On-Sensor Background Event Suppression with FeFETs

Leo Liu
Stanford University
Stanford, CA
liuleo@stanford.edu

Brandon Cai
Stanford University
Stanford, CA
bjcai@stanford.edu

Kwabena Boahen
Stanford University
Stanford, CA
boahen@stanford.edu

Abstract

Event cameras promise microsecond latency, but background activity from leakage currents and egomotion congests the readout network and inflates latency and jitter. Existing suppression algorithms, which depend on timestamping, precludes pixel-level integration and thus must be relegated to the chip periphery, occurring after event readout. To circumvent this bottleneck, we propose to use dense ferroelectric FETs to detect spatiotemporally-correlated events and reject noise without timestamping. Our simulations show that over 90% of leakage- and egomotion-induced events are suppressed while preserving over 80% of foreground events. That cuts a 140 million events per second stream 14-fold to 10 million events per second, and consequently reduces latency and jitter by a factor 1,000 and 2,500, respectively.

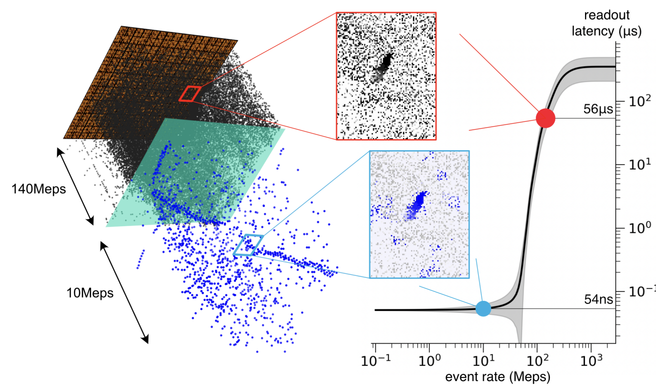

Figure 1. **Suppressing the event rate by a factor of 14 reduces readout latency by a factor of 1000.**
Left – A 1024×1024 pixel array (brown) outputs 140Meps to a 256×256 logic core chip (teal) that reduces event-rate 14-fold to 10Meps by suppressing background events. *Insets* – before (black) and after (blue) suppression.
Right – Reducing the event-rate from 140Meps to 10Meps would cut readout latency 1000-fold, from 56µs (red) to 54ns (teal), and jitter (gray fill) 2,500-fold, from 23µs to 9ns [5].

1. Event Cameras

Events generated from background activity overwhelm an event camera's readout network and balloon readout latency and jitter. When an event camera is stationary or the scene is uncluttered, an on-chip network could read out a moving object's associated events in as little as 50ns [5]. But leakage and egomotion in a cluttered background produces tens of events per second (eps) from every pixel and thus a megapixel array outputs tens of millions of events per second. These events congest the network to increase both latency and jitter 1,000-fold from 50ns to 56µs and 2,500-fold from 9ns to 23µs, respectively (Fig. 1, right).

Though existing algorithms can suppress over 95% of background events while preserving 90% of object events, their reliance on timestamping precludes implementation within the pixel array. Handcrafted spatiotemporal filters classify a pixel's event as object or background by differencing its timestamp against those of recent events at neighboring pixels, either thresholding that difference against a programmable constant or counting how many neighbors fired within a specified time window [3]. But microsecond timestamps are too large to embed into every pixel, and thus these algorithms must execute at the chip's periphery [4] or on an off-chip processor [6]. Consequently, these filters occur after event readout and therefore cannot mitigate latency or jitter due to congestion.

We propose to filter background events prior to readout by densely (3D) integrating the pixel array with a processor chip that interfaces tiles of 4-by-4 pixels with equivalently-sized logic cores leveraging dense and efficient ferroelectric field-effect transistors (FeFETs) to detect spatiotemporally correlated events.[1] Along a row or column of four pixels, a pixel's event pulses a corresponding gate of a four-gate FeFET. This FeFET only conducts current if its gates are pulsed in exact order from source to drain, corresponding to the spatiotemporally ordered events that a moving edge produces (Fig. 2). To track both left- and rightward motion,

[1] Appendix A details FeFET operation.

