Event Structural Valley: A Unified Theoretical and Practical Framework for Event Camera Autofocus

Supplementary Material

A. Theoretical Derivation of Event-Rate Behavior

This appendix provides complete derivations supporting the dual-peak–valley behavior of the event-rate curve. We first derive the blur-dependent activation region from the temporal contrast threshold, then prove the rise–peak–fall behavior on canonical primitives (single point and thin line), and finally discuss extensions to multi-feature and textured scenes.

A.1. Setup and Assumptions

We use log-intensity $L = \log I$ and denote by L_σ the image blurred by a Gaussian kernel of standard deviation $\sigma \geq 0$. We assume:

(A1) Short-window sweep. Events are counted within a short time window Δt in which the blur scale changes by $\Delta\sigma$.

(A2) Sweep-dominant variation. Within Δt, the dominant contribution to ΔL is induced by the blur-scale change (lens motion). Other sources such as illumination flicker or strong object motion are treated as perturbations and are mitigated by the ESVA regularization in the main paper.

(A3) Locally Gaussian defocus. Defocus blur is approximated by Gaussian convolution, which is standard and sufficient for the qualitative rise–peak–fall shape.

A.2. From Temporal Threshold to an Activation Region

Starting from the event condition $|\Delta L(\mathbf{x}, t)| \geq C$, we apply the short-window approximation:

$$\Delta L(\mathbf{x}; \sigma) \approx \frac{\partial L_\sigma(\mathbf{x})}{\partial \sigma}\, \Delta\sigma. \tag{1}$$

Thus, event triggering implies

$$\left|\frac{\partial L_\sigma(\mathbf{x})}{\partial \sigma}\right| \geq \theta(C), \qquad \theta(C) \coloneqq \frac{C}{|\Delta\sigma|}. \tag{2}$$

Define the activation region

$$\Omega(\sigma) \coloneqq \left\{\mathbf{x} : \left|\frac{\partial L_\sigma(\mathbf{x})}{\partial \sigma}\right| \geq \theta(C)\right\}. \tag{3}$$

In a discrete sensor, the event count in the window is proportional to the number of activated pixels; in continuous notation:

$$R(\sigma) \propto \mathrm{meas}\big(\Omega(\sigma)\big). \tag{4}$$

A.3. Single-Point Scene: Rise–Peak–Fall and a Closed-Form Maximizer

We begin with a single bright impulse because it permits an explicit characterization of $\Omega(\sigma)$ and reveals the intrinsic non-monotonic dependence on σ.

Scene model. Consider a 1D impulse scene $I(x) = A\,\delta(x)$ with amplitude $A > 0$. Under Gaussian blur with standard deviation $\sigma > 0$,

$$I_\sigma(x) = (I * h_\sigma)(x) = A\, h_\sigma(x), \tag{5}$$

where $h_\sigma(x) = \frac{1}{\sqrt{2\pi}\,\sigma} \exp\!\left(-\frac{x^2}{2\sigma^2}\right)$. Using $L = \log I$, for this primitive it is standard to work directly with intensity I_σ since the qualitative activation geometry is identical; the threshold $\theta(C)$ can be interpreted as an effective contrast bound in the same manner.

Derivative with respect to blur scale. Differentiating I_σ with respect to σ gives

$$\frac{\partial I_\sigma(x)}{\partial \sigma} = \frac{A}{\sqrt{2\pi}\,\sigma^2}\left(\frac{x^2}{\sigma^2} - 1\right)\exp\!\left(-\frac{x^2}{2\sigma^2}\right). \tag{6}$$

Activation set and bounding radii. Let $u = x/\sigma$. Then Eq. (6) becomes

$$\left|\frac{\partial I_\sigma(x)}{\partial \sigma}\right| = \frac{A}{\sqrt{2\pi}\,\sigma^2}\left|u^2 - 1\right| e^{-u^2/2}. \tag{7}$$

The activation condition $|\partial I_\sigma / \partial\sigma| \geq \theta(C)$ is equivalent to

$$\left|u^2 - 1\right| e^{-u^2/2} \geq \frac{\theta(C)\sqrt{2\pi}\,\sigma^2}{A}. \tag{8}$$

A direct closed-form boundary for $\Omega(\sigma)$ is algebraically involved because of the factor $|u^2 - 1|$. However, the rise–peak–fall behavior follows from tight bounds on $|u^2 - 1|$ over a dominant region:

- For $|u| \leq 1$, we have $0 \leq |u^2 - 1| \leq 1$.
- For $|u| \leq 1/\sqrt{2}$, we have $|u^2 - 1| \geq 1/2$.

