Pixel-wise programmability enables dynamic high-SNR cameras for high-speed microscopy

Jie Zhang*[1,3], Jonathan Newman[1,3], Zeguan Wang[2,3], Yong Qian[2,3], Wei Guo[1,3], Zhe Sage Chen[4], Changyang Linghu[5], Ralph Etienne-Cummings[6], Eric Fossum[7], Edward Boyden[2,3], and Matthew Wilson[1,3]

[1]Picower Institute for Learning and Memory, MIT, Cambridge, MA, USA
[2]McGovern Institute for Brain Research, MIT, Cambridge, MA, USA
[3]Department of Brain and Cognitive Sciences, MIT, Cambridge, MA, USA
[4]Department of Psychiatry, NYU Grossman School of Medicine, New York, NY, USA
[5]Department of Cell and Developmental Biology, University of Michigan, Ann Arbor, MI, USA
[6]Department of Electrical and Computer Engineering, Johns Hopkins University, Baltimore, MD, USA
[7]Thayer School of Engineering, Dartmouth College, Hanover, NH, USA
*Correspondence to: jzhang41@mit.edu

Abstract

High-speed widefield fluorescence microscopy has the potential to capture biological processes with exceptional spatiotemporal resolution. However, conventional cameras suffer from low signal-to-noise ratio (SNR) at high frame rates, limiting their ability to detect faint fluorescent events. Here we introduce an image sensor where each pixel has individually programmable sampling speed and phase, so that pixels can be arranged to simultaneously sample at high speed with a high SNR. In high-speed voltage imaging experiments, our image sensor significantly increases the output SNR compared to a low-noise scientific CMOS camera (~2-3 folds). This SNR gain enables the detection of weak neuronal action potentials and subthreshold activities missed by the standard scientific CMOS cameras. Our proposed camera with flexible pixel exposure configurations offers versatile sampling strategies to improve signal quality in various experimental conditions.

Main text

Widefield fluorescence microscopy enables direct observation of physiological processes, such as cell signaling and local chemical concentrations. In neuroscience, this method has revolutionized the study of the neural basis of behavior by capturing the dynamics of thousands of cells in behaving animals[1]. However, challenges still remain in using functional fluorescence microscopy to sample fast neural dynamics[2–6], such as the membrane potential that reflects neural activity occurring at millisecond timescales. High sampling speeds (>1 kHz) of fluorescent indicators increase pixel readout noise and limit the number of fluorescent photons integrated during each sampling period. These factors lead to a low signal-to-noise ratio (SNR) for detecting fluorescent activities associated with spiking events and subthreshold activity. High-speed imaging setups can use high excitation power to maintain the SNR, but suffer from many caveats such as photobleaching, heat, and cytotoxicity. Currently, experiments

at kHz camera speed were conducted only in head-fixed animals using high-performance but bulky microscopes and were limited to a few minutes in duration[5,6], which was insufficient to observe many behavioral states and their associated neural mechanisms.

The SNR limitation of image sensors at high speed stems from the trade-off at each pixel. Fast sampling speeds lead to high readout noise, shortened exposure time, and fewer collected photons, all of which will inevitably lower the SNR. Despite this trade-off, physiological signals, such as the membrane voltage of a cell soma, are often spatially extended, and the same signal can be captured by multiple pixels within a region of interest (ROI). This spatial redundancy makes it possible to use multiple pixels at varying speeds and phases in a complementary manner to sample the ROI with both high SNR and high temporal resolution. However, the standard CMOS image sensors do not have flexible pixel-level exposure programmability to allow such configuration. Instead, all the pixels must have the same exposure and are integrated concurrently (global shutter) or sequentially in lines (rolling shutter). Such inflexibility prevents the sensor from utilizing the spatial redundancy of the scene to overcome the SNR speed trade-off inherent to individual pixels.

In our image sensor, the programmable pixel-wise exposure CMOS (PE-CMOS), individual pixels can be separately exposed. This unique feature allows for various sampling speeds and phases at each pixel, making it possible to coordinate their configurations to optimize the signal quality of the ROI. To illustrate the advantage of the PE-CMOS compared to a CMOS sensor, we showed a simulated example of their expected responses at sampling a fluorescent signal from a neuron expressing genetically encoded voltage indicators (GEVI) (**Fig. 1A**). The CMOS sensor with uniform exposure suffers from the SNR-speed trade-off, as high frame rates result in low SNR and low frame rates lead to signal aliasing (**Fig. 1B**). The PE-CMOS can utilize pixel-wise programmability to optimize the pixels' sampling pattern within an ROI. In one configuration, the PE-CMOS staggers pixels exposure in time to acquire fast-spiking events at multiple phases (**Fig. 1C**). This approach enhanced the ROI's temporal resolution without increasing the sampling rate or compromising the SNR. In another configuration, the PE-CMOS sampled the ROI with pixels at different speeds, capturing high-frequency events (spiking activity) and weak signals (subthreshold potentials) that are difficult to acquire simultaneously at a fixed frame rate (**Fig. 1C**).

The PE-CMOS sensor design enables pixel-wise programmability without compromising pixel sensitivity, which is determined by the percentage of effective photodiode area within each pixel (fill factor). Prior approaches to achieve pixel configurability required a high number of pixel-level circuits[7–9] (~25 transistors), which occupied valuable pixel area and reduced the photodiode fill factor (38% and 45% for pixel pitches of 12.6 and 11.2 μm, respectively[8,9]). In contrast, the PE-CMOS achieves pixel-wise programmability using only six transistors per pixel, enabling high sensitivity (75% fill factor for an even smaller pixel pitch of 10 μm), comparable to high-performance CMOS sensors (**Fig. 1D, E, Fig. S1** and **Methods**). PE-CMOS allows for nearly instantaneous updates of pixel configurations via on-chip control, enabling the possibility to optimize local spatiotemporal resolution and SNR in real time based on various sampling patterns (**Fig. S2**).

We compared the PE-CMOS performance with a scientific CMOS camera (Hamamatsu Orca Flash 4.0 v3 sCMOS) at sampling high-speed fluorescent signals. Both sensors were used to simultaneously image (through a 50/50 beamsplitter) the spontaneous activity of cultured mouse hippocampal neurons expressing the ASAP3 GEVI protein[3] (**Fig. 2**). In terms of temporal resolution, it is essential to obtain a minimum of three samples to accurately represent the shape of a spike over time, with one taken at the peak and two at the troughs. Given the half-width of the action potential generated ASAP3 spikes (~6 ms), the sampling period must be less than 2 ms. We, therefore, set the sCMOS camera to sample the signal at 800 Hz with a 1.25 ms exposure duration. For the PE-CMOS camera, we used a longer exposure time of 5 ms to integrate over the full spike width for achieving an improved SNR. To ensure sufficient temporal resolution in capturing the spikes, we staggered the pixel sampling with phase offsets in multiples of 1.25 ms (**Fig. 1C**). Although each pixel has a low sampling speed of 200 Hz, the PE-CMOS can acquire the ROI at an equivalent of 800 Hz with pixels at different phases.

Before comparing the PE-CMOS and sCMOS's SNR performance, we first show that the staggered PE-CMOS sampling configuration achieves high temporal resolution even with a low per-pixel sampling rate. This is seen directly at the PE-CMOS outputs of different phases (**Fig. 2C**). With a 5 ms exposure and a slow sampling rate of 200 Hz, individual pixels are at a sub-Nyquist rate to sample ASAP3 spikes with 6 ms half-width. This could lead to aliasing, causing the samples to miss the spike at its peak (one example is indicated by the black arrow in **Fig. 2C**, under phase $\Delta = 0$ ms). Nevertheless, by applying various phase shifts to the surrounding pixels, the spike exceeding the temporal resolution of a single phase was guaranteed to be captured by phase-shifted pixels (**Fig. 2C,** black arrow, phase $\Delta = 1.25$ ms). We then interpolated across the ROI and obtained an 800 Hz equivalent time series capturing all the spikes (**Fig. 2C**). Multi-phasic sampling's ability to resolve aliasing was further recapitulated in a benchtop experiment, where phase-shifted PE-CMOS pixels sampled at 250 Hz accurately captured a fast spike train (8 ms inter-spike interval) at a 1000 Hz temporal resolution (**Fig. S3, S4,** and **Methods)**.

The ability to achieve high temporal resolution with lower sampling rate allows PE-CMOS to use prolonged pixel exposure to improve the SNR. The PE-CMOS outperforms the sCMOS with superior SNR under the same temporal resolution. The SNR performance difference is particularly evident in the PE-CMOS ability to detect low-SNR subthreshold potentials and spiking events. One example illustrates this advantage, as a spike is distinctly visible in the PE-CMOS signal (indicated by the magenta arrow in **Fig. 2F**), while the same event is indiscernible in the simultaneously recorded sCMOS ROI time series due to a low SNR (**Fig. 2E, F**). Moreover, the PE-CMOS outputs provide a higher SNR when capturing weak subthreshold activities. An example is marked by the black circle in **Fig. 2B, C**. Furthermore, PE-CMOS maintained superior SNR performance over the sCMOS even in smaller ROIs covering parts of the cell membrane (**Fig. 2G, H**). Many events with sufficient SNR (> 5) in the PE-CMOS outputs were obscured by the noise floor in the sCMOS outputs (**Fig. 2H**). To quantify the spike SNR difference between the PE-CMOS and sCMOS, we isolated the putative spiking events at either sensor's output (>5 SNR) and compared their SNR (**Fig. 2J,** Wilcoxon rank-sum test for equal median, $n = 93$ events, $p < 10^{-8}$). The putative spike events obtained

from the PE-CMOS consistently have higher SNR than that obtained from the sCMOS, with an average SNR gain of a factor of two **(Fig. 2J**).

The superior SNR performance of the PE-CMOS compared to the sCMOS is attributed to the optimally chosen pixel exposure. We modeled the relationship between pixel exposure and SNR using the known GEVI time constant of an action potential (**Fig. S5A** and **Methods**). The pixel output is then computed as the area under the GEVI spike, and the major noise sources are photon shot noise and circuit readout noise (**Fig. S5A** and **Methods**). Pixel SNR increases with exposure, attributed to two reasons: first, the photon shot noise is reduced relative to the signal as more photons are integrated by the pixel, and second, the readout noise decreases due to a slower sampling rate (**Fig. S5B** and **Methods**). The most optimal exposure to maximize SNR is around 4 – 8 ms, roughly matching the width of a GEVI spike. Pixel exposure extending longer than the spike width has a diminishing benefit on the SNR. To sample ASAP3 spikes, our model predicted that increasing exposure from 1.25 to 5 ms corresponds to ~2-3 fold of spike SNR improvement depending on the pixel read noise, base fluorescent, and the fluorescence signal dF/F (**Fig. S5C**). The model prediction matched our empirical measurements (**Fig. 2J**), where the PE-CMOS (T_E = 5 ms) results in a spike SNR gain of greater than two compared to the sCMOS (T_E = 1.25 ms). This is a significant signal quality improvement without increasing excitation power and allows for extracting meaningful physiological signals that conventional sCMOS cameras cannot.