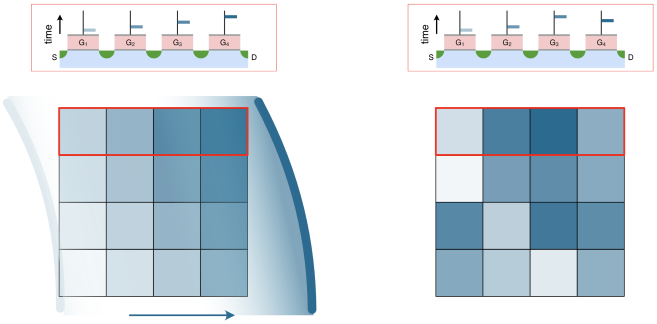

Figure 2. **A four-gate FeFET detects spatiotemporally correlated events and rejects uncorrelated events.**
Left – As an edge sweeps from left to right across a 4-by-4 tile, its pixels fire events in that left-to-right order (shades of blue; darker pixels fired events more recently). In the top row of pixels (outlined red), a 4-gate FeFET receives gate pulses whose spatial order (G1, G2, G3, G4, from source to drain) matches their temporal order. Thus, this FeFET conducts current.
Right – Random orders of events are unlikely to arrive in left-to-right order. And thus this FeFET does not conduct current.

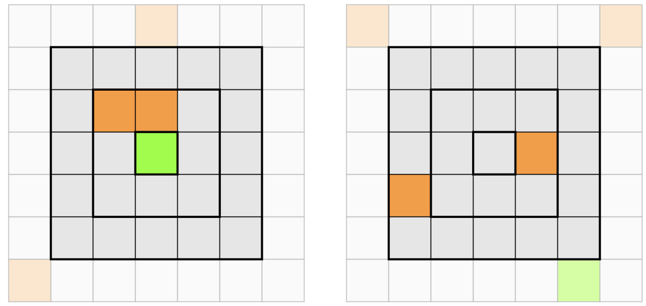

Figure 3. **A core's state depends on the states of neighbors in its inner and outer annulus spanning 3-by-3 and 5-by-5 16-pixel patches, respectively.**
Left – When a 4-gate FeFET detects motion, its core transitions out of IDLE (gray) to CHECK (orange). If two nearest neighbors are in CHECK (inner annulus), then it transitions from CHECK to ACTIVE (green).
Right – When a second-nearest neighbor (outer annulus) transitions to CHECK, this core is reset back to IDLE.

we place two FeFETs oriented in opposition and OR events from the same column. Doing the same for vertical motion yields four FeFETs in total and at 0.25μm^2 per gate totaling a mere 4μm^2 per sixteen-pixel tile.

By computing using these compact FeFETs' outputs as primitives, our algorithm suppresses over 90% of background events (induced by leakage or egomotion) while preserving over 85% of object events *before* event readout. That reduces a 140Meps (5.7Meps object, 69.0Meps leakage, and 65.1Meps egomotion) event stream 14-fold to merely 10Meps, and thus latency and jitter reduces 1,000-fold, from 56μs to 54ns and 2,500-fold, from 23μs to 9ns, respectively (see Fig 1, *left*).

Our contributions are as follows. Section 2 presents how a four-gate FeFET differentiates correlated and uncorrelated events to suppress leakage noise. Section 3 details how multiple tiles coordinate to suppresses egomotion. Finally, Section 4 presents simulated results and compares these to the state-of-the-art.

2. Suppressing Leakage Noise with FeFETs

While a moving object's events pulse a FeFET's four gates consecutively, noisy events pulse its gates randomly. As an object's edge sweeps across four consecutive pixels, a corresponding FeFET's gates are pulsed sequentially in an escalating but non-monotonic order (e.g. gates 1, 2, 3, 4 are pulsed in order **1**-**2**-1-2-**3**-2-3-**4**), as is characteristic of a spatially-diffuse edge. In contrast, random events pulse gates with uniform probability. To differentiate these two cases, a 4-gate FeFET only detects the order of the initial event (bolded) from each pixel. A completed sequence is communicated to digital logic circuitry within the core, which resets this FeFET's state. When an out-of-order initial event occurs, this FeFET self-resets without communicating. With misdetection occurring at a mere 4.2% probability (chance of one in-order sequence out of $4! = 24$ equally-possible permutations), that correctly rejects spatiotemporally-uncorrelated activity with 95.8% accuracy.

While a moving object's events produce a contiguous cluster of active cores, noisy events stimulate spatially incoherent detections dispersed across the array. We differentiate these two scenarios by setting each core to one of three states (IDLE, CHECK, ACTIVE, Fig. 3). Detecting an in-order sequence transitions a core's state from IDLE to a CHECK state, where it checks if at least two neighbors within an eight-tile inner annulus (128 pixels) have detected the same motion and are also in CHECK. Exceeding that threshold transitions this core's state from CHECK to ACTIVE, thereby confirming a moving object spanning multiple tiles. Failing that threshold resets a sixteen-tile outer annulus (256 pixels) to IDLE—as there are no nearby objects—rejecting uncorrelated activity. Under random activity, a tile transitions erroneously from IDLE to CHECK with 4.2% probability and from CHECK to ACTIVE with 4.1% probability. But is reset back to IDLE by any of sixteen cores in its outer ring with probability 49%. At steady state, cores are falsely active with probability 7.7% and 92.3% of events caused by noise are suppressed.