Using these, define an *outer* (superset) activation region $\Omega_+(\sigma)$ by replacing $|u^2-1|$ with its upper bound 1 on $|u| \le 1$:

$$\Omega(\sigma) \subseteq \Omega_+(\sigma) := \left\{ x : \exp\left(-\frac{x^2}{2\sigma^2}\right) \ge \frac{\theta(C)\sqrt{2\pi}\,\sigma^2}{A} \right\}. \tag{9}$$

Similarly, define an *inner* (subset) activation region $\Omega_-(\sigma)$ by using the lower bound $|u^2-1| \ge 1/2$ for $|u| \le 1/\sqrt{2}$:

$$\Omega_-(\sigma) := \left\{ x : \exp\left(-\frac{x^2}{2\sigma^2}\right) \ge \frac{2\,\theta(C)\sqrt{2\pi}\,\sigma^2}{A} \right\} \subseteq \Omega(\sigma). \tag{10}$$

Both $\Omega_+(\sigma)$ and $\Omega_-(\sigma)$ are symmetric intervals $|x| \le r_\pm(\sigma)$ with radii

$$r_\pm(\sigma) = \sqrt{-2\sigma^2 \ln\left(\frac{c_\pm\,\theta(C)\sqrt{2\pi}\,\sigma^2}{A}\right)}, \tag{11}$$

where $c_+ = 1,\ \ c_- = 2$. Therefore, the activation measure satisfies

$$2r_-(\sigma) \;\le\; \mathrm{meas}\big(\Omega(\sigma)\big) \;\le\; 2r_+(\sigma). \tag{12}$$

Rise–peak–fall and existence of a maximizer at $\sigma > 0$. Define

$$\psi_c(\sigma) := -2\sigma^2 \ln\left(\frac{c\,\theta(C)\sqrt{2\pi}\,\sigma^2}{A}\right), \qquad c > 0, \tag{13}$$

so that $r_\pm(\sigma) = \sqrt{\psi_{c_\pm}(\sigma)}$ when the log argument is in $(0,1)$. Differentiating ψ_c yields

$$\psi_c'(\sigma) = -4\sigma\left[\ln\left(\frac{c\,\theta(C)\sqrt{2\pi}\,\sigma^2}{A}\right) + 1\right]. \tag{14}$$

Thus ψ_c has a unique stationary point at

$$\sigma_c^\star = \sqrt{\frac{A}{c\,\theta(C)\sqrt{2\pi}\,e}}, \tag{15}$$

and $\psi_c(\sigma)$ increases for $\sigma < \sigma_c^\star$ and decreases for $\sigma > \sigma_c^\star$. Consequently, both bounds $r_-(\sigma)$ and $r_+(\sigma)$ are single-peaked with maxima at positive blur scales. By the sandwich relation in Eq. (12), $\mathrm{meas}(\Omega(\sigma))$ must also be non-monotonic and achieves its maximum at some $\sigma^\star > 0$. Moreover, since $\mathrm{meas}(\Omega(\sigma)) \to 0$ as $\sigma \to 0^+$ and also decays for large σ (the right-hand side in Eq. (8) grows with σ^2), the rise–peak–fall behavior holds.

Closed-form maximizer (tight approximation). If we adopt the common tight approximation that the dominant activation boundary is governed by the exponential term in Eq. (8) (equivalently, taking $c = 1$ as a representative constant in Eq. (15)), we obtain the closed-form characteristic scale

$$\sigma^* \approx \sqrt{\frac{A}{\theta(C)\sqrt{2\pi e}}}. \tag{16}$$

This expression matches the scaling observed in practice and is sufficient for explaining why the event rate is maximized at a nonzero defocus level.

Thin-line scene (2D). For a vertical thin bright line $I(x,y) = A\,\delta(x)$ of length L_y, blurring acts only along x and the above 1D analysis replicates along y:

$$\mathrm{meas}\big(\Omega(\sigma)\big) \propto L_y \cdot 2r(\sigma), \tag{17}$$

so the rise–peak–fall behavior and the maximizer location remain unchanged up to a constant factor.

A.4. Extension to Multi-Feature and Natural Scenes

Sparse multi-point scene. Consider K isolated impulses

$$I(x,y) = \sum_{i=1}^{K} A_i\,\delta(x - x_i, y - y_i).$$

After Gaussian blur, I_σ is the sum of shifted Gaussians. Under the sparse regime where activation regions weakly overlap, the total activation measure approximately adds:

$$\mathrm{meas}\big(\Omega(\sigma)\big) \;\approx\; \sum_{i=1}^{K} \mathrm{meas}\big(\Omega_i(\sigma)\big), \tag{18}$$

where $\Omega_i(\sigma)$ denotes the activation region induced by the ith primitive. Since each $\mathrm{meas}(\Omega_i(\sigma))$ is rise–peak–fall with a maximum at $\sigma > 0$ (Sec. A.3), the aggregate remains non-monotonic and exhibits the same qualitative profile. The exact peak location depends on the distribution of amplitudes $\{A_i\}$ and overlap effects.

