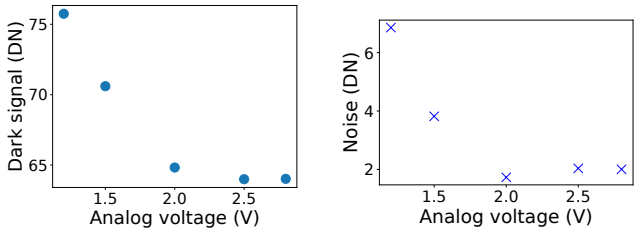
While the exact pixel and ADC designs might vary for different image sensors, the fundamental underlying phenomenon of how the circuits behave at different analog voltage levels would remain the same. Furthermore, the underlying phenomena is agnostic to the shutter type, rolling or global, of the image sensor as the shutter type does not affect the charge transfer process and ADC operating mechanism.

2.3 Analog voltage scaling for energy efficiency

Towards addressing the energy efficiency bottleneck, we propose to scale analog voltage to promote camera energy efficiency. We study the energy and fidelity implications of voltage scaling by characterizing three different image sensors from three popular manufacturers, as shown in Table 1. These image sensors vary in terms of resolution, shutter type, and pixel designs, and are representative of cameras available in typical IoT and mobile systems. In this section, we mainly present the findings of RPi camera characterization; while other cameras revealed similar insights.

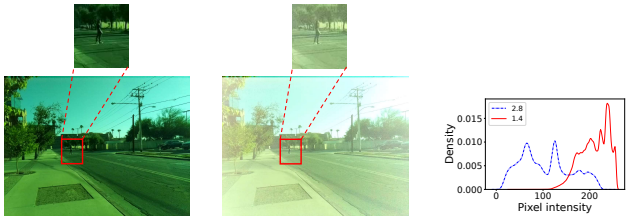
We observe that image sensor manufacturers usually take a conservative approach when it comes to specifying the recommended operating analog voltage for stable camera operation. They typically recommend a 10% tolerance band around nominal voltage in the datasheets. In reality, however, we find that we can undervolt the analog power rail to a value that is far below the recommended limit, as shown in Table 1, without affecting basic camera functionality. For instance, we can aggressively undervolt the RPi's analog power rail by up to 50% of its nominal value with the camera still streaming pixels.

The extent to which we can undervolt a camera strongly depends on sensor design and the sensor's sensitivity to analog voltage changes. Among the three sensors we characterize, we notice the RPi camera can be undervolted by



(a) Brightness (b) Noise

Figure 4: Sensor brightness and noise levels increase with lower analog voltage. Increased brightness is due to decreased ADC output swing and increased noise due to elevated shot noise levels.



(a) Image (2.8 V) (b) Image (1.4 V) (c) Histograms

Figure 5: Images captured at different sensor voltages and their histograms. Low voltage image is brighter and grainier than its high voltage counterpart. This is also reflected in the shift in mean and variance width in the histogram. Images appear green because they are captured raw, i.e., without white balance.

a large extent compared to other ones. In addition to variability across cameras from different vendors, we also notice variability across cameras from the same vendor, as shown in Table 1. This is due to the silicon lottery [21]: One RPi camera can undervolt slightly better than the other and vice versa.

2.3.1 Energy implications of voltage scaling. Aggressive undervolting leads to significant energy savings, as dynamic power quadratically varies with voltage ($P \propto V^2$). Since analog circuits are slow on technology scaling, dynamic power still dominates the overall power consumption by contributing to about two thirds of total power. On the other hand, static power, which is the less dominant power source in analog electronics, is largely unaffected by voltage scaling.

To understand how camera power varies with analog voltage scaling, we intercept the analog power rail of the RPi camera and instead connect it to an external variable power supply. To compute overall power, we measure the current drawn from different power rails while we vary analog voltage. We find that undervolting the camera voltage by 50%, i.e. from 2.8 V to 1.4 V, yields more than 50% camera power savings, as shown in Fig. 3.

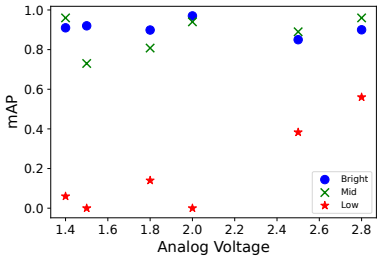
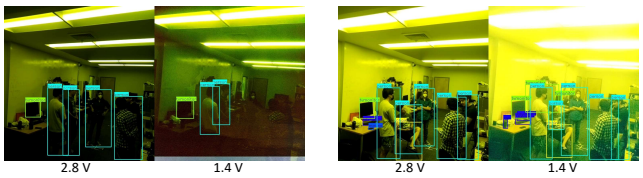


Figure 6: Brighter and noisier images do not significantly affect visual task fidelity in well-lit scenes.



(a) Dimly-lit (b) Well-lit

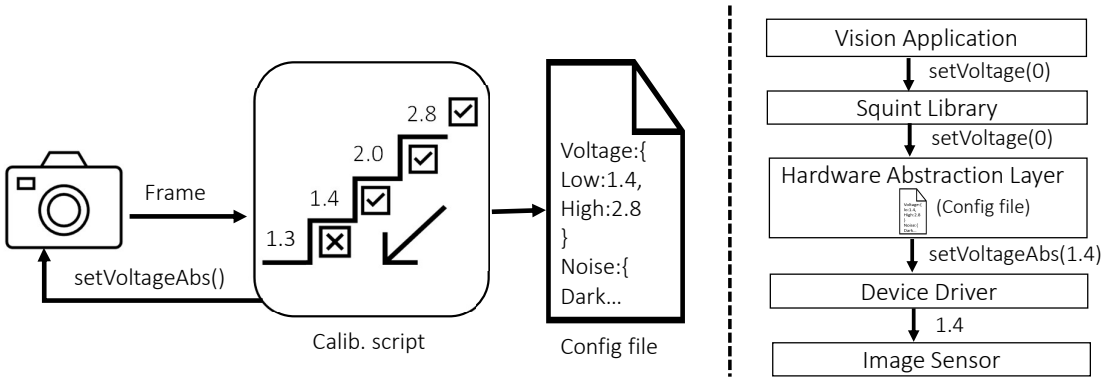
Figure 7: High fidelity is still needed for precise object detection in challenging scenarios. Low fidelity suffices otherwise.

2.3.2 Image fidelity implications of voltage scaling. While aggressively reducing analog voltage helps significantly reduce camera power, it also impairs image quality by making images brighter and noisier. Increased brightness is due to reduction in ADC dynamic range. That is, image sensors usually have a SAR ADC and lowering the ADC's reference voltage results in lowering the output signal swing. This forces pixels to saturate early, brightening the image and decreasing the overall contrast.

Increased noise is due to more pixel shot noise triggered by reduction in pixel bias voltage as shot noise exponentially varies with the negative of bias voltage. In addition to brightness and noise artifacts, we also notice pixelation artifacts appearing in images. This is due to pixels randomly getting turned off possibly due to timing errors caused by insufficient supply voltage.

We experimentally validate these claims by measuring the standard dark signal [22] and temporal noise [23] of the RPi camera while we vary sensor analog voltage. We capture 25 frames per voltage setting and average pixel values across frames to determine the average frame. We average the pixel values of the average frame to compute the dark signal. For temporal noise, we compute variance of pixel values across frames to determine the variance frame. Then we average all the values in the variance frame to compute the temporal noise.