The PE-CMOS can also be used to optimize pixel sampling patterns based on temporal features of the ROI. As we demonstrated, the optimal exposure to maximize SNR depends on the signal's shape (**Fig. S5**). However, an ROI may contain events of diverse temporal characteristics. For example, a neuron may exhibit transient spiking events and subthreshold activities of different intensities and frequencies. In such cases, using a fixed pixel exposure cannot ensure the optimal detection of all signals representing important and distinct physiological mechanisms. A better strategy is to sample the pixels of the ROI at different exposure and speeds. To illustrate this, we configured the PE-CMOS to sample the ASAP3 activity of a cell soma with pixel exposure ranging from 15.4 ms down to 1.9 ms (**Fig. 3A, B**). This configuration samples an ROI at 64-520 Hz, aiming to simultaneously capture fast spike bursts, low SNR spikes, and subthreshold activities. The ROI time series showed that fast pixels (520 Hz) can separate high-SNR burst events in time but cannot reliably detect slower and weaker GEVI activities due to low SNR (**Fig. 3D, E, F**). On the other hand, pixels sampling at moderate (260 Hz) and slow (<130 Hz) speed improved the SNR for low-frequency signals and detected low SNR candidate events for subthreshold activities or low SNR spikes. Combining pixels at different speeds, we reconstructed a 1040 Hz equivalent ROI time series through interpolation (**Methods**), capturing high-SNR spike bursts, low SNR spikes, and weak slow-varying subthreshold activities using a 260 Hz sensor readout speed **(Fig. 3C-F**).

The advantages of programmable pixel-wise exposure extend beyond fluorescence microscopy to general imaging applications, particularly in capturing scenes with a high dynamic range of motion and light intensity. For example, the time-staggered pixel pattern (**Fig. 1C**) can track high-speed motion in low light conditions with long exposures. For instance, the PE-CMOS sensor with 25 Hz sample rate resolves a motion of 100 frame/s (**Fig. S6**). Another example is

the use of multiple exposure pixel patterns to capture high dynamic range scenes (**Fig. S7**), which have been traditionally difficult to acquire using a single image sensor and often require multiple sensors with additional optics[10].

In conclusion, the PE-CMOS sensor enables several promising future applications. Its CMOS-based design offers scalability to higher pixel numbers and speed to meet the demand of various heterogenous imaging applications. For instance, the PE-CMOS allows configurations of pixel exposures to optimally match the temporal characteristics of dendritic potentials evoked by different neurotransmitters[11] at various synapses. Additionally, using pixels sampled at different speeds, the PE-CMOS may enable us to simultaneously image multiple fluorescent indicators with different intensity ranges and time constants, such as calcium and GEVI signals[12]. Finally, the PE-CMOS power efficiency, resulting from its ability to achieve high temporal resolution with low sampling speed, makes it suitable for miniaturized microscopes[13–16] to enable kHz GEVI imaging from freely moving rodents. Under a sensor speed of 100 to 250 Hz, which is typically limited to capturing slow fluorescent dynamics like calcium indicators, PE-CMOS has the potential to enable GEVI imaging at 400-1000 Hz. Ultimately, this will provide an opportunity to directly image *in vivo* neural voltage activity to investigate fast-spiking activities[17] underlying complex behavior.

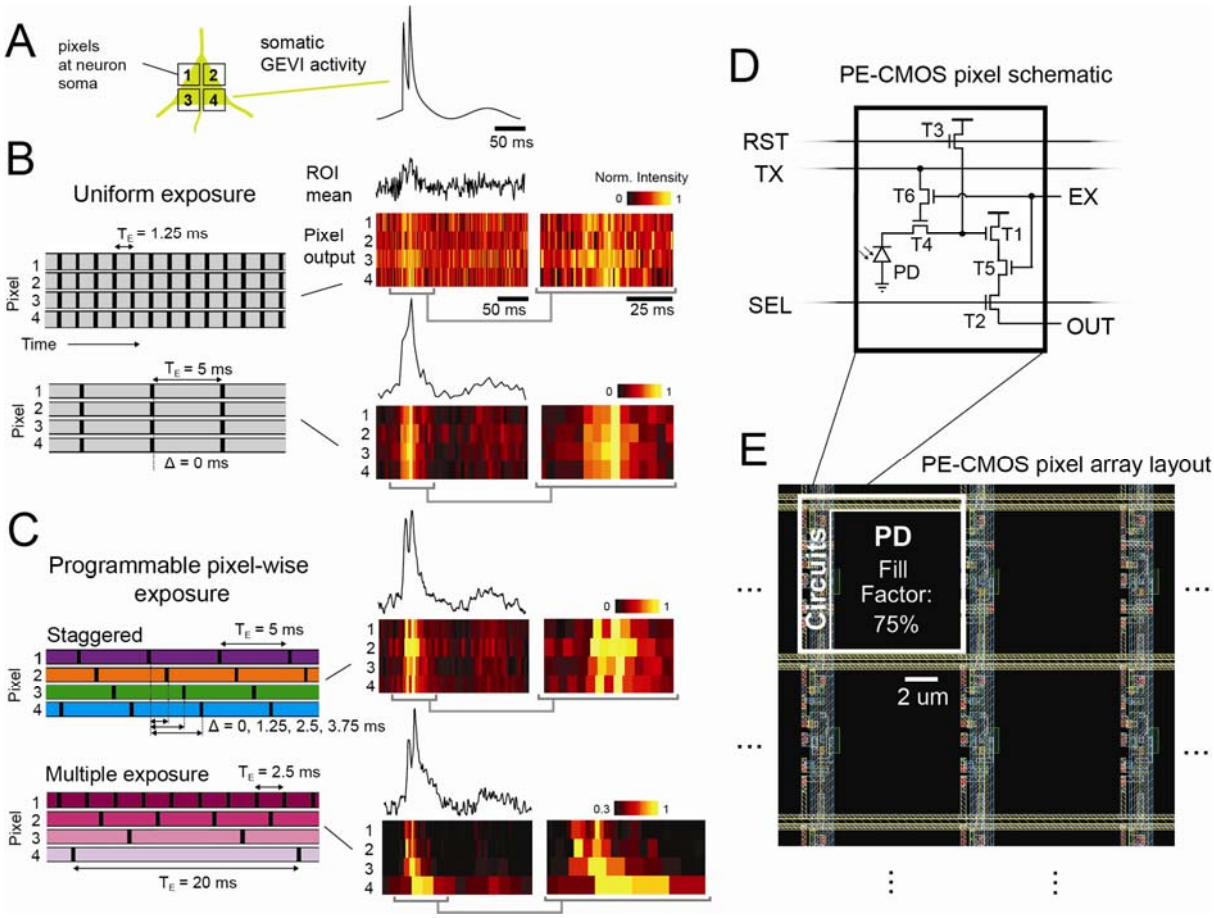


Fig. 1: Programmable pixel-wise exposure enables flexible pixel configuration to sample an ROI with high SNR and temporal resolution.

A, Pixels within an ROI capture spatiotemporally-correlated physiological activity, such as somatic GEVI signals.

B, Simulated CMOS pixel outputs with uniform exposure (T_E) face the trade between SNR and temporal resolution. Short T_E (1.25 ms) provides high temporal resolution but low SNR. Long T_E (5 ms) enhances SNR but suffers from aliasing due to low sample rate, causing spikes (10 ms interspike interval) to be indiscernible. Pixel outputs are normalized row-wise. Gray brackets: the zoomed-in view of the pixel outputs.

C, Simulated pixel outputs of the PE-CMOS. Pixel-wise exposure allows pixels to sample at different speeds and phases. Two examples: in the staggered configuration, the pixels sample the spiking activity with prolonged T_E (5 ms) at multiple phases with offsets of (Δ = 0, 1,25, 2.5, 3.75 ms). This configuration maintains SNR and prevents aliasing, as the interspike interval exceeding the temporal resolution of a single phase is captured by phase-shifted pixels. In the multiple exposure configuration, the ROI is sampled with pixels at different speeds, resolving high-frequency spiking activity and slow varying subthreshold potentials that are challenging to acquire simultaneously at a fixed sampling rate.

D-E, the PE-CMOS pixel schematic and **E,** the pixel layout. The design achieves programmable pixel-wise exposure while maximizing the photodiode (PD) fill factor for high optical sensitivity.