Natural textured scenes. A textured scene can be locally decomposed into a mixture of edge-like and impulse-like primitives. Under mild conditions (bounded variation and finite-energy local structures), the activation measure integrates local contributions that are each non-monotonic in σ. Therefore, the global event-rate curve $R(\sigma) \propto \mathrm{meas}(\Omega(\sigma))$ preserves a rise–peak–fall trend, although the precise maximizer may shift with texture statistics, local contrast, and overlap.

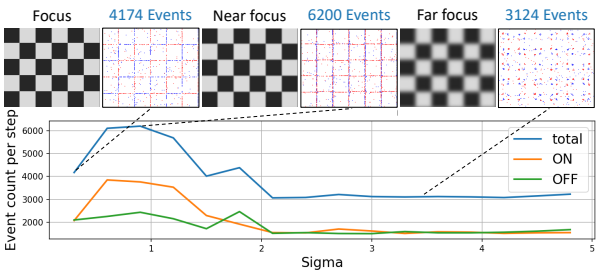

Figure 1. Controlled simulation with a one-sided defocus trajectory. Top: images and simulated events at focus, near-focus, and far-focus. Bottom: event count per step versus blur scale σ (total, ON, OFF). The event rate increases and then decreases as σ grows from 0, corresponding to one half of the M-shaped response in a full focus sweep.

A.5. From Rise–Peak–Fall to Dual-Peak–Valley in a One-Way Sweep

Let $\sigma(t)$ denote the defocus scale during a one-way sweep that crosses the best focus plane. Since σ is a nonnegative measure of defocus, it satisfies the trajectory

$$\sigma(t) \downarrow 0 \uparrow \sigma(t), \tag{19}$$

with $\sigma = 0$ at best focus. Because $R(\sigma)$ has a local minimum at $\sigma = 0$ and achieves maxima at $\sigma^\star > 0$ on both sides (Sec. A.3), the observed curve $R(f)$ exhibits two dominant peaks separated by a valley. This establishes the dual-peak–valley geometry used by ESVA in the main paper.

Summary. Starting from the temporal threshold C, we derived an equivalent activation criterion in terms of $\partial L_\sigma / \partial \sigma$ and an effective threshold $\theta(C) = C/|\Delta\sigma|$. For canonical primitives, the activation measure is provably non-monotonic in σ, with maxima at nonzero defocus and a local minimum at $\sigma = 0$. Combined with the defocus trajectory of a one-way sweep, this yields the dual-peak–valley event-rate structure that underpins valley-seeking autofocus.

B. Additional Experiments and Analyses

This supplementary section complements the main paper with additional analyses to make the methodology and evaluation more complete. We provide a controlled simulation to isolate the causal mechanism behind the dual-peak-valley behavior, extended reliability and tail statistics with significance tests, validation of the confidence score S, and a parameter sensitivity study. These results help characterize ESVA's robustness and applicability beyond the main-text metrics.

B.1. Controlled Simulation Under Explicitly Defined Conditions

We conduct a controlled simulation to validate the predicted rise–peak–fall behavior of event rate under defocus, while eliminating confounding factors such as scene motion and illumination drift. We use a static checkerboard pattern and synthesize a one-sided defocus trajectory by applying a Gaussian PSF with an explicitly defined blur scale σ, which increases from the in-focus state ($\sigma = 0$) to progressively defocused states. Events are generated by a standard contrast-threshold model on log-intensity: $V = \log(I + \epsilon)$ and $\Delta V = V_{\sigma+\Delta\sigma} - V_\sigma$; ON events fire when $\Delta V \geq C_{\text{on}}$ and OFF events fire when $\Delta V \leq -C_{\text{off}}$. We additionally model a refractory period and sparse isolated noise to better match real sensors.

As shown in Fig. 1, along this one-sided defocus path the event rate first increases and then decreases as σ grows from 0. Intuitively, mild defocus expands effective edge support and causes more pixels to cross the contrast threshold, while heavy defocus attenuates local contrast changes and suppresses threshold crossings. When a full sweep crosses the best focus and then moves away on the opposite side, the symmetric counterpart forms the second peak, yielding the dual-peak–valley (M-shaped) response.

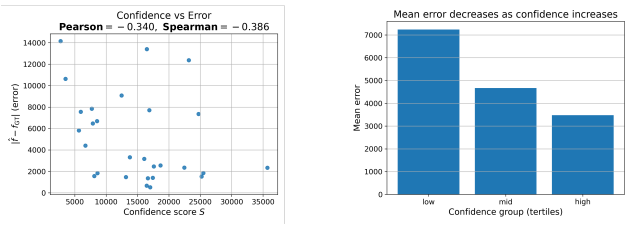

Figure 2. (a) Confidence score S versus absolute focus error on EVK4. Pearson $= -0.340$, Spearman $= -0.386$. (b) Mean autofocus error for samples grouped by confidence tertiles.