As shown in Fig. 4, we notice the dark signal and the temporal noise substantially increase as we decrease the sensor



(a) Calibration **(b) Execution**

Figure 8: Phases of Squint. During calibration, Squint sweeps voltage supplied to the camera to determine the undervolting limit, as well as to profile noise. This information is stored in a configuration file. During execution, the system uses the configuration file to translate the developer's high-level fidelity request to actual low-level camera voltage using the standard camera system stack.

voltage. We also noticed the sensor being insensitive to voltage changes until a certain point, which could be due to strong voltage regulation inside the sensor. The brightness and noise increase is also reflected in the histograms of actual images, as shown in Fig. 5. We can see that the peaks shift rightwards and the distribution spreads outwards as we switch from high voltage to low voltage.

On a side note, we can also see that the actual images in Fig. 5 appear more green because they are captured RAW by turning off most image signal processing (ISP) stages. Notably, we turn off automatic white balance (AWB) to avoid color based artifacts produced by AWB. AWB algorithms [24–27] estimate color temperature of a scene based on average image intensity in order to determine appropriate gains that need to be applied for individual color channels of an image. As a result, AWB perceives brightness increase due to undervolting as change in scene brightness and apply more gain to the blue color channel resulting in bluish artifacts.

2.3.3 Task fidelity implications of voltage scaling. While aggressive undervolting impairs image fidelity by adding brightness and noise, we find that brightness and noise increases do not affect visual task fidelity in most situations. However, we see a noticeable degradation in task fidelity under challenging scene environments, e.g., low-light scenes, as shown in Fig. 6. This creates an opportunity to situationally adapt camera voltage based on scene illumination needs. That is, we can aggressively undervolt the camera most of the time for rapid power savings and occasionally bring the voltage back to nominal under poor lighting conditions. Furthermore, we observe that there is no incentive to operate the camera at intermediate voltages. This is because for well-lit scenes, vision tasks perform reasonably well even at 1.4 V, and for dimly-lit scenes, vision tasks perform reasonably well only at 2.8 V, as shown in Fig. 6.

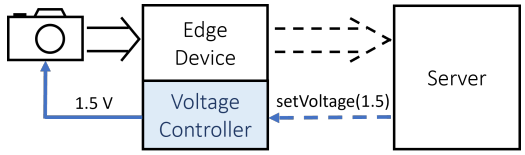
The insensitivity of vision tasks to brightness and noise increase in most situations stems from the fact that neural networks running behind these tasks are usually trained to make themselves immune to intensity and noise changes [28]. That being said, vision tasks still need high fidelity for performing precise object detection, especially under challenging conditions such as dimly-lit scenes, as shown in Fig. 7a. For less challenging scenes, we observe that low-fidelity images provide comparable task performance to their high-fidelity counterparts, as shown in Figs. 7b, 15a, 16a. For this analysis, we experimentally capture images at different voltage settings from the RPi camera for three different lighting conditions. We then feed the images to a YOLO-based people detection task and compare the detection results against ground truth to determine accuracy.

2.4 Motivational observations

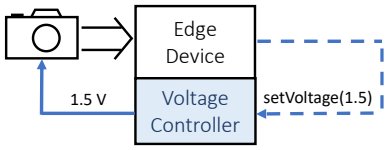
To summarize, we have the following insights for image sensor voltage scaling:

- We can aggressively undervolt image sensors far beyond specified operating ranges in the datasheets.
- Aggressive undervolting results in significant camera power savings.
- Aggressive undervolting also makes images brighter and noisier due to the reduced upper bound of the analog voltage range.
- Raised brightness and noise levels in images does not significantly affect vision task accuracy in most situations.

These observations motivate the need for dynamic voltage scaling strategies for image sensors to save energy at sufficient task fidelity.



(a) Server-based



(b) Edge-based

Figure 9: Different Squint system configurations. Server-based: vision apps running on server issue voltage request to the voltage controller. Edge- based: vision apps run on the edge device itself and directly issue requests to the voltage controller.

3 DESIGN

Design aspirations: We set three goals that guide the design of system support for sensor dynamic voltage scaling through our hardware and software extensions.

- Seamless: Vision applications should be able to reconfigure camera voltage on a frame-to-frame basis without any frame drops
- Lightweight: Our hardware extension should be lightweight with minimal system overhead
- Flexible: Our runtime should allow developers to flexibly specify camera voltage needs on a per-frame basis.

Calibration and Execution Phases: We propose two phases for Squint to seamlessly work with existing and future visual computing systems, as shown in Fig. 8.

(1) To account for variations across sensors, the system goes through a one-time calibration phase to determine the actual voltage limits and noise profile of the image sensor.
(2) The system then enters an execution phase to guide policy based voltage selection based on the calibration information.

(1) Calibration phase: During this phase, the system runs a script that automatically sweeps through different camera voltages and determines what set of voltages are usable. The script starts with the default nominal voltage specified in the datasheet and keeps decreasing the voltage in steps of one tenth of a volt until the camera freezes, at which point the camera voltage is restored to nominal. While it does so, the script also captures several frames for noise assessment. The system commits the voltage limits and noise data into a configuration file, which could be stored in the hardware abstraction layer of the system stack.

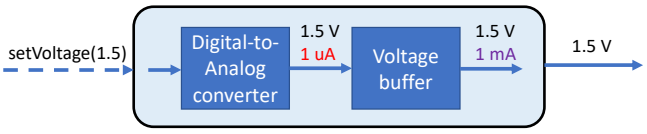


Figure 10: Voltage controller comprises a DAC to translate digital voltage requests into corresponding analog voltage and a voltage buffer to generate sufficient current flow for reliable camera operation.

(2) Execution phase: During this phase, the system uses the configuration file generated during the calibration step to translate the application's high-level fidelity requirements into actual set of voltages that needs to be applied to the camera. These voltages are sent to the voltage controller which enacts them on the camera using standard device-driver mechanisms.

System support for Squint: To efficiently run the calibration and execution phases, we build system support for Squint by extending the existing IoT pipeline with necessary hardware and software extensions. These extensions could be integrated into different system incarnations, such as server-based and edge-based, depending on where the vision application runs in the system, as shown in Fig. 9. The resulting Squint system architecture centers around the idea of adaptively scaling camera voltage for vision application usage. We describe the design of our voltage controller and its integration with the existing IoT streaming pipeline, as shown in Fig. 9. We also discuss the software runtime for vision application developers to leverage the voltage controller through policy specification.

3.1 Programmable Voltage Controller

The voltage controller module, shown in Fig. 10, intercepts the incoming voltage reconfiguration command from a server and produces appropriate analog voltage to power the camera. The vision application specifies voltage needs using the designed runtime support in §3.2.

We design a fairly simple voltage controller comprising a digital to analog converter (DAC) and a voltage buffer. The DAC programmatically takes desired voltage from vision application and converts that into equivalent analog voltage to power the camera. The DAC is supported by the voltage buffer which is a unity-gain amplifier that generates sufficient current for stable camera operation.