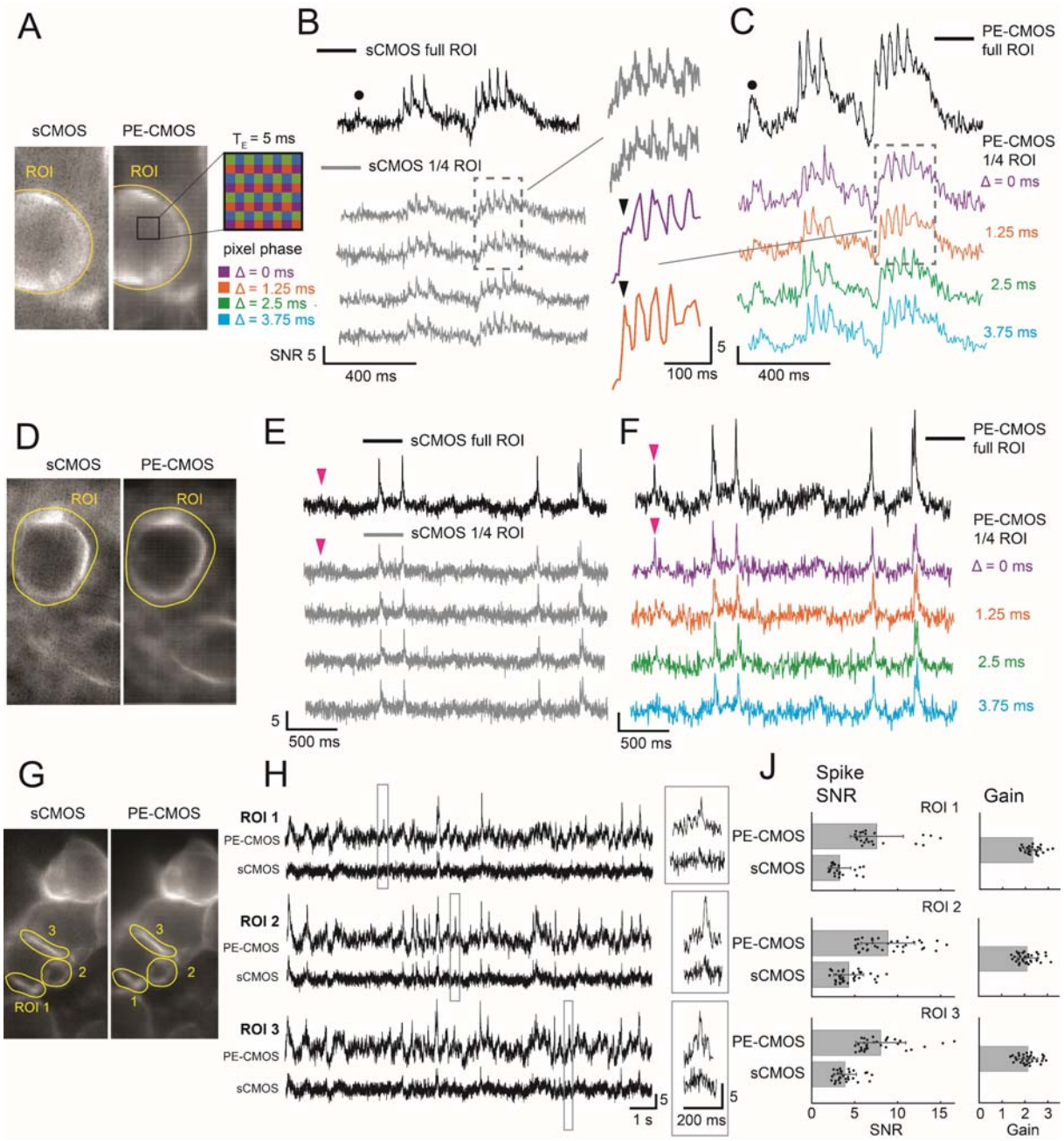


Fig. 2: Comparison of the PE-CMOS with an sCMOS camera.

A, Maximum intensity projection of the sCMOS (Hamamatsu Orca Flash 4.0 v3) and the PE-CMOS videos of a cultured neuron expressing the ASAP3 GEVI protein. All vertical scales of SNR are 5 in all subfigures.

B, ROI time series from the sCMOS sampled at 800 Hz with pixel exposure (T_E) of 1.25 ms. Black trace: ROI time series. Gray trace: the time series each with 1/4 pixels of the ROI. Plotted signals are inverted from raw samples for visualization.

C, simultaneously imaged ROI time series of the PE-CMOS. Colored trace: the time series of phase-shifted pixels at offsets (Δ) of 0, 1.25, 2.5, and 3.75 ms each contain 1/4 pixels of the ROI. All pixels are sampled at 200 Hz with T_E = 5 ms. Black trace: the interpolated ROI time series with 800 Hz equivalent sample rate. Black arrows: An example showing a spike exceeding the temporal resolution of a single phase is captured by phase-shifted pixels. Black circles: an example subthreshold event barely discernable in sCMOS is visible in the pCMOS output.

D-F: same at panels A-C with an example showing a spike captured by the PE-CMOS but not resolvable in the sCMOS output due to low SNR (marked by the magenta arrow).

G, H, comparison of signal quality from smaller ROIs covering parts of the cell membrane. Gray boxes: zoomed-in view of a few examples of putative spiking events.

J, SNR of putative spikes events from ROIs in panel G. A putative spiking event is recorded when the signals from either output exceed SNR > 5 (Wilcoxon rank-sum test for equal medians, n = 93 events,$p < 10^{-8}$). The gain is calculated as the spike SNR in the PE-CMOS divide by the SNR in the sCMOS.

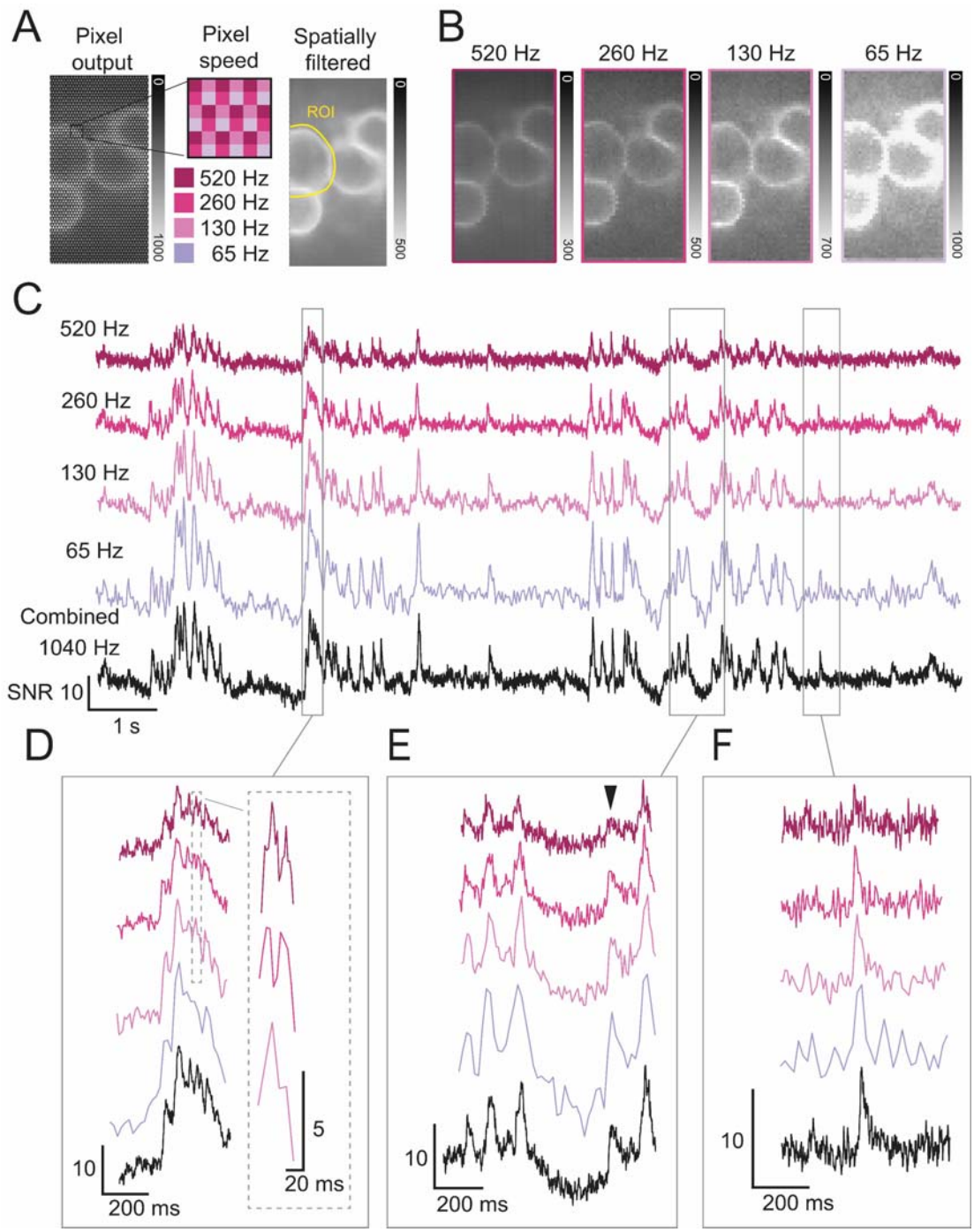


Fig. 3: Multi-exposure/speed sampling to simultaneously maximize SNR of spiking events and subthreshold activity.

A, Maximum intensity projection of the PE-CMOS videos, raw and filtered (2 x 2 box filter) output at full spatial resolution. Intensity is measured by digital bits (range: 0 – 1023).

B, Maximum intensity projection divided into four sub-frames according to pixel sampling speed, each with 1/4 spatial resolution.

C, The ROI time series from pixels of different speeds (colored trace). Black trace: a 1040 Hz equivalent signal interpolated across all ROI pixels.

D, Fast sampling pixels (520 Hz) resolves high-SNR spike bursts.

E–F, Pixels with more prolonged exposure (T_E = 2.8 – 5.7 ms) improves SNR to detect weak subthreshold activity (black arrow) and **F,** low SNR spike.

The vertical scale of SNR is 10 unless otherwise noted.

Methods

Sensor architecture and circuit design

The PE-CMOS chip was fabricated using the 180 nm CMOS image sensor process from XFAB, a commercial foundry service. Two prototypes were produced with pixel arrays containing a 6.5 µm (Chip A) and 10 µm (Chip B) pixel pitch (**Fig. S1A**). The pixel array is arranged in rows and columns, with readout circuitry at each column (**Fig. S1B**). Each column readout circuitry comprises a programmable gain amplifier (PGA), correlated-double sampling (CDS) circuits, and an analog-to-digital converter (ADC). The PGA offers a variable gain of 8 - 64x to the pixel output. The amplified signal is then processed by the CDS circuits to reduce flicker noise before being sampled by the 10-bit ADC. The row-wise signals (RST, TX, and SEL) multiplex pixels to the readout circuits at each column. These signals also control pixel reset (RST), charge transfer (TX), and row-wise multiplex (SEL) operations. In a single readout period (0.9 ms in our implementation), each column readout circuit selects pixels of the column that end the exposure and convert their outputs to digital bits that are then sent off-chip.

The PE-CMOS pixel comprises 6 transistors (T1 - T6) and one pinned photodiode (PD). T1 to T4 form the standard rolling shutter 4T-pixel design[18]. During pixel exposure, the PD converts incoming photons into electrons. When the PD is ready to be sampled, T3 (controlled by RST) first resets the voltage on the floating diffusion (FD) node. T4 (controlled by TX) then moves the electrons from the PD to FD, producing a change in voltage. The voltage signal is buffered by T1 and connected to the readout circuits by T2 (controlled by SEL), where the column readout circuitry samples it.

To enable pixel-wise exposure programmability, we incorporated two additional transistors (T5 and T6) into the PE-CMOS pixel design. T5 and T6 operate as switches. Whenever the row selection signals select K pixels for readout (K = 8 in Chip A, K = 4 in Chip B), the pixel with T5 and T6 turned on will terminate its exposure and be sampled by the column circuitry. The remaining pixels with their respective T5 and T6 off continue their exposure. These two transistors are controlled by column bus lines EX<1:K> placed at each column. Pixel-wise exposure control can be achieved by synchronizing the EX signals with row signals (RST, SEL and TX). An example pixel-wise operation diagram is shown in **Fig. S1D**.