Following an object's motion, noise around the cluster's perimeter deactivates the remnant cluster of active cores (Fig. 4). As noisy events randomly toggle cores near an active cluster from IDLE to CHECK, cores at the edge of

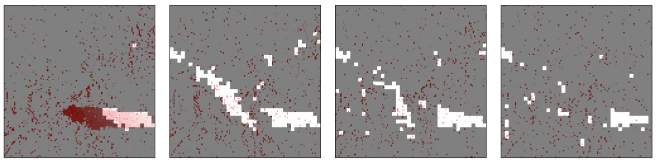

Figure 4. **Following the onset of a moving bee, active tiles are stochastically reset back to inactive.**
A bee moving right-to-left generates events (red) that leave behind a trail of active tiles (clear overlay). Without further stimulus, state transitions of nearby tiles deactivate this trail back to the ambient level within 100ms.

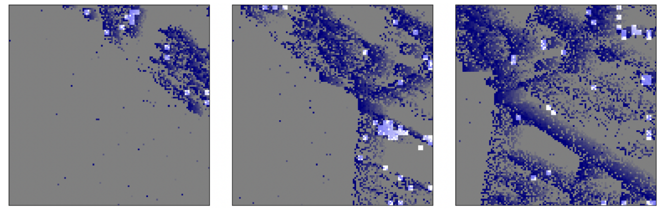

Figure 5. **Egomotion induces noise-like activity in tiles.**
Screenshots are taken at 10, 20, and 30ms after the onset of an egomotion-only event stream [1]. Overlaid are active tiles (clear) and inactive tiles (gray). Despite strongly correlated activity, only 3%, 8%, and 6% of tiles are active at these respective moments in time.

that cluster are stochastically reset at a rate proportional to both the length of the cluster's perimeter and the ambient event rate (λ Hz/pixel), which toggles cores from IDLE to CHECK at a rate $0.042 \times 4 \times \lambda = \lambda/6$ tiles/s. Hence an object that activates an N^2-tile cluster with a perimeter approximately equal to N tiles is deactivated $6N/\lambda$ seconds/tile. Integrating across the entire cluster thus yields $6N^2/\lambda$ seconds to reset this cluster. This inverse dependence on λ accelerates the clearance of stale clusters under heavy noise, thus conferring robustness across a wide range of object and background event rates.

3. Suppressing Egomotion Events

Egomotion triggers Poissonian events along every visible edge (Fig. 5). When every edge shifts by one pixel, each pixel generates an event after an exponentially-distributed delay, owing to pixel mismatch. Hence activity across an entire edge appears like Poisson noise. In the moments that follow, pixels continue to generate Poisson events either from further contrast gradations or from noise.

Because edges appear like noise, long edges characteristic in egomotion are suppressed while short edges characteristic in object motion are let through. This filter condition is dictated by the radius of the inner and outer annulus of each core's neighborhood: edges that stretch across the outer annulus are suppressed while edges that fit in the inner annulus are passed through. Given the exponential timing of events, the false-positive probability that a long edge triggers two-of-eight detections in the inner annulus before it triggers any-of-sixteen in the outer annulus is a mere 7.4%, thus predicting a 92.6% suppression rate.

4. Results

Our preliminary simulations predict that this algorithm suppresses 92% of noisy events and classifies object and noise events with 92% accuracy. We merged a recorded stream of 57K object events containing little noise, 690K synthetic Poisson leakage events, and 651K egomotion events [1] into a single 140Meps event stream (see Fig 1, *inset*). We emulated our proposed architecture in Python and read out events that occur only within active tiles. We then counted the number of object and background events in this output. Out of a total of 58K object events, 50K were successfully read out (85%) while a total of 1.3M events of both leakage and egomotion events are suppressed (97%). Altogether, that reduces the total number of events 14-fold.

5. Conclusion

We have presented a lightweight yet effective spatiotemporal filter that uses dense multi-gate FeFETs to detect correlated events and reject uncorrelated events. Leveraging these devices, we detailed a tiled architecture whose tiles each independently receive events from a cluster of 4-by-4 pixels, connect their outputs to FeFETs, and coordinate with each other to suppress background events. Our simulations reveal that the event rate could be reduced up to 14-fold. And with this reduced rate latency and jitter are reduced 1,000-fold and 2,500-fold, respectively.