3.1.1 Integration with edge device. The voltage controller could be easily integrated as a standalone integrated circuit on top of the edge device. Alternatively, we could repurpose existing DVFS controllers inside SoCs of edge devices to generate desired voltages for the camera. That would entail

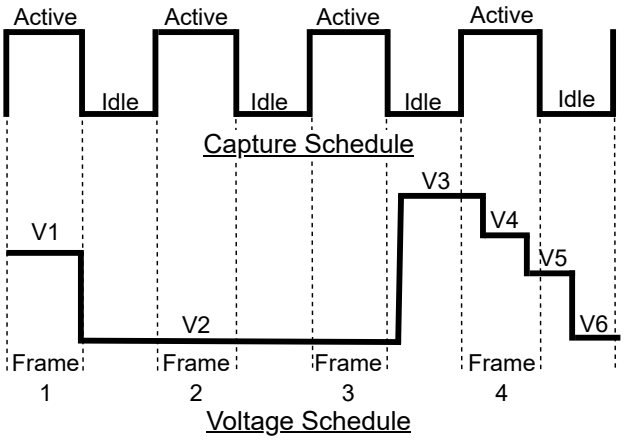


Figure 11: Timing different voltages in tandem with camera capture schedule leads to different fidelity patterns across the frames and within a frame.

adding a few extra power modes to meet camera power supply requirements.

3.2 Developer support

We develop runtime support to allow the developers to flexibly specify voltages. This consists of a `SetCamVoltage()` function for developers to set a voltage. Voltage can be set on a per-frame basis or persist across frames. A runtime service receives these calls to send the voltage to the voltage controller.

```
SetCamVoltage(float val, string spec="SNR/Rel/Abs")
```

In order to facilitate development, we provide multiple levels of abstraction for voltage specification. We provide sensor-agnostic ways to configure voltage such as through "SNR" and relative voltage specification ("Rel"), allowing developers to be unconcerned with the intricacies of imaging and power trends. Additionally, we also provide sensor-specific ways to configure voltage to allow developers to directly set absolute ("Abs") camera voltages.

For relative voltage specification, we let developers specify their desired voltage in [0, 1] range with 0 denoting the low voltage limit and 1 denoting the high voltage limit, regardless of the camera. Our decision to abstract this information is grounded in the observation of the monotonicity of imaging (Fig. 4) and power (Fig. 3) trends.

Our runtime service translates the specification into appropriate hardware voltage with the help of the configuration file generated during the calibration phase of the system. The final voltage is then enacted to the sensor hardware through device driver, similar to how it is done in standard Android and iOS mobile systems architecture.

3.2.1 Integration with CameraAPI and sysfs interface. Our voltage configuration method could be easily integrated as part of standard CameraAPI. The CameraAPI is an application programming interface that allows software developers to interact with a device's camera hardware, including controlling camera settings such as exposure, focus, and zoom, in a platform agnostic manner. By integrating voltage control into the existing CameraAPI, developers can easily configure voltage through a familiar interface.

```
VideoCapture cap(0);
cap.set(cv::CAP_PROP_EXPOSURE, 0.0167);
cap.set(cv::CAP_PROP_VOLTAGE, 0.5);
```

The above snippet shows an example integration using the popular OpenCV-based Linux CameraAPI. In this approach, voltage can be defined as a camera property and set to a desired value, similar to other camera settings such as exposure. This method of integration can also be adapted to Android and iOS cameraAPIs.

Alternatively, voltage control can be integrated as part of sysfs interface to allow developers to configure and query sensor voltage through a standardized interface controlled by the operating system. This approach involves modifying the camera device driver to create voltage as a sysfs file node, which can then be exported to userspace using the sysfs interface. Developers can also directly specify voltage scheduling policy, instead of voltage, to enable automatic voltage control by OS, reducing developer burden.

```
echo 0.5 > /sys/devices/camera/sensor/voltage
echo "Perf" > /sys/devices/camera/sensor/policy
```

3.2.2 Voltage scheduling policies. Developers can build various policies that autonomously guide the voltage configuration, in the similar spirit to issuing voltage configurations with the Linux DVFS API on desktop systems and Android's Frequency API on mobile systems. Policies can incur different system overheads, leading to system trade-offs.

A policy should predict voltage progression with time as well as image quality requirements to maximize task accuracy. A simple policy could occasionally operate a camera at high voltage to detect frames and undervolt the camera for tracking the detected objects across frames. Developers can also introduce improved application-specific proxies with other prediction strategies, e.g., with optical flow based voltage reconfiguration.

Notably, depending on how the voltage schedule is aligned with the camera's capture schedule, different fidelity granularities could be achieved within a single image, as shown in Fig. 11. Modern cameras employ a rolling shutter mechanism that exposes and streams one line of pixels at a time. For coarse-grained fidelity control, the system should sustain

Table 2: System components in the IoT streaming pipeline

Component	Specification
Camera	RPi Cam v2.1, Sony IMX219, 4K @ 60 fps
Edge device	RPi v 3B, quad-core Cortex-A53, 1GB RAM
Network	300 Mbps Ethernet
Server	Intel Xeon CPU E5-1620, 32 GB RAM

the voltage for the entirety of the active frame capture duration. On the other hand, for fine-grained fidelity control, the system could apply different voltages to different parts of the frame.

4 IMPLEMENTATION

Platform: We build our IoT system (Fig. 12a) as a gstreamer pipeline around RPi to be representative of a standard video streaming system. As shown in Table 2, we use RPi v2 camera [29] as the capture device, connected to a RPi v3B. The RPi forms the edge device that consumes the incoming pixel stream from the camera and compresses the stream into a compact representation using standard H264 compression techniques. We run a real-time streaming protocol (RTSP) server [30] on the Pi that transmits the compressed pixel stream from the edge device over Ethernet to the server. Finally, we utilize a server class Intel Xeon CPU desktop that decompresses the compressed pixel stream for performing video analytics, e.g., people detection. The entire pipeline runs in real-time at 24 fps.

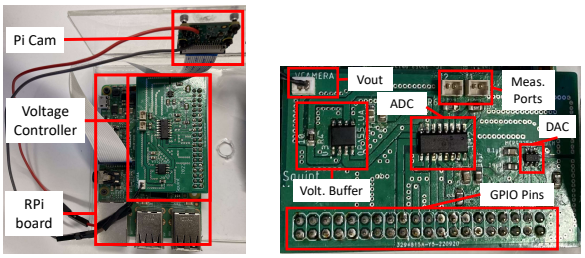
Voltage controller: We design our voltage controller (Fig. 12b) as a fully programmable module by using off-the-shelf components. Specifically, we use Microchip's MCP4725 12-bit DAC [31] as our digital-to-analog converter and Texas Instruments OPA551PA operational amplifier [32] to build the voltage buffer that generates sufficient current for stable camera operation. We integrate the voltage controller as a standalone module, powered and interfaced through the I2C interface of the Pi. The voltage controller takes voltage configuration commands from the Pi and generates the desired voltage to power the camera's analog voltage rail.

Runtime: We implement our runtime as a filesystem interface (§3.2) using a stateless networking service protocol. Specifically, we implement a RESTful API [33, 34] to handle voltage configuration requests from the vision applications running on the server. A simple web service running on the Pi forwards these voltage configuration requests to the voltage controller over the I2C interface.