Pixel-wise programmability allows pixels in PE-CMOS to form custom pixel-sampling patterns. Some examples are shown in **Fig. S2**. We configured the pixels into a tiled arrangement for the *in vitro* GEVI imaging experiments to provide a uniform spatiotemporal sampling of cell ROI. Every arbitrary 2 x 2 window in the tiled pattern contains four pixels sampling at four different phases or temporal resolutions (**Fig. S1A, B**). A random spatial pattern might be desirable for other applications, such as compressed sensing imaging. The PE-CMOS pixels can also be arranged in ROI dependent spatial patterns to further optimize the SNR (**Fig. S1A**). Since GEVI proteins are expressed mainly on the cell membrane, we can surround the bright outer contour of the cell ROI with fast pixels having short T_E, while using slower pixels within the dimmer region of the ROI to enhance the SNR.

Finding the optimal exposure to capture GEVI spikes

The exposure duration (T_E) of a pixel has a direct impact on its SNR, and the most optimal T_E for maximizing the SNR depends on the shape of the signal. We created a model to demonstrate how the SNR and T_E are correlated when sampling a GEVI spike. By using this model, we can select an optimal T_E that maximizes the SNR.

We first model the shape of the GEVI signal, $v(t)$, during a spiking event, which is approximated as a combination of double exponential functions[2–4]:

$$v(t) = \begin{cases} A_R e^{t/\tau_{R_fast}} + B_R e^{t/\tau_{R_slow}} & \text{when } t < 0 \\ A_F e^{-t/\tau_{F_fast}} + B_F e^{-t/\tau_{F_slow}} & \text{when } t \geq 0 \end{cases} \quad (1)$$

where τ_{Rfast} and τ_{Rslow} characterize the spike rise times at fast and slow timescales, respectively; and τ_{Ffast} and τ_{Fslow} are the respective fall time. A and B are two scaling constants of the GEVI indicators. **Fig. S5A** plots the shape of $v(t)$ with parameters of three different GEVI indicators (ASAP3[3], Archon[4], and Voltron[2]), with the spike peak normalized to 10% dF/F.

When a pixel samples $v(t)$ with an exposure duration of T_E, the resulting output, $y[n]$, is obtained as discrete samples of the convolution between $v(t)$ and the exposure, $e(t)$:

$$y(t) = e(t) * v(t), \qquad e(t) = \begin{Bmatrix} 1 & 0 \leq t \leq T_E \\ 0 & otherwise \end{Bmatrix} \quad (2)$$

$$y[n] = y(nT_E) + \sigma[n] \quad (3)$$

where $\sigma[n]$ is the sampled noise containing both shot noise and circuit read noise. We assume that the shot noise is described by a Poisson distribution with a rate factor λ equal to the signal at the pixel. As such, its power is expressed as:

$$N_{shot} = y[n] + F \cdot T_E$$

where F denotes the baseline fluorescent signal, represented here as a constant for simplicity.

The read noise of a circuit comprises two components: flicker noise and thermal noise. In pixel design, the correlated double sampling technique can effectively eliminate the flicker noise, which is more prominent at low frequencies due to its 1/f noise density profile. Therefore, we can estimate the circuit read noise using only thermal noise that has a white power spectrum over the circuit bandwidth, with the power given by:

$$N_{read} \approx N_{thermal} = N_0/T_E$$

Where N_0 is a noise density measured in V/Hz, and $1/T_E$ is the circuit bandwidth. The magnitude of thermal noise depends on the circuit bandwidth, which is inversely proportional to the exposure time of the pixel.

We can now derive an analytic expression of the peak pixel SNR with respect to T_E:

$$SNR = \sqrt{\frac{S}{N_{shot} + N_{read}}} = \sqrt{\frac{\max\{y\}^2}{N_0/T_E + \max\{y\} + F \cdot T_E}} \tag{4}$$

$$\approx \frac{\max\{y\}}{\sqrt{N_0/T_E + F \cdot T_E}} \quad \text{when } F \gg \max\{y\}$$

Here, the SNR is defined as the ratio between signal amplitude and the root-mean-squared noise. We plotted **Eq. 4** for various GEVI indicators for exposure time T_E (**Fig. S5B**, parameters: F = 20e^-, peak dF/F = 10%, N_0 = 2.6 e^-_{rms}). For comparison across different GEVI indicators, we normalize the SNR by dividing it by the peak value. For all the GEVI indicators, the SNR improves rapidly with longer T_E but drops gradually as extra integration time adds more shot noise than signal power due to the baseline fluorescent. To maximize the SNR for GEVI spikes, the most optimal T_E is 4-8 ms, approximately the half-width of the action potential.

Using phase offset pixel-wise exposures to increase temporal resolution

Maximizing the SNR requires a pixel exposure duration approximately equal to the spike width (**Fig. S5**). However, this can limit the pixel's temporal resolution and potentially create ambiguity in detecting spiking events. To demonstrate this, we conducted a benchtop experiment using a LED[19] light source to simulate a spike burst with each spike having a 4 ms half-width and an 8 ms interspike interval (**Fig. S3A**). To accurately detect this spiking event, it is necessary to have enough time samples to capture the spike shape and interspike interval. At a minimum, we need three temporal samples to capture a spike event (one sample at the peak and one on the trough on either side of the peak) or an interspike interval. However, prioritizing SNR maximization using a pixel exposure duration of 4 ms will not provide enough samples to detect the spikes in this example and their separations unequivocally. Sampling this signal at 250 Hz may result in aliasing, making it impossible to distinguish between the spikes (**Fig. S3B**).

Introducing phase shifts between pixels in PE-CMOS enables us to enhance the temporal resolution of an ROI without increasing the sampling rate of individual pixels. For example, using four pixels, we can sample the ROI time-series at four phases (**Fig. S3C**). This approach helps to overcome aliasing issues that may arise when sampling the ROI without pixel phase offsets. Spikes that are not accurately detected at one phase can be captured at phase-shifted

pixels (**Fig. S3D**). Interpolating across these pixels allows us to reconstruct spiking events at 1000 Hz resolution, with eight samples separating the spike peaks (**Fig. S3F**).

In this example, the phase-shifted sampling pattern enables us to increase the bandwidth of the ROI by a factor of four, without the need to increase the sampling speed of each individual pixel. At a sampling rate of 250 Hz, a single pixel's bandwidth is limited to 125 Hz. Frequencies above 125 Hz fold into the pixel's bandwidth and cause aliasing (**Fig. S3E**), which makes it difficult to distinguish between frequency components. By introducing phase-offset pixels, the combined outputs of four pixels sample the ROI at a much higher frequency, effectively extending the bandwidth to 500 Hz, four times that of the individual pixel, and avoiding the aliasing effects that occur with single pixels (**Fig. S3E**).

Using phase-shifted pixels enhances the bandwidth of the sampled ROI time series. However, it does not eliminate the narrowband attenuations (seen at 250 and 500 Hz in **Fig. S3E**), since these are inherent to the exposure function, $e(t)$. While these attenuations may not affect the detection of spikes with a broad frequency spectrum, they may pose a problem when the ROI contains specific signals of interest that fall into these frequency ranges. The PE-CMOS can employ pixels with varied exposure durations within the ROI (**Fig. S4**). These pixels sample the ROI at different speeds with varying $e(t)$. While certain frequencies may be attenuated by pixels with a specific $e(t)$, they are preserved by other pixels with different exposure lengths. Consequently, the frequency spectrum of the ROI displays no narrowband attenuation (**Fig. S4**).

ROI signal interpolation

We employ linear interpolation to compute the ROI time-series from the pixel outputs. For an ROI containing N pixels, we can write each pixel's output as:

$$\boldsymbol{y}_n = \boldsymbol{E}_n \boldsymbol{v}_n \qquad (5)$$

where $\boldsymbol{E}_n \in \mathbb{R}^{M \times L}$ denotes the sampling matrix representing pixel exposure and sampling operation that converts the fluorescent signal $\boldsymbol{v}_m \in \mathbb{R}^L$ into the pixel outputs, $\boldsymbol{y}_n \in \mathbb{R}^M$, of length M.

We then define the average fluorescent signal of the ROI: $\overline{\boldsymbol{v}_{ROI}} = \frac{1}{N}\sum_{n=1}^{N} \boldsymbol{v}_m$ and rewrite Eq. 5 as:

$$\boldsymbol{y}_m = \boldsymbol{E}_m \, \overline{\boldsymbol{v}_{ROI}} + \boldsymbol{E}_m(\boldsymbol{v}_{\boldsymbol{m}} - \overline{\boldsymbol{v}_{ROI}})$$

where $\boldsymbol{E}_m(\boldsymbol{v}_{\boldsymbol{m}} - \overline{\boldsymbol{v}_{ROI}})$ denotes the difference between $\boldsymbol{v}_{\boldsymbol{m}}$ and the ROI average. By definition, the sum $\sum_{n=1}^{M} |\boldsymbol{E}_m(\boldsymbol{v}_{\boldsymbol{m}} - \overline{\boldsymbol{v}_{ROI}})|$ should be minimized when $\overline{\boldsymbol{v}_{ROI}}$ is the average of the ROI.

By rewriting the difference term, $\varphi_m = \boldsymbol{E}_m(\boldsymbol{v}_{\boldsymbol{m}} - \overline{\boldsymbol{v}_{ROI}})$, we can concatenate the output of all the pixels in the ROI into one vector:

$$\begin{bmatrix} \boldsymbol{y_1} \\ \vdots \\ \boldsymbol{y_N} \end{bmatrix} = \begin{bmatrix} \boldsymbol{E_1} \\ \vdots \\ \boldsymbol{E_N} \end{bmatrix} \overline{\boldsymbol{v}_{ROI}} + \begin{bmatrix} \boldsymbol{\varphi_1} \\ \vdots \\ \boldsymbol{\varphi_N} \end{bmatrix}$$

If we define $\boldsymbol{y} = \begin{bmatrix} \boldsymbol{y_1} \\ \vdots \\ \boldsymbol{y_N} \end{bmatrix}$, $\boldsymbol{E} = \begin{bmatrix} \boldsymbol{E_1} \\ \vdots \\ \boldsymbol{E_N} \end{bmatrix}$ and $\boldsymbol{\varphi} = \begin{bmatrix} \boldsymbol{\varphi_1} \\ \vdots \\ \boldsymbol{\varphi_N} \end{bmatrix}$, the pixel outputs become:

$$\boldsymbol{y} = \boldsymbol{E}\overline{\boldsymbol{v}_{ROI}} + \boldsymbol{\varphi} \qquad (6)$$

Given Eq. 6, we can find $\overline{\boldsymbol{v}_{ROI}}$ using a ridge regression by enforcing a L_2 norm penalty:

$$\overline{\boldsymbol{v}_{ROI}} = min_{v_{ROI}} \frac{1}{2} \|y - \boldsymbol{E}\overline{\boldsymbol{v}_{ROI}}\|_2^2 + \lambda_{ridge} \|\overline{\boldsymbol{v}_{ROI}}\|_2^2$$

and further derive its closed-form solution:

$$\hat{\boldsymbol{v}}_{ROI} = \left(\boldsymbol{E}^T\boldsymbol{E} + \lambda_{ridge}\boldsymbol{I}\right)^{-1}\boldsymbol{E}^T\boldsymbol{y} \qquad (7)$$

where $\hat{\boldsymbol{v}}_{ROI}$ is the least-squared estimate of $\overline{\boldsymbol{v}_{ROI}}$, and the hyperparameter, λ_{ridge}, controls the trade-off between the L_2-norm of v_{ROI} and reconstruction error. Note that when the value of λ_{ridge} becomes very large, the term $\left(\boldsymbol{E}^T\boldsymbol{E} + \lambda_{ridge}\boldsymbol{I}\right)^{-1}$ will approximate a scaled identity matrix and **Eq. 7** becomes

$$\hat{\boldsymbol{v}}_{ROI} \approx \frac{1}{\lambda_{ridge}}\boldsymbol{E}^T\,\boldsymbol{y}$$

which is equal to up-sampling and interpolation of the pixel outputs $\boldsymbol{y}$.