A. FeFET Operation

A multi-gate FeFET's selective response to a sequence of voltage pulses applied to its gates arises from two requirements for flipping polarization of charge dipoles in a ferroelectric layer: (i) An electric field that exceeds the ferroelectric layer's coercive field. (ii) A supply of electrons to a p-doped channel operating in inversion mode (i.e., with negative charge carriers). When a pulse of voltage applies an electric field across the gate stack that exceeds the coercive field, charge dipoles flip and the channel becomes conductive (Fig. 6, *left*). But for the voltage to produce an electric field across the gate-stack—rather than across the channel—electrons must be supplied from the source. That occurs only if the first pulse is applied to the first gate, whose source is grounded.

Flipping the first gate's dipoles establishes a conductive

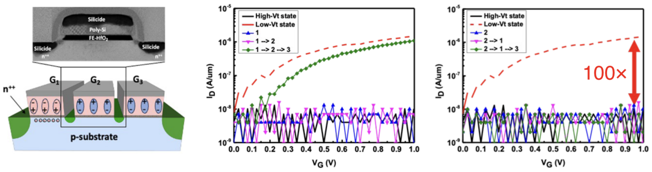

Figure 6. **Experimental measurements of a 3-gate FeFET fabricated by GlobalFoundries [2].**
Left – A three-gate FeFET after a voltage pulse strikes the first transistor's gate (G1), flips charge dipoles in its ferroelectric layer, and attracts electrons from its n-doped source into its p-doped channel. To manufacture this string in a 28-nm gate-first CMOS process, GlobalFoundries deposited ferro-electric-doped hafnium-dioxide (FE-HfO2), titanium nitride (TiN, not shown), polysilicon (Poly-Si), and silicide layers on a p-doped 300-mm silicon wafer. *Center* – When voltage pulses strike the manufactured FeFET string's gates consecutively (G1 $\rightarrow$ G2 $\rightarrow$ G3), its threshold-voltage drops from 1.2V (black curve) to 0.1V (green curve). *Right* – When the first pulse strikes the second gate and the second pulse strikes the first gate (G2 $\rightarrow$ G1 $\rightarrow$ G3), the string's threshold remains above 1.0V (black and green curves) and it conducts 100 times less current (red double-arrow at $V_G = 1$V).

channel to supply electrons to the second gate's (virtual) source. That allows this gate's dipoles to be flipped when it is pulsed. That extends the conductive channel to third gate's (virtual) source, and so on. Thus, if all the gates are pulsed consecutively, then all the dipoles are flipped and the string's current increases a hundred-fold (Fig. 6, *center*). But if the first and second pulses are swapped, for example, then the second gate's dipoles are not flipped and the string's current does not increase (Fig. 6, *right*).

References

[1] Levi Burner, Anton Mitrokhin, Cornelia Fermüller, and Yiannis Aloimonos. Evimo2: An event camera dataset for motion segmentation, optical flow, structure from motion, and visual inertial odometry in indoor scenes with monocular or stereo algorithms, 2022. 3

[2] Hugo J.-Y. Chen, Matthew Beauchamp, Kasidit Toprasertpong, Fei Huang, Louis Le Coeur, Thorgund Nemec, H.-S. Philip Wong, and Kwabena Boahen. Multi-gate fefet discriminates spatiotemporal pulse sequences for dendrocentric learning. In *2023 International Electron Devices Meeting (IEDM)*, pages 1–4, 2023. 4

[3] Shasha Guo and Tobi Delbruck. Low cost and latency event camera background activity denoising. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 45(1):785–795, 2023. 1

[4] Alireza Khodamoradi and Ryan Kastner. $o(n)$-space spatiotemporal filter for reducing noise in neuromorphic vision sensors. *IEEE Transactions on Emerging Topics in Computing*, 9(1):15–23, 2021. 1

[5] Leo Liu and Kwabena Boahen. Hierarchical event readout with asynchronous pipelined opportunistic merges. In *2025 29th IEEE International Symposium on Asynchronous Circuits and Systems (ASYNC)*, pages 108–117, 2025. 1

[6] A. Rios-Navarro, S. Guo, G Abarajithan, K. Vijayakumar, A. Linares-Barranco, T. Aarrestad, R. Kastner, and T. Delbruck. Within-camera multilayer perceptron dvs denoising. In *2023 IEEE/CVF Conference on Computer Vision and Pattern Recognition Workshops (CVPRW)*, pages 3933–3942, 2023. 1