4.1 RPi based system integration

4.2 Workloads

We study YOLO-based people detection that localizes and tracks people in a scene to allow for counting analytics and



(a) Squint **(b) Voltage Controller**

Figure 12: Squint system setup

potential security responses. This would be representative of multiple IoT video analytics applications such as counting people entering and leaving a trail entrance and analyzing shoppers' buying activity [35] for automated checkout. Additionally, we perform a preliminary investigation of OpenCV marker based camera pose estimation that is the backbone for augmented reality tracking tasks. This would be representative of multi-camera AR systems such as Microsoft Hololens [36].

4.2.1 Benchmarks. Since existing datasets contain images taken at nominal camera voltage and the undervolting image artifacts are difficult to simulate, we construct our own dataset based on real image captures at different camera voltages. Specifically we stage our dataset representing two real-life scenarios:

- Entrance detection: people entering and leaving a scene
- Person tracking: people randomly standing around in a room holding casual discussions in places such as office cafeterias

Our dataset contains variations in lighting, object density, and object proximity, making it a viable proxy for real life scenarios. We recruit five people and take their consent for staging the dataset. We instruct the participants to move in specific patterns that emulate the entrance detection and the person tracking in cafeteria scenarios mentioned above. During the movements, we vary the ambient light using a controllable analog light source. Specifically, each capture consists of six minutes with two minutes of high light, then two minutes of medium light, and finally two minutes of low light.

We use our experimental RPi camera to stage the dataset at different voltages, while the participants emulate the scenarios. In addition to the experimental camera, we also use a ground truth camera that always operates at nominal voltage for ground truth captures. We place the experimental and ground truth cameras in close proximity so that they see the same scene with negligible parallax. Overall, the dataset includes 73,066 images.

For the marker based pose estimation, we construct a limited dataset using the same setup as above. For this workload, we place the marker at three different distances and three different orientations with respect to the camera.

4.2.2 Baselines. We test our workloads against the following baselines:

- Squint (Sq-X): The system supplies the desired voltage to the camera based on the vision application's fidelity needs. Here "X" denotes the policy used for voltage scaling. Notably, the system operates the camera at extreme voltages, i.e., 2.8 V and 1.4 V, for RPi camera, since there is no incentive to operate in between, as discussed in §2.3.3..
- Reduced resolution (RR): Instead of modulating voltage, the system changes capture resolution based on the fidelity needs. Analogous to Squint, the system configures the camera at 1080p for high fidelity and at 480p for low fidelity.

4.2.3 Metrics. Here we discuss different evaluation metrics.

Task accuracy: We use the standard metrics from computer vision literature to evaluate the workloads. For the people detection workload, we use intersection over union (IoU) as the metric. IoU measures the amount of overlap between estimated and ground truth detections. A detection is considered a true positive (TP), if the IoU score is above a certain threshold. Otherwise, it is considered a false positive (FP). The final detection accuracy is obtained by determining the number of true positives among all the detections, i.e., TP/(TP + FP), for each frame, which is known as mean average precision (mAP). The final mAP is obtained by averaging individual mAP scores across all frames. On the other hand, for marker-based camera pose estimation workload, we use translational error and rotational error as the metrics. These are derived by computing the L2 norm of estimated pose against ground truth pose.

Power: We sample different camera power rails every 0.1 ms to measure their respective current draws. We multiply the current draw with the voltage of each rail and sum them over all the rails to get the sensor power. We analyze the sensor power readings across different samples to report power trends.

Overhead: We measure the voltage reconfiguration time of our API by measuring the time difference between voltage reconfiguration request sent by the vision application and acknowledgement notification received from the Pi's web service. We measure the reconfiguration time across multiple voltage requests to report trends. On the other hand, to report power overhead of the voltage controller module, we use worst case power estimates from the component datasheets.

4.2.4 Policies/Parameter choices. We evaluate the following policies, in a similar spirit as CPU's DVFS policies [37, 38].

Performance (Perf): The system sets the sensor voltage to the highest possible setting, providing maximum task performance but potentially consuming more power and generating more heat. This is representative of status-quo.

Powersave (Pwr): The system sets the sensor voltage to the lowest possible setting to save power.

Random (Rand): To demonstrate the effectiveness of flexible voltage reconfiguration, we map a random policy whereby the system generates a random number from 1.4 to 2.8, and supplies that voltage to the image sensor.

Lightutil (Light): The system uses ambient light sensor's reading to determine the optimal voltage setting for the current workload. The system uses a lighting - voltage lookup table that is obtained from sensor characterization to guide the voltage selection process.

For our evaluation, the system occasionally captures a high fidelity image, computes its average pixel intensity, and uses it as a proxy to estimate the ambient lighting of a scene. If the light estimate is greater than a certain threshold, the system supplies 1.4 V to the sensor; otherwise it supplies 2.8 V to the sensor, as guided by our characterization curves.

On-Demand (OD): The system sets the sensor voltage to the lowest possible setting when the scene is idle and scale the voltage up as scene complexity increases. Specifically, the system computes IoU score across detection results of previous frames to determine the staticness/dynamism of a scene's objects and uses the IoU score to appropriately determine the desired sensor voltage.

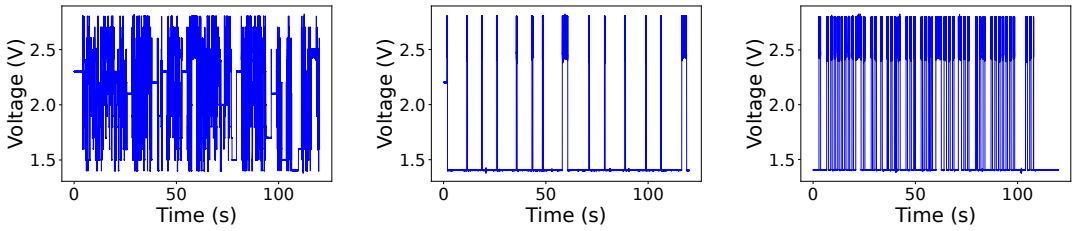
A larger IoU score across frames indicates that the scene is fairly static. In such scenarios, the system supplies 1.4 V to the camera to continue to track the existing objects. On the other hand, a lower IoU score across frames indicates large motion in the scene, including people entering and exiting the scene. In such scenarios, the system supplies 2.8 V to the camera for high quality detections.

Distance based (Dist): We use a distance based policy for marker based camera pose estimation workload, whereby the system adapts voltage based on proximity of marker with respect to the camera. The system supplies a low voltage (1.4 V) to the camera when the marker is close to the camera, and supplies a high voltage (2.8 V) when the marker is located far from the camera for precision.

5 EVALUATION

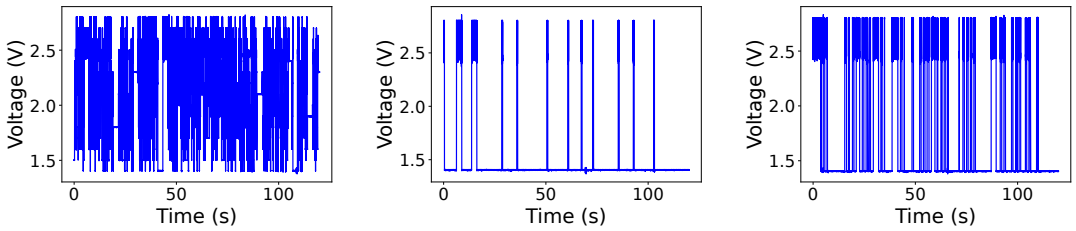
5.1 The usage of Squint is flexible, seamless, and performant

Our runtime flexibly allows apps to configure full swing of sensor voltages on a frame-to-frame basis without any restrictions. The voltage schedule for random policy shown in Fig. 13 indicates that the API is flexible enough to randomly



(a) Random **(b) Lightutil** **(c) On-Demand**

Figure 13: Bright Light Voltage schedules generated by different policies for *Entrance detection* dataset.