Voltage imaging in dissociated neuron cultures

Neuron culture preparation and AAV transduction: All procedures involving animals were performed in accordance with National Institute of Health Guide for Laboratory Animals and approved by the Massachusetts Institute of Technology Committee on Animal. Dissociated hippocampal neurons were prepared from postnatal day 0 or 1 Swiss Webster mice (Taconic) without regard to sex following the previous protocol (https://www.nature.com/articles/nmeth.2836). Briefly, dissected hippocampal tissue was digested with 50 units of papain (Worthington Biochem) for 6 to 8 min, and the digestion was stopped with ovomucoid trypsin inhibitor (Worthington Biochem). Cells were plated at a density of 40,000 – 60,000 per glass coverslip coated with Matrigel (BD Biosciences). Neurons were seeded in 200 µl plating medium containing MEM (Life Technologies), glucose (33 mM, Sigma), transferrin (0.01%, Sigma), Hepes (10 mM, Sigma), Glutagro (2 mM, Corning), Insulin (0.13%, Millipore), B27 supplement (2%, Gibco), and heat inactivated FBS (7.5%, Corning). After cell adhesion, additional plating medium was added. AraC (0.002 mM, Sigma) was added when glia density was 50–70% of confluence. Neurons were grown at 37 °C and 5% CO_2 in a humidified atmosphere.

We transduce cultured neurons at 5 – 7 days in vitro (DIV) by administering ~ 10^{10} viral particles of AAV9-hSyn-ASAP3 (Janelia Viral Tools) per well (of 24-well plate). Voltage imaging was performed 7 to 14 days after transduction.

***Microscope setup*:** Cultured hippocampal neurons expressing ASAP3 were imaged on a customized upright fluorescent microscope with a 20x 1.0NA objective lens (Olympus). The light from a 470 nm LED (Thorlabs) was cleaned with a 469/35nm band pass filter (Semrock) for excitation. A 488nm long pass dichroic mirror and a 496nm long pass filter were used for illumination and emission. Using a 50/50 beam splitter (Thorlabs), the sample image was evenly split onto both the Hamamatsu sCMOS camera and our PE-CMOS sensor for side-by-side comparison.

***ROI time series extraction*:** We hand selected the ROI using the max projection image of the PE-CMOS and sCMOS videos. Due to differences in pixel size (PE-CMOS: 10µm pixel pitch, sCMOS: 6.5µm), sCMOS ROI contains approximately ~2.3x more pixels than that of the PE-CMOS ROI. PE-CMOS pixels are arranged in the tile configuration (**Fig. S2**).

For the sCMOS, we computed the ROI time series by averaging all the pixels values. We then filtered the ROI time-series using a band-pass filter to remove photo-bleaching effects and high-frequency content too close to the sampling Nyquist frequency (Filter parameters: 4th-order IIR, with high-pass cut-off at 0.5 Hz and low-pass cut-off at 360 Hz, 90% of the Nyquist bandwidth, 400 Hz, half of pixel sampling speed). For visualization purpose, we invert the time series as ASAP3 GEVI has negative going action potential waveforms.

For the PE-CMOS, we first computed sub-ROI time series for pixels with the same T_E and Δ. We then filtered each sub-ROI time series using a band-pass filter to remove photo-bleaching effects and high-frequency content too close to the sampling Nyquist frequency. The filter parameters used in **Fig. 2** are: 4th-order IIR, with high-pass cut-off at 0.5 Hz and low-pass cut-off at 90 Hz, 90% of the pixel Nyquist bandwidth 100 Hz. The same filter is used in **Fig. 3**, but with low-pass cut-off changed to 234 Hz, 117 Hz, 58 Hz, and 29 Hz, corresponding to different pixel Nyquist bandwidth 260, 130, 65 and 33 Hz. The ROI time series representing the ROI average are interpolated from sub-ROI time-series using method described earlier. No additional filters are used for the PE-CMOS ROI time-series.

To compare the sCMOS output to that of the PE-CMOS at different Δ in **Fig. 2**, which compose of ~1/4 pixels of the ROI, we computed four sCMOS sub-ROI time series each using only 1/4 ROI pixels extracted through spatial downsampling. The average of these 4 sub-ROI time series is equal to the ROI time series.

***Noise estimation*:** To estimate the noise of the times series, we find a section of the signal with at least 2 seconds in duration that does not contain any spikes. We then high-pass filter this trace to remove slow subthreshold oscillations (Filter parameters: 4th-order IIR, f_{cutoff} = 50 Hz).

The root-mean squared (rms) noise is then calculated as the standard deviation of the filtered time-series. This value is referred to as 1 SNR in the figures and text.

***Putative spike detection*:** ROI times-series are first high-pass filtered to remove high amplitude oscillations at low frequency (Filter parameters: 4th IIR filter, f_{cutoff} = 2 Hz). From the filtered PE-CMOS time series, we detected events with SNR > 5 and record their SNR and time of occurrence. We then record the highest SNR in the sCMOS times series within the corresponding time window. We set a 30 ms window centering at the peak time of PE-CMOS to avoid mismatching the peak SNR between PE-CMOS and sCMOS time series.

General imaging applications using PE-CMOS

We demonstrated two general imaging examples that benefits from using the PE-CMOS. In the first example, the PE-CMOS pixels are configured to have fixed exposure with T_E = 40 ms and Δ = 0, 10, 20, 30 ms (**Fig. S6A**) This corresponds to pixel sampling speed of 25 Hz. We imaged the motion of bouncing ball with static background (**Fig. S6B**). Staggered exposure embeds the motion within full resolution frames (**Fig. S6C**). We split each frame into four subframe organized according to pixel phase, which yields a 100 Hz equivalent video at 1/4 spatial resolution. From the high-speed video, we can track the motion of the bouncing ball with temporal precision of 10 ms (**Fig. 2E**), which is not possible using a conventional camera where pixels are uniformly sampled at 25Hz.

In the second examples, the PE-CMOS is configured to capture high-dynamic range videos using pixels of different exposures and speeds (**Fig. S7**). Within a 2 × 2 patch, pixels are configured to have exposure ranging from T_E = 8, 16, 32 and 64 ms (**Fig. S2A**). This results in a multi-spatiotemporal sampling of the scene at 125, 63, 31 and 16 Hz, simultaneously capturing fast motion and enhancing the SNR of the dark background regions (**Fig. S7B**). We can computationally reconstruct an HDR frame from a single PSE output by smoothing it with a 2 × 2 box filter (**Fig. S7C, D**).

References

1. Kim, T. H. & Schnitzer, M. J. Fluorescence imaging of large-scale neural ensemble dynamics. *Cell* **185**, (2022).
2. Abdelfattah, A. S. *et al.* Bright and photostable chemigenetic indicators for extended in vivo voltage imaging. *Science (80-.).* (2019). doi:10.1126/science.aav6416
3. Villette, V. *et al.* Ultrafast Two-Photon Imaging of a High-Gain Voltage Indicator in Awake Behaving Mice. *Cell* **179**, (2019).
4. Piatkevich, K. D. *et al.* A robotic multidimensional directed evolution approach applied to fluorescent voltage reporters article. *Nat. Chem. Biol.* (2018). doi:10.1038/s41589-018-0004-9
5. Kannan, M. *et al.* Dual-polarity voltage imaging of the concurrent dynamics of multiple neuron types. *Science (80-.).* **378**, (2022).
6. Fan, L. Z. *et al.* All-optical physiology resolves a synaptic basis for behavioral timescale plasticity. *Cell* **186**, (2023).

7. Zhang, J., Xiong, T., Tran, T., Chin, S. & Etienne-Cummings, R. Compact all-CMOS spatiotemporal compressive sensing video camera with pixel-wise coded exposure. *Opt. Express* **24**, 9013 (2016).
8. Sarhangnejad, N. *et al.* Dual-Tap Pipelined-Code-Memory Coded-Exposure-Pixel CMOS Image Sensor for Multi-Exposure Single-Frame Computational Imaging. *Dig. Tech. Pap. - IEEE Int. Solid-State Circuits Conf.* **2019**-**Febru**, 102–104 (2019).
9. Luo, Y. & Mirabbasi, S. A 30-fps 192 × 192 CMOS Image Sensor With Per-Frame Spatial-Temporal Coded Exposure for Compressive Focal-Stack Depth Sensing. *IEEE J. Solid-State Circuits* **57**, (2022).
10. Tocci, M. D., Kiser, C., Tocci, N. & Sen, P. A versatile HDR video production system. *ACM Trans. Graph.* (2011). doi:10.1145/2010324.1964936
11. Nevian, T., Larkum, M. E., Polsky, A. & Schiller, J. Properties of basal dendrites of layer 5 pyramidal neurons: A direct patch-clamp recording study. *Nat. Neurosci.* **10**, (2007).
12. Nguyen, C. *et al.* Simultaneous voltage and calcium imaging and optogenetic stimulation with high sensitivity and a wide field of view. *Biomed. Opt. Express* **10**, (2019).
13. Ghosh, K. K. *et al.* Miniaturized integration of a fluorescence microscope. *Nat. Methods* **8**, 871–878 (2011).
14. Juneau, J. *et al.* MiniFAST: A sensitive and fast miniaturized microscope for in vivo neural recording. *bioRxiv* (2020).
15. Guo, C. *et al.* Miniscope-LFOV: A large field of view, single cell resolution, miniature microscope for wired and wire-free imaging of neural dynamics in freely behaving animals. *bioRxiv* (2021).
16. Scherrer, J. R., Lynch, G. F., Zhang, J. J. & Fee, M. S. An optical design enabling lightweight and large field-of-view head-mounted microscopes. *Nat. Methods* (2023). doi:10.1038/s41592-023-01806-1
17. Lee, A. K. & Wilson, M. A. Memory of sequential experience in the hippocampus during slow wave sleep. *Neuron* **36**, 1183–1194 (2002).
18. Fossum, E. & Hondongwa, D. B. A Review of the Pinned Photodiode for CCD and CMOS Image Sensors. *IEEE J. Electron Devices Soc.* **2**, 33–43 (2014).
19. Newman, J. P. *et al.* Optogenetic feedback control of neural activity. *Elife* **4**, (2015).

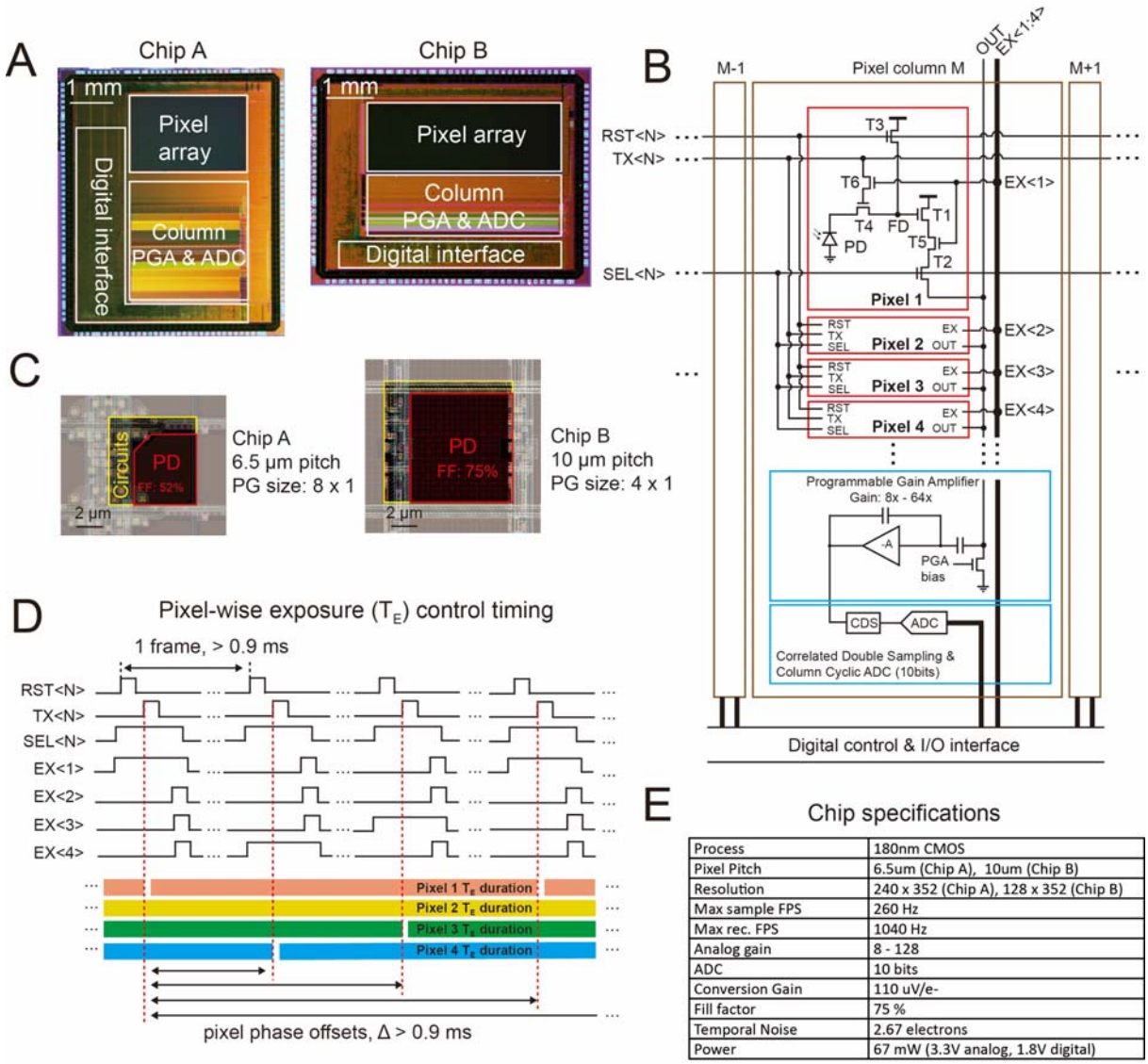
Acknowledgments

We thank S. Chin and G. Brown for discussion and development of additional sparse devolution algorithms. We thank J. Liu, G.F. Lynch, and A. Khalifa for comments on the manuscript. The chip fabrication was funded by support by the Louis B. Thalheimer Fund for Translational Research, and the NIH (R21 EY028381-01). J.Z., Z.C. and M.A.W acknowledges additional funding through the NIH (MH118928).

Author contributions

J.Z., J.P.N, M.A.W conceptualize the ideas. J.Z., J.P.N and C.L. conducted proof of concept testing. J.Z. designed the image sensor with input from E.F. and R.E.C. J.Z. and J.P.N implemented the acquisition hardware and firmware to steam the imaging data. J.Z., J.P.N, Z.W., and Y.Q. conducted the in vitro experiments using cultured neurons. J.Z., W.G. and Z. C. processed the imaging data. All authors contributed to the manuscript.

Supplementary Figures



Process	180nm CMOS
Pixel Pitch	6.5um (Chip A), 10um (Chip B)
Resolution	240 x 352 (Chip A), 128 x 352 (Chip B)
Max sample FPS	260 Hz
Max rec. FPS	1040 Hz
Analog gain	8 - 128
ADC	10 bits
Conversion Gain	110 uV/e-
Fill factor	75 %
Temporal Noise	2.67 electrons
Power	67 mW (3.3V analog, 1.8V digital)

Fig. S1: PE-CMOS design detail.
A, Micrograpy of two PE-CMOS prototypes implemented in a 180 nm CMOS image sensor process. Chip A and B feature 6.5 and 10 µm pixels, respectively.
B, PE-CMOS circuits schematic for Chip B. Pixels are organized in pixel groups (PG) consisting of K row and 1 column (K = 8 in Chip A, 4 in Chip B). Each column shares one amplifier and ADC. The pixel design features 6 transistors (T1 – T6) to provide pixel-wise exposure control while maximizing the pixel fill factor.
C, Pixel layouts (FF: photodiode fill factor).
D, Example timing control for programmable pixel-wise exposure. A pixel only ends the exposure when its corresponding TX, SEL, and EX signal are high. Its exposure continues otherwise. Pixel-wise exposure control can be achieved by programming column-wise signal EX in coordination with row-wise signals (TX, SEL and EX).
E, sensor specifications table.

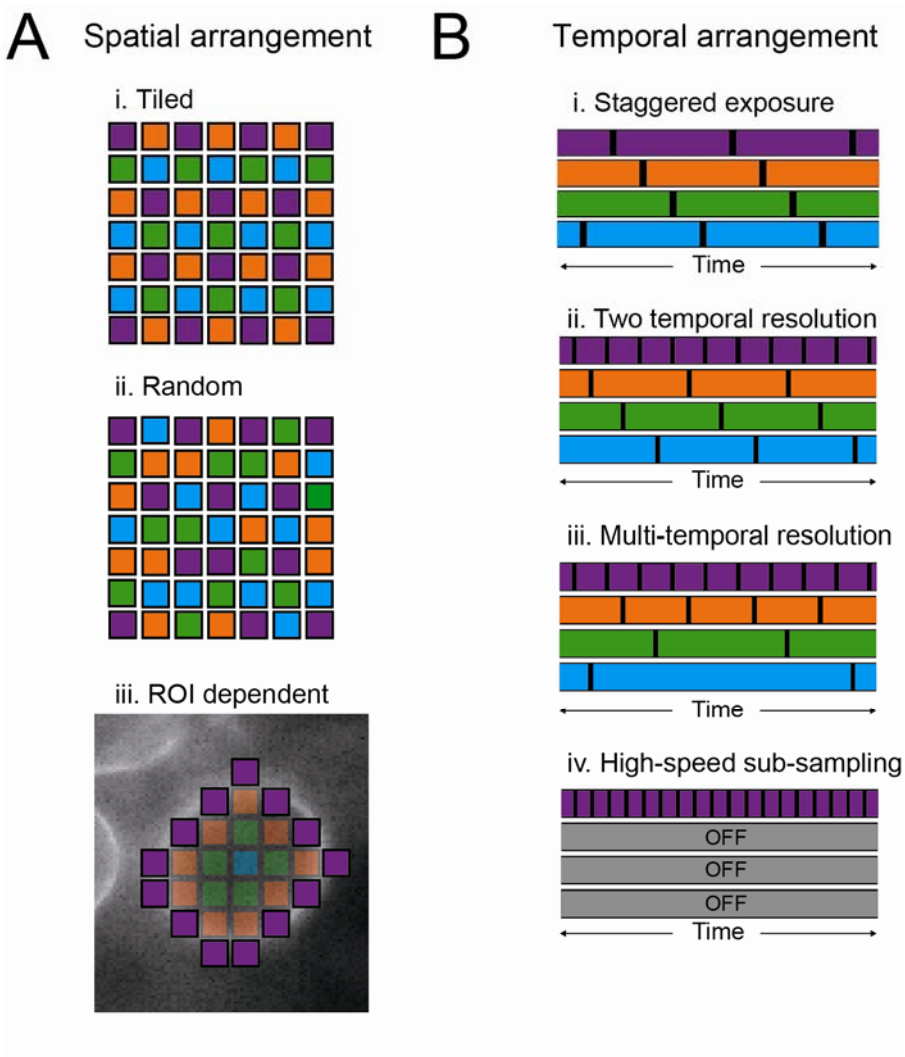


Fig. S2: Examples of pixel sampling patterns enabled by the PE-CMOS

A, spatial arrangement examples. Pixels of the same color have identical exposure duration (T_E) and phase offsets (Δ). In addition to a tiled or random arrangement, pixels can be arranged according to the spatial features of the biological scene. In A.iii, pixels at the bright regions of the cell membrane are configured to have short T_E with a high sampling rate, while pixels in the dim regions use long T_E with a slow sampling rate to enhance the SNR.

B, Temporal arrangement examples. B.i: pixel can have identical T_E with different Δ. B.ii and B.iii: Pixels with different T_E and Δ. B.iv: A subset of pixels is configured for the highest sampling rate while the rest are turned off.