(a) Random **(b) Lightutil** **(c) On-Demand**

Figure 14: Bright Light Voltage schedules generated by different policies for *Person tracking* dataset.

configure sensor voltage anywhere from 1.4 V to 2.8 V. On the other hand, the system operates the image sensor at extreme voltages, either 1.4 V or 2.8 V, for Lightutil and On-Demand policies, as shown in Fig. 13, because there is no power/performance incentive to operate at an intermediate voltage, as mentioned in §2.3.3.

We find that the system chooses a different voltage schedule for different policies for a given scenario. For instance, we can see On-Demand policy schedule has a denser schedule than Lightutil, because it's motion adaptive, and motion changes are frequent in both the entrance detection and the office cafeteria scenes. Lightutil on the other hand has a sparser schedule because it's light adaptive and light changes are less frequent in the scenes.

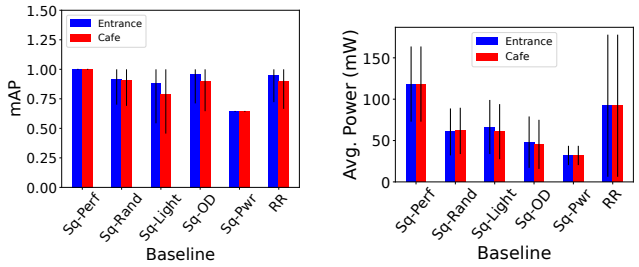
In addition to variation across policies, we find that the same policy selects a different voltage schedule for different scenarios. For instance, we can see On-Demand having a denser schedule for the entrance detection scenario (Fig. 13) compared to the person tracking in cafeteria scenario (Fig. 14), as there is large motion involved in the former due to people constantly entering and leaving the scene as opposed to the cafeteria one where people are fairly static.

While the runtime configures different voltages for different policies, we find that Squint is able to stream frames seamlessly without any noticeable frame drops. By doing so, Squint is able to provide a consistent real-time pipeline performance of 24 fps for all the evaluated workloads.

Notably, we find that voltage configuration requests executed while the sensor is actively capturing an image results in different parts of the frame possessing different qualities. This is due to rolling shutter sensor operation whereby different lines of pixels are read at different voltage configurations. We notice such artifacts during the dataset collection with more artifacts appearing when rapidly changing voltages. This opens up an opportunity to design more fine-grained voltage scheduling policies that can sample different parts of a frame, as mentioned in §3.2. We plan to study the fidelity effects of such policies deeply in a future work.

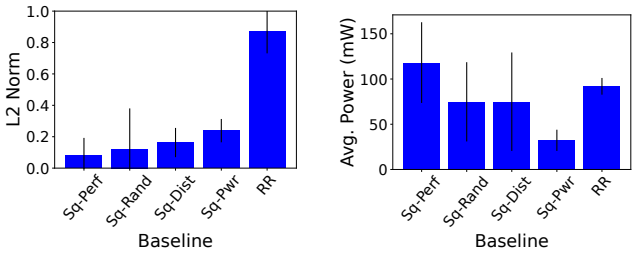
Algorithms are still reliable. We find that task accuracy is fairly maintained for people detection workload, while policies choose different voltage schedules during system execution, as shown in Fig. 15a. Notably, On-Demand has slightly better accuracy than Lighutil because it's adaptive to both light and motion in the scene. Random policy has comparable performance to Lightutil because the system randomly modulates the sensor supply voltage with an average operating voltage landing in the middle. At this average intermediate voltage, the frame quality is comparable to the frame quality at a nominal voltage. Lower accuracy for Powersave policy is due to continuous low voltage operation, regardless of scene dynamics. On the other hand, reduced resolution (RR) baseline also provides consistent accuracy as neural networks, by design, are typically trained to perform well at low input resolutions, e.g., 300x300 for YOLO.

From Fig. 15a, we also notice that the accuracy varies quite a lot within a policy as indicated by the large error bars. This variation is mainly because of varying scene dynamics pertaining to occlusion, object proximity, and contrast difference in the dataset. Specifically, we notice the difficult scene situations such as people occluding one another, people standing



(a) Task accuracy **(b) Power**

Figure 15: People detection task



(a) Task accuracy **(b) Power**

Figure 16: Camera pose detection task.

far from the camera, and natural contrast differences due to non-uniform illumination in the room result in lower task accuracy, and vice-versa.

For marker based pose estimation workload, we find the accuracy is also maintained for distance based policy, "Sq-Dist", as the system adapts sensor voltage based on the marker's proximity with respect to the camera, as shown in Fig. 16a. On the other hand, reduced resolution baseline adapts sensor resolution based on marker distance from the camera. We can see that reduced resolution baseline has a larger error than Squint, mainly because higher resolution does not help with pose tracking in dimly-lit office environments. On the other hand, Squint based imaging helps with tracking because undervolting adds brightness to the signal.

5.2 The usage of Squint is energy-efficient

We find that Squint imaging leads to significant camera power savings – up to 73% for our evaluated policies – due to aggressive undervolting. The actual amount of power saved directly depends on the voltage schedule chosen by a policy. A sparser voltage schedule with more low voltage operation, such as the one with Lightutil policy under "Bright" light, results in more savings, whereas a denser voltage schedule with more high voltage operation, such as the one with On-Demand policy under "Bright" light, results in relatively less savings. However, if we consider all lighting conditions, Lightutil has an overall higher power consumption than On-Demand, as shown in in Fig. 15b. This is because Lightutil always selects high voltage under low-light conditions, whereas On-Demand selects both high and low voltage based on scene dynamics. We observe similar trends for camera pose estimation workload as can be seen from Fig. 16b.

Fig. 17 shows the experimental power trace corresponding to two different voltages. As we can see, for a given supply voltage, the image sensor alternates between active and idle states with different timing and power profiles for each of those states. While the system switches the sensor's supply voltage, we notice only the power profiles of active/idle states getting shifted keeping their timing intact. Decreasing the active state duration through aggressive standby techniques [11] would lead to even more savings and we will deeply study such mechanisms as future work.

Meanwhile, we find that reducing capture resolution (RR) baseline also results in noticeable sensor power savings. This is because the sensor samples fewer pixels at lower resolutions, thereby reducing burden on analog readout. Image sensors typically adopt a binning mechanism for image subsampling, whereby low-resolution images are achieved by averaging pixels over a window. This means all pixels in the imaging array are activated, typically at a higher resolution, but only a few of them are readout after an averaging process based on the desired resolution. Since readout is far more power hungry than pixels, reducing resolution helps with energy efficiency. Our voltage scaling techniques could be combined with resolution adaptation techniques to compound energy savings.