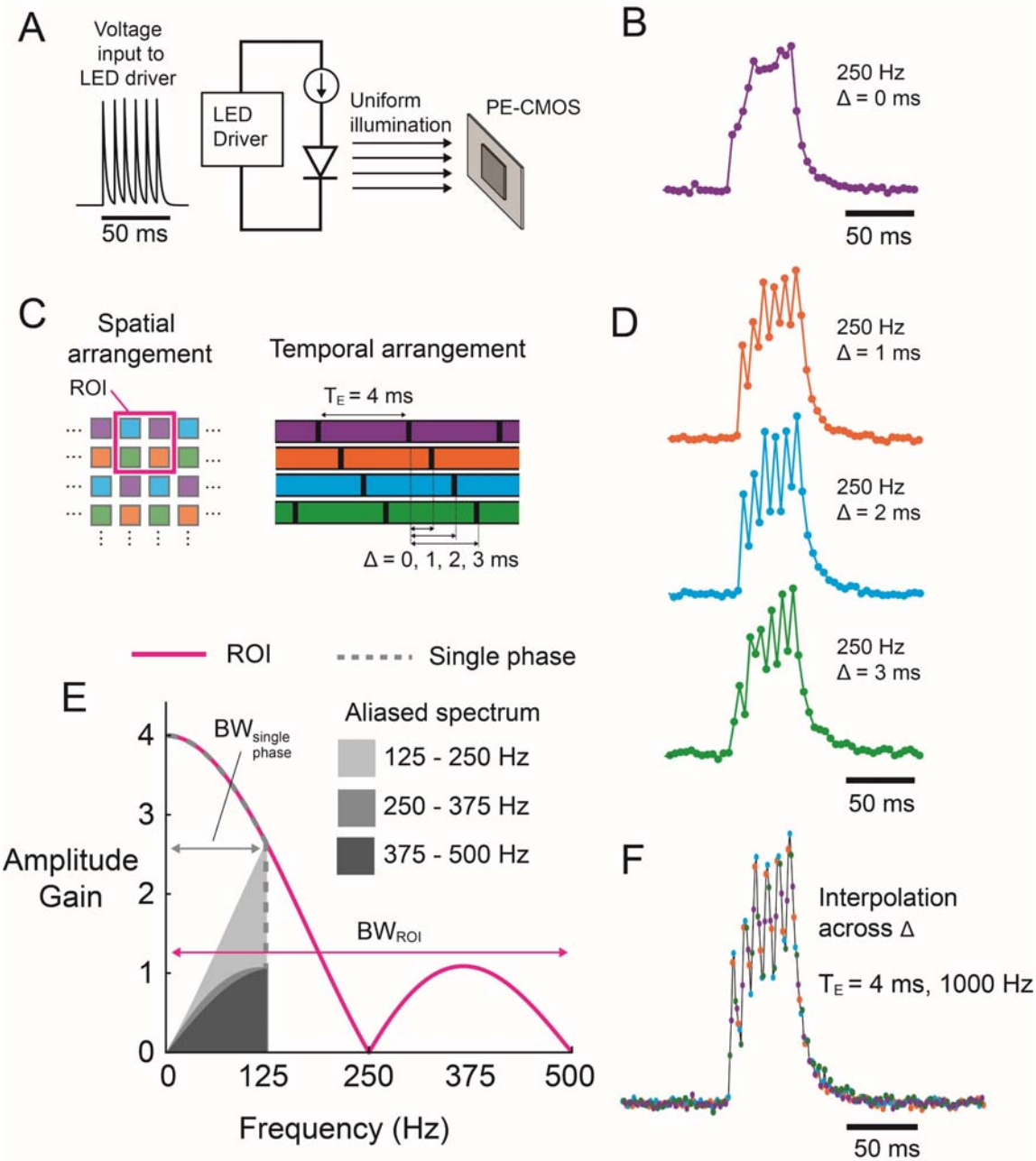


Fig. S3: A benchtop example showing phase offset pixels can improve ROI temporal resolution

A, a test setup using an LED to illuminate the PE-CMOS sensor uniformly. The LED is driven with a current mimicking GEVI spike burst events (6 spikes, 8 ms inter-spike interval).

B, pixels sampling at 250 Hz with exposure (T_E) of 4 ms without phase offset (Δ = 0 ms) are unable to resolve individual spikes due to the slow sampling rate (purple trace).

C, PE-CMOS sampling configuration with various phase offset pixels. The ROI consists of pixels with phase offsets Δ of 0, 1, 2, and 3 ms (color traces).

D, PE-CMOS pixels at different phases together samples the fast-spiking activity that is unsolvable using a single phase.

E, the frequency response of pixels without phase offsets vs. pixels with phase offsets. In this example, the single phase's bandwidth is 125 Hz, with distortion added by frequency aliasing (gray color). With phase offset, PE-CMOS extends the bandwidth of the ROI by 4 times (500 Hz) and avoids aliasing effects (magenta trace).

F, an 1000 Hz equivalent high-speed time series (8 samples between two consecutive spike peaks) from interpolating PE-CMOS pixels at different phases.

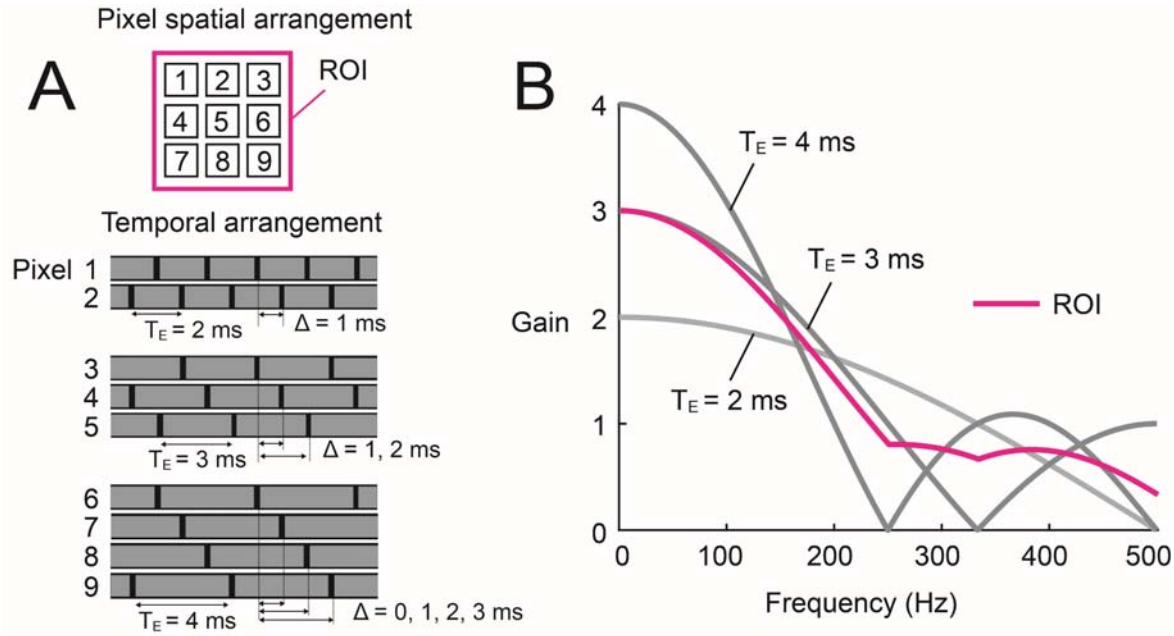


Fig. S4: PE-CMOS frequency response of ROI containing multiple pixel exposure.
A, An ROI with 9 pixel configured with T_E = 2, 3, and 4 ms and phase offsets of Δ = 0, 1, 2, and 3 ms.
B, The combined ROI's frequency response eliminates narrow band attenuations at different T_E 's frequenecy response.

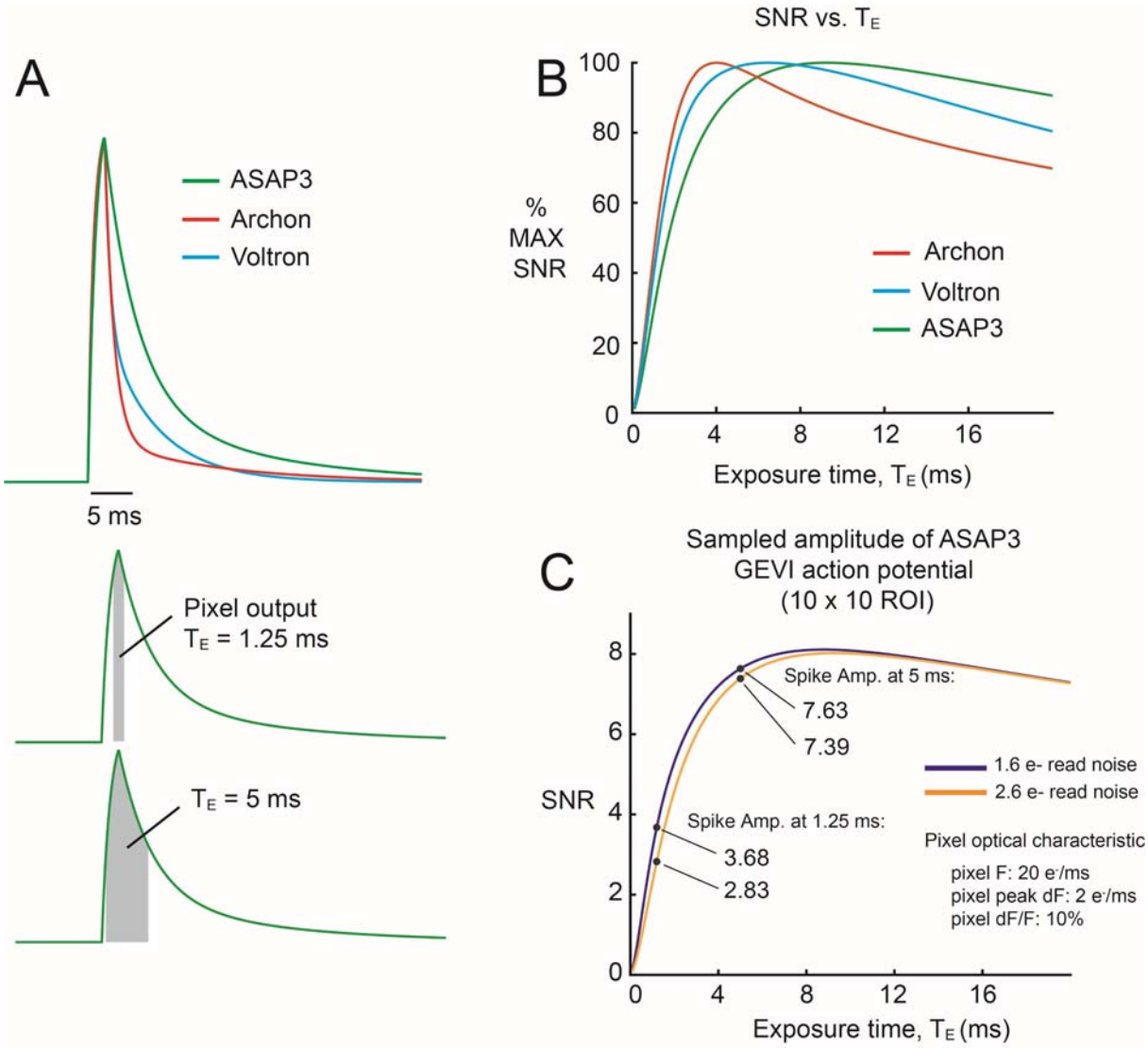


Fig. S5: Modeling the relationship between pixel exposure and SNR of sampling GEVI action potentials

A, Spikes modeled using timing parameters of several GEVI indicators. A pixel integrates fluorescence photons for exposure time, T_E, which equals to the area under the curve.