System-power saving estimates: Aggressive voltage scaling in Squint can yield approximately 10% power savings for XR and IoT systems. While this may seem modest, combining Squint with end-to-end system optimization techniques such as sparse sensing holds the potential for significantly higher savings. We provide first-order estimates.

XR systems such as the Meta Quest2 headset use four cameras for inside-out tracking, consuming 120-130 mW each, totaling 500 mW for the 4000 mW [39] system, while running common use-cases such as Beat Saber. However, aggressive voltage scaling can reduce camera power consumption by an order of magnitude, bringing the total camera power down to 50 mW. With a 90% occurrence of voltage scaling, the average camera power would be around 95 mW, resulting in approximately 10% system power savings.

In TrailGuard AI [40], an IoT system widely used for wildlife tracking and anti-poaching, a motion sensor-triggered camera and low-power vision processing unit are employed. The overall system consumes around 700 mW, with the camera contributing roughly 100 mW. By applying voltage scaling, the camera's power draw can be reduced to just 10 mW. With a 90% occurrence of voltage scaling, this results in an

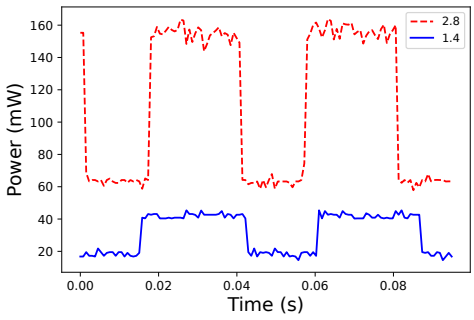


Figure 17: Power trace for sensor voltage modes.

average camera power consumption of approximately 20 mW and overall system power savings of around 11%.

5.3 The Hardware and Software extensions are Lightweight

Voltage controller is power and area efficient. Our voltage controller consumes < 1 mW of power during system operation, which is negligibly small in comparison to the camera power savings. We built our voltage controller as a standalone PCB module mainly with three components, which could be easily integrated into existing camera systems. Miniaturization efforts such as building the voltage controller into an integrated circuit form factor using the state-of-the-art fabrication technology would lead to much higher power and area efficiency.

Voltage configuration API is performant. We find that our API takes about 10 ms for switching the camera voltage. Since 10 ms is within a frame period (33 ms), the API enables the system to seamlessly switch sensor voltages on a frame-by-frame basis.

6 RELATED WORK

Dynamic voltage scaling in CPUs: The idea to dynamically scale a component's voltage has been extensively used in microprocessors since the early 1990s [6, 41, 42]. For CPUs, undervolting helps conserve processor power [43] and cools down the chip, but at the cost of performance. Overvolting, on the other hand, helps increase processor performance allowing for higher frequency operation. CPU voltage scaling is usually combined with clock frequency scaling for power and performance gains, and the mechanisms to enable such gains are popularly referred to as DVFS [44, 45] in the architecture community.

Similar to CPU undervolting, we undervolt image sensors to conserve power. Furthermore, while undervolting CPUs degrades application's performance, undervolting image sensors degrades application's imaging fidelity. Finally, unlike CPU voltage scaling techniques which adapt the voltage of *digital* circuits, our sensor voltage scaling techniques adapt the voltage of *analog* circuits for significant power savings.

Table 3: Comparison of low-power imaging systems

	Max. Res	Frame rate	Res. Reconfig	Pwr/Battery Savings
WISPcam [48]	176x144	0.001 fps	No	N/A
Battery-Free HD Video [49]	1080p	60 fps	No	1000X
Wireless Comp. Vision [50]	160x120	1 fps	Yes	62%
Glimpse [51]	480p	30 fps	No	10X
Camroptera [52]	160x120	N/A	No	N/A
Squint (Ours)	4K	60 fps	Yes	10X

Sensors for power efficiency: Designing image sensors that can operate on lower analog power supplies has been an emerging trend in the image sensors community to promote sensor's power efficiency. Samsung recently designed two such imagers [8, 9] with different resolution and pixel sizes [7], which use 2.2 V supply to power all their analog circuitry. Since lowering pixel voltage causes fidelity issues stemming from decreased voltage swing and backflow of charge into photodiodes, several circuit level changes have been done to compensate for those issues.

Many researchers in academia designed power-efficient sensors as well. Choi et al [10] designed a dual voltage-mode camera that can operate in high-fidelity imaging mode (3.3 V) and low-power vision mode (0.9 V) to save power. Osawa et al [46] designed a low-power IoT camera that has power consumption proportional to frame rate, using a hierarchical column multiplexer and variable capacitance buffer. LiKamWa et al [11] explore clock scaling and aggressive standby techniques for sensor energy efficiency. Murakami et al [47] proposes an approach to dynamically scale frequencies and voltages of image sensor's digital circuits within a frame to save power during vertical blanking period.

Similar to these imagers, we also lower the sensor's analog voltage to save power consumption. Different from these imagers, our undervolting mechanisms do not need explicit changes to the sensor design and can be directly applied on off-the-shelf cameras.

Systems for efficient visual computing: Many researchers proposed visual systems that can programmatically configure different sensors' parameters to achieve desirable system trade-offs. Stagioni [53] adapts the temperature of a 3D stacked sensor by varying the amount of near-sensor processing, to trade thermal noise with system energy. Stagioni migrates the vision task between near-sensor and far-sensor VPU in a duty-cycled fashion to regulate temperature.

Banner [54] adapts sensor resolution on a frame-level granularity to trade task accuracy with system energy, by avoiding repeated memory allocation. Josephson et al. [50]

adapt imaging resolution based on situational needs in addition to low-power backscatter streaming, improving the battery efficiency by 62%, as shown in Table 3. Naderiparizi et al. [48, 49] built a battery-free video streaming system by eliminating expensive ADC and codecs from camera module. Instead of adapting sensor resolution, Glimpse [51] and Camaroptera [52] selectively captures and processes key frames and discards the rest.

Rhythmic Pixel Regions [55] adapts sensor resolution on a region-level granularity for fine-grained accuracy-energy trade-offs, through a encoder-decoder architecture. Our work is in the similar spirit as these works but we reconfigure the sensor's analog voltage to trade sensor energy with imaging fidelity.

7 FUTURE DIRECTIONS

Network friendliness: Many IoT systems deployed in the wild are constrained by network bandwidth and use ultra-low bandwidth protocols such as LoRA to communicate between edge device and server. Our voltage reconfiguration API requires only a few bytes of data per request, thereby not burdening the network interface. To make our system even more network friendly, we will explore better data transmission protocols and intelligent ways to configure voltage. For instance, the system could occasionally send voltage requests that hold for a bunch of frames instead of sending requests on a per-frame basis.

Voltage scheduling policies: While we mainly explore policies that generate coarse-grained voltage schedules enabling fidelity control on a per-frame basis, our system allows generating fine-grained voltage schedules enabling fidelity control on a per-line basis within a frame. Such fine-grained schedules have to be generated in tandem with rolling shutter based camera capture schedules to allow sampling different parts of the frame at different desirable fidelities, which could be interesting for applications such as semantic segmentation. This fine-grained voltage scheduling would also lead to more sensor power savings, and we will investigate policies that would enable such control as future work.