B, Pixel SNR with respect to exposure time, T_E, for spikes modeled in panel A. Each curve is normalized with the maximum SNR set to 100%. The SNR increases fast with longer T_E but drops gradually as extra integration time adds more shot noise than signal power due to baseline fluorescent.

C, SNR curves at sampling the ASAP3 spikes. Two curves showing the SNR of pixels with different read noise (ROI: 10 x 10 pixels). Despite the PE-CMOS (orange curve) having slightly higher read noise of 2.6 electrons (e^-) rms than the sCMOS (blue curve, 1.6 e^- rms), longer T_E (5 ms) still enables the PE-CMOS to have >2 times better SNR than the sCMOS with shorter T_E (1.25 ms).

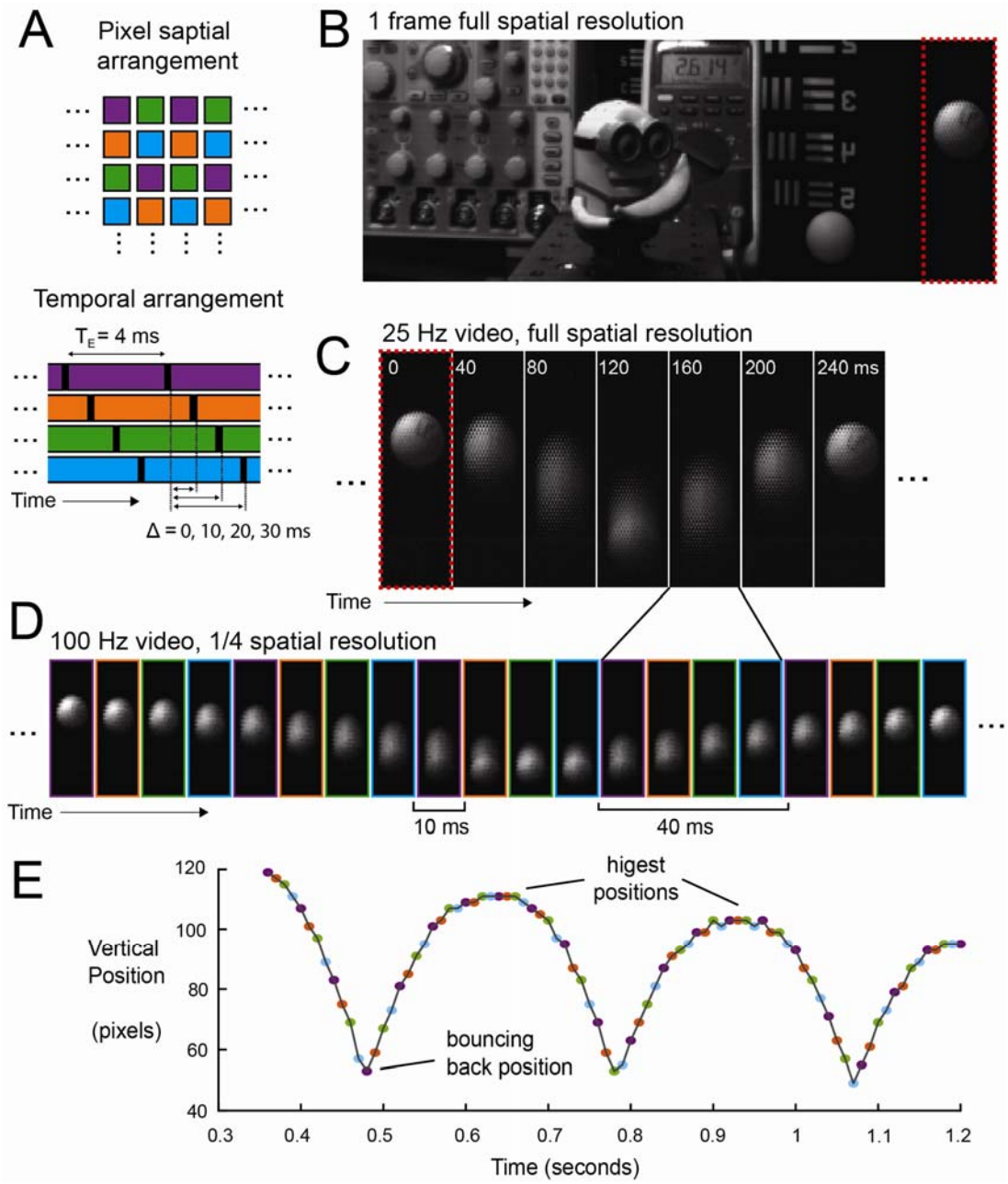


Fig. S6: Applying the PE-CMOS to general imaging application to resolve fast motion with low sampling rate.

A, a staggered pixel configuration with exposure (T_E) of 40 ms and phase offset (Δ) = 0, 10, 20, 30 ms. Pixel of the same color has identical exposure and phase offset.

B, A single frame at full spatial resolution from a video at 25 Hz speed.

C, Seven frames taken from a 25 Hz video of the section highlighted by the dotted red line in panel B. The bouncing ball's motion is embedded within the spatial patterns at full each.

D, splitting each full-resolution frame into four sub-frames, each with 1/4 resolution to resolve a 100 Hz equivalent video.

E, Resolving the position of the bouncing ball at 100 Hz resolution while each pixel is sampled at 25 Hz, results in 4-fold gain in temporal resolution by using phase-shifted pixels.

.

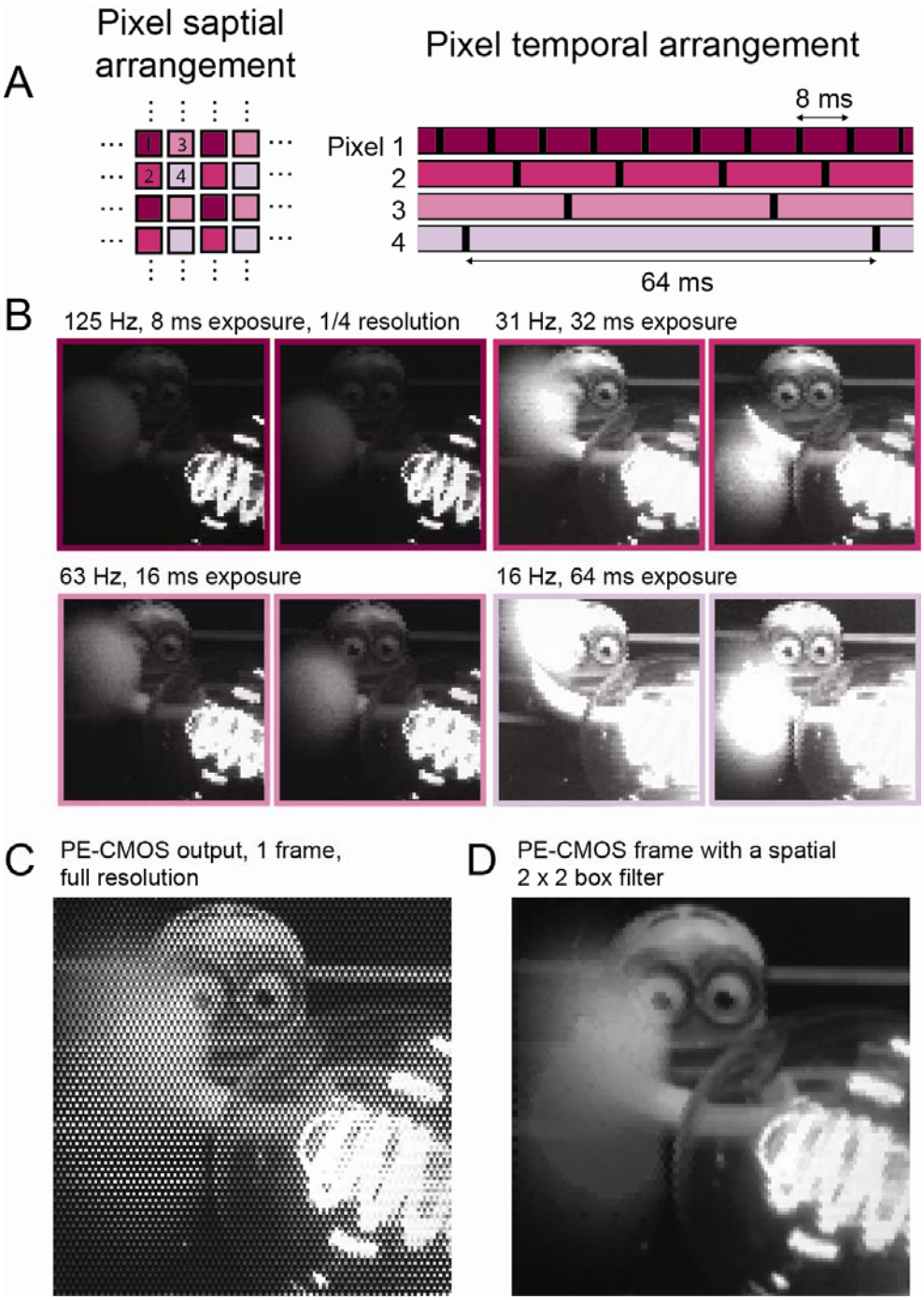


Fig. S7: Multi-spatiotemporal sampling using the PE-CMOS captures high dynamic range videos

A, Spatially varying pixel configuration T_E = 8 – 64 ms.

B, The PE-CMOS simultaneously samples the scene at 64 Hz – 125 Hz at 1/4 spatial resolution, resolving fast motion, avoiding over-exposure and enhancing the SNR of the dark regions.

Data availability

Code and data for the figures can be found in the public repository: https://github.com/jz0229/PE-CMOS