Our policies also operate the sensor at extreme voltage levels as there is no incentive to operate the sensor at an intermediate voltage level for vision task. However, as mentioned in the characterization section, intermediate voltage level might be useful in low-light situations whereby undervolting adds brightness to the signal without adding too much noise. We will investigate the usefulness of intermediate voltage control in applications that can tolerate some accuracy loss as future work.

Image restoration: In this work, we focus on capturing raw sensor imagery without extensive image processing on undervolted images. However, we recognize the potential of AI-based image restoration techniques, particularly denoising, in mitigating undervolting artifacts. It is essential to consider the trade-offs associated with these denoising techniques, as they tend to be resource-intensive in terms of power consumption and latency, especially when deployed on mobile devices. We will study the quality and system implications of denoising techniques as future work.

Dynamic frequency scaling: While we exclusively focus on studying the effects of voltage scaling on the sensor's analog circuitry in this work, as a natural extension, we would like to study the power and performance implications of clock frequency scaling on the sensor's digital circuitry. Notably, we would want to explore how analog voltage scaling would work in tandem with digital clock scaling to provide desired fidelity and frame rate needed for vision applications.

Voltage scaling on other camera types: In this work, we extensively studied the implications of voltage scaling on 2D image sensors. We will extend this study to other sensor modalities such as depth cameras and event cameras. Specifically, we will study how voltage scaling would affect different sensing and readout patterns of off-the-shelf depth and event cameras in the context of emerging 3D vision and eye tracking applications, respectively.

8 CONCLUSION

Existing visual computing systems supply fixed voltage to power the sensor's analog circuitry, limiting the sensor's energy efficiency. Squint overcomes the limitation by allowing systems to aggressively undervolt sensor's analog voltage for significant efficiency gains. We build system support to dynamically adapt sensor voltage from vision applications through a programmable voltage controller hardware and a voltage configuration API software. With our hardware and software interfaces, we demonstrate the efficacy of voltage scaling for image sensors through different scheduling policies. We foresee our work as early steps towards imaging-aware voltage scheduling techniques for IoT and AR systems.

ACKNOWLEDGEMENTS

We thank Andrew Berkovich, Reid Pinkham, and Song Chen for insightful discussions on how voltage scaling deeply affects analog circuits in cameras. They provided an image sensor designer perspective that helped us write a strong motivation section. We also thank the anonymous reviewers and the paper shepherd for their detailed feedback. This material is based upon work supported by the National Science Foundation under grants 1909663 and 1942844.

REFERENCES

[1] Deepak Vasisht, Zerina Kapetanovic, Jongho Won, Xinxin Jin, Ranveer Chandra, Sudipta Sinha, Ashish Kapoor, Madhusudhan Sudarshan, and

Sean Stratman. {FarmBeats}: An {IoT} platform for {Data-Driven} agriculture. In *14th USENIX Symp. on Networked Systems Design and Implementation (NSDI 17)*, 2021.

[2] Mauricio Delbracio, Damien Kelly, Michael S. Brown, and Peyman Milanfar. Mobile computational photography: A tour. *Annual Review of Vision Science*, 2021.

[3] Cindy S Bick, Inhee Lee, Trevor Coote, Amanda E Haponski, David Blaauw, and Diarmaid Ó Foighil. Millimeter-sized smart sensors reveal that a solar refuge protects tree snail partula hyalina from extirpation. *Nature's Communications Biology*, 2021.

[4] Vikram Iyer, Ali Najafi, Johannes James, Sawyer Fuller, and Shyamnath Gollakota. Wireless steerable vision for live insects and insect-scale robots. *Science robotics*, 2020.

[5] Tom's guide. Magic Leap 2 release date, price and all the new features. https://www.tomsguide.com/news/magic-leap-2-everything-weve-heard-so-far.

[6] Mark Weiser, Brent Welch, Alan Demers, and Scott Shenker. Scheduling for reduced cpu energy. In *Mobile Computing*, pages 449–471. Springer, 1994.

[7] Image Sensors World. Era of 2.2V Supply Coming? https://image-sensors-world.blogspot.com/2021/09/era-of-22v-supply-coming.html.

[8] Image Sensors World. Samsung Announces 200MP 0.64um Pixel Mobile Sensor. https://image-sensors-world.blogspot.com/2021/09/samsung-announces-200mp-064um-pixel.html.

[9] Image Sensors World. Samsung Announces 50MP 1um Pixel Sensor with All-Direction PDAF. https://image-sensors-world.blogspot.com/2021/09/samsung-announces-50mp-1um-pixel-sensor.html.

[10] Jaehyuk Choi, Jungsoon Shin, Dongwu Kang, and Du-Sik Park. Always-on cmos image sensor for mobile and wearable devices. *IEEE Journal of Solid-State Circuits (JSSC)*, 2015.

[11] Robert LiKamWa, Bodhi Priyantha, Matthai Philipose, Lin Zhong, and Paramvir Bahl. Energy characterization and optimization of image sensing toward continuous mobile vision. In *Proc. of the 11th annual international conf on Mobile systems, applications, and services*, 2013.

[12] Kazuya Kitamura, Toshihisa Watabe, Takehide Sawamoto, Tomohiko Kosugi, Tomoyuki Akahori, Tetsuya Iida, Keigo Isobe, Takashi Watanabe, Hiroshi Shimamoto, Hiroshi Ohtake, et al. A 33-megapixel 120-frames-per-second 2.5-watt cmos image sensor with column-parallel two-stage cyclic analog-to-digital converters. *IEEE Transactions on Electron Devices*, 2012.

[13] Jaehyuk Choi, Seokjun Park, Jihyun Cho, and Euisik Yoon. An energy/illumination-adaptive cmos image sensor with reconfigurable modes of operations. *IEEE Journal of Solid-State Circuits*, 2015.

[14] Isao Takayanagi, Miho Shirakawa, Koji Mitani, Masayuki Sugawara, Steinar Iversen, Jørgen Moholt, Junichi Nakamura, and Eric R Fossum. A 1.25-inch 60-frames/s 8.3-m-pixel digital-output cmos image sensor. *IEEE Journal of Solid-State Circuits*, 2005.

[15] Che-I Lin, Cheng-Hsiao Lai, and Ya-Chin King. A four transistor cmos active pixel sensor with high dynamic range operation. In *Proceedings of Asia-Pacific Conference on Advanced System Integrated Circuits*. IEEE, 2004.

[16] Eric R Fossum and Donald B Hondongwa. A review of the pinned photodiode for ccd and cmos image sensors. *IEEE Journal of the electron devices society*, 2014.

[17] Sony. IMX219PQH5-C. https://www.arducam.com/downloads/modules/RaspberryPi_camera/IMX219DS.PDF.

[18] OmniVision. OV5647 datasheet. https://www.uctronics.com/download/Image_Sensor/OV5647_DS.pdf.

[19] OmniVision. PYTHON 1.3/0.5/0.3 MegaPixels Global Shutter CMOS Image Sensors. https://www.onsemi.com/pdf/datasheet/noip1sn1300a-d.pdf.

[20] R Jacob Baker. *CMOS: circuit design, layout, and simulation*. John Wiley & Sons, 2019.

[21] John L Hennessey and David A Patterson. *Computer Architecture: A Quantitative Approach*. Elsevier, 2018.

[22] Albert Theuwissen. How To Measure Temporal Noise in Dark ? https://harvestimaging.com/blog/?p=994.

[23] Albert Theuwissen. How to measure the average dark signal ? https://harvestimaging.com/blog/?p=795.

[24] Marc Ebner. Combining white-patch retinex and the gray world assumption to achieve color constancy for multiple illuminants. In *Joint Pattern Recognition Symposium*, pages 60–67. Springer, 2003.

[25] M Ebner. The gray world assumption. *Color Constancy; John Wiley & Sons: Chichester, UK*, pages 106–108, 2007.

[26] Dongliang Cheng, Dilip K Prasad, and Michael S Brown. Illuminant estimation for color constancy: why spatial-domain methods work and the role of the color distribution. *JOSA A*, 2014.

[27] Joost Van De Weijer, Theo Gevers, and Arjan Gijsenij. Edge-based color constancy. *IEEE Transactions on image processing*, 2007.

[28] Vadim Ziyadinov and Maxim Tereshonok. Noise immunity and robustness study of image recognition using a convolutional neural network. *Sensors*, 2022.

[29] Sony. IMX219 datasheet. https://github.com/rellimmot/Sony-IMX219-Raspberry-Pi-V2-CMOS/blob/master/RASPBERRY%20PI%20CAMERA%20V2%20DATASHEET%20IMX219PQH5_7.0.0_Datasheet_XXX.PDF.

[30] Micheal Promonet. v4l2rtspserver. https://github.com/mpromonet/v4l2rtspserver.

[31] Microchip. MCP4725, 12-Bit Digital-to-Analog Converter with EEPROM Memory in SOT-23-6. https://ww1.microchip.com/downloads/aemDocuments/documents/OTH/ProductDocuments/DataSheets/22039d.pdf.

[32] Texas Instruments. OPA55x High-Voltage, High-Current Operational Amplifiers. https://www.ti.com/lit/gpn/OPA551.

[33] Wikipedia. Representational state transfer. https://en.wikipedia.org/wiki/Representational_state_transfer.

[34] Roy Thomas Fielding. *Architectural styles and the design of network-based software architectures*. University of California, Irvine, 2000.

[35] Wikipedia. Amazon Go. https://en.wikipedia.org/wiki/Amazon_Go.

[36] Wikipedia. Microsoft HoloLens. https://en.wikipedia.org/wiki/Microsoft_HoloLens.

[37] Frank Dehmelt. Adaptive (dynamic) voltage (frequency) scaling—motivation and implementation. *Texas Instruments – Application Report*, pages 1–10, 2014.

[38] Mohamed Elgebaly and Manoj Sachdev. Variation-aware adaptive voltage scaling system. *IEEE Tran. on Very Large Scale Integration (VLSI) Systems*, 2007.

[39] THE HOME HACKS DIY. How Much Power (Watts) Does a VR Headset Use? https://www.thehomehacksdiy.com/how-much-power-watts-does-a-vr-headset-use.

[40] TrailGuard-AI-in-the-News. TrailGuard AI. https://www.resolve.ngo/blog/TrailGuard-AI-in-the-News.htm.

[41] Anantha P Chandrakasan, Samuel Sheng, and Robert W Brodersen. Low-power cmos digital design. *IEICE Transactions on Electronics*, 1992.

[42] Mark A Horowitz. Self-clocked structures for low power systems. *ARPA semi-annual report*, 1993.

[43] Masakatsu Nakai, Satoshi Akui, Katsunori Seno, Tetsumasa Meguro, Takahiro Seki, Tetsuo Kondo, Akihiko Hashiguchi, Hirokazu Kawahara, Kazuo Kumano, and Masayuki Shimura. Dynamic voltage and frequency management for a low-power embedded microprocessor. *IEEE journal of solid-state Circuits*, 2005.

[44] Jan M Rabaey, AP Chandrakasan, and B Nikolic. Digital integrated circuits, vol. 996, 1996.

[45] Wonyoung Kim, Meeta S Gupta, Gu-Yeon Wei, and David Brooks. System level analysis of fast, per-core dvfs using on-chip switching regulators. In *IEEE 14th Int Symp. on High Performance Computer Architecture*, 2008.

[46] Masato Osawa, Shuzo Hiraide, Shunsuke Suzuki, Hideki Kato, Kosei Tamiya, Yasunari Harada, Kuba Raczkowski, Jean-Luc Bacq, Peter Van Wesemael, Mingxu Liu, et al. An adaptive frame image sensor with fine-grained power management for ultra-low power internet of things application. In *IEEE 44th European Solid State Circuits Conference (ESSCIRC)*, 2018.

[47] Hirotaka Murakami, Eric Bohannon, John Childs, Grace Gui, Eric Moule, Katsuhiko Hanzawa, Tomofumi Koda, Chiaki Takano, Toshimasa Shimizu, Yuki Takizawa, et al. A 4.9 mpixel programmable-resolution multi-purpose cmos image sensor for computer vision. In *IEEE Int. Solid-State Circuits Conference (ISSCC)*, 2022.

[48] Saman Naderiparizi, Aaron N Parks, Zerina Kapetanovic, Benjamin Ransford, and Joshua R Smith. Wispcam: A battery-free rfid camera. In *IEEE Intl. Conf on RFID (RFID)*, 2015.

[49] Saman Naderiparizi, Mehrdad Hessar, Vamsi Talla, Shyamnath Gollakota, and Joshua R Smith. Towards {Battery-Free}{HD} video streaming. In *15th USENIX Symp. on Networked Systems Design and Implementation (NSDI)*, 2018.

[50] Colleen Josephson, Lei Yang, Pengyu Zhang, and Sachin Katti. Wireless computer vision using commodity radios. In *Proc. of the 18th Intl. Conf. on Information Processing in Sensor Networks (IPSN)*, 2019.

[51] Saman Naderiparizi, Pengyu Zhang, Matthai Philipose, Bodhi Priyantha, Jie Liu, and Deepak Ganesan. Glimpse: A programmable early-discard camera architecture for continuous mobile vision. In *Proc. of the 15th Annual Intl Conf. on Mobile Systems, Applications, and Services (MobiSys)*, 2017.

[52] Matteo Nardello, Harsh Desai, Davide Brunelli, and Brandon Lucia. Camaroptera: A batteryless long-range remote visual sensing system. In *Proc. of the 7th Intl. Workshop on Energy Harvesting & Energy-Neutral Sensing Systems*, 2019.

[53] Venkatesh Kodukula, Saad Katrawala, Britton Jones, Carole-Jean Wu, and Robert LiKamWa. Dynamic temperature management of near-sensor processing for energy-efficient high-fidelity imaging. *Sensors*, 2021.

[54] Jinhan Hu, Alexander Shearer, Saranya Rajagopalan, and Robert LiKamWa. Banner: An image sensor reconfiguration framework for seamless resolution-based tradeoffs. In *Proc. of the 17th Annual Int Conf on Mobile Systems, Applications, and Services*, 2019.

[55] Venkatesh Kodukula, Alexander Shearer, Van Nguyen, Srinivas Lingutla, Yifei Liu, and Robert LiKamWa. Rhythmic pixel regions: multi-resolution visual sensing system towards high-precision visual computing at low power. In *Proceedings of the 26th ACM International Conference on Architectural Support for Programming Languages and Operating Systems*, 2021.