B.2. Reliability, Tail Robustness, and Statistical Significance

Beyond mean accuracy, we report dispersion, tail robustness, and paired statistical significance. Table 1 summarizes central tendency (mean/median), dispersion (std), tail behavior (90th percentile and top-10% failures), and paired Wilcoxon signed-rank tests with Holm correction on EVK4. ESVA improves both average and tail performance over representative baselines and yields statistically significant gains over ER+EGS, OLE'23, and PBF.

Metric	Ours	ER+EGS	OLE'23	PBF	ELP
Mean	**4.24**	17.47	10.17	9.71	4.48
Std	**3.83**	15.22	9.21	10.26	4.21
Median	**3.74**	13.32	6.53	4.66	4.58
90th perc.	**10.18**	24.76	26.49	27.47	8.85
Top-10% fail.	**0/28**	5/28	5/28	4/28	0/28
Wilcoxon p (raw)	–	1.51e-05	1.22e-03	4.56e-03	0.99
Wilcoxon p (Holm)	–	**6.06e-05**	**3.66e-03**	**9.13e-03**	0.99

Table 1. EVK4 reliability and significance analysis (unit: ms).

B.3. Confidence Score S as a Reliability Indicator

We analyze whether the confidence score S reflects focusing reliability. Fig. 2(a) shows the relationship between S and absolute focus error on EVK4, with Pearson correlation -0.340 and Spearman correlation -0.386, indicating a moderate negative association. Figure 2(b) and Table 2 further groups sequences by S (low/mid/high) and reports error statistics: low-S cases exhibit substantially larger mean error and worse tail behavior, supporting S as a practical warning indicator.

Group S	**n**	**S_mean**	**err_mean**	**err_90p**
Low-S	9	6304	7238	11337
Mid-S	9	14499	4669	9944
High-S	10	22738	3476	7859

Table 2. Error statistics conditioned on confidence score S (EVK4).

B.4. Failure Boundaries by Illumination and Motion

To characterize failure boundaries, Table 3 reports large-error cases (top 10% errors) grouped by illumination and motion on EVK4. Large errors occur predominantly under low-light conditions, suggesting that weak effective edges and reduced event activity are primary failure factors for this structural prior.

Scene (n)	**Large_err**	**S_mean**	**err_mean**	**err_90p**
dark static (7)	28.6%	18648.1	6910.0	12779.4
dark motion (7)	14.3%	15639.9	5736.9	12040.8
bright motion (7)	0.0%	12947.6	4265.4	7148.2
bright static (7)	0.0%	11994.5	3362.6	6911.2

Table 3. Tail large-error cases (top 10% errors) by lighting and motion on EVK4.

B.5. Parameter Sensitivity

We evaluate sensitivity by independently varying σ_s, τ_c, and η around the default settings. Table 4 summarizes the min/max relative change of mean error when varying one parameter at a time. Across DAVIS, EVK4, and SYN, ESVA remains stable over wide parameter ranges, indicating low sensitivity to precise parameter tuning.

Dataset	$\boldsymbol{\sigma_s}$	$\boldsymbol{\tau_c}$	$\boldsymbol{\eta}$
DAVIS	[-25.8%, +0.0%]	[-27.5%, +13.1%]	[-45.6%, +63.5%]
EVK4	[-65.4%, +0.0%]	[-32.9%, +33.0%]	[+0.0%, +37.6%]
SYN	[-29.5%, +34.0%]	[-55.2%, +63.7%]	[-17.5%, +73.4%]

Table 4. Parameter sensitivity: min/max relative change of mean error when varying one parameter at a time around the default setting.

C. Related Event-Camera Autofocus Methods

C.1. ER+EGS [4]: Event-Rate Maximization with Golden-Section Search

This approach uses the event rate as a sharpness metric and searches on the time/focus axis via an event-domain golden-section strategy. The best focus corresponds to the maximum of the global score:

$$t^\star = \arg\max_t \; F_{\mathrm{ev}}(t), \qquad F_{\mathrm{ev}}(t) = \sum_{x\in\Omega} R_e^2(x,t,\Delta t). \tag{20}$$

Here $R_e(x,t,\Delta t)$ is the average number of events within the event accumulation interval with the length Δt $[t - \Delta t/2,\, t + \Delta t/2]$, and Ω is the camera domain.

Both theoretical analyses and empirical observations in existing work [1–3] further indicate that the focus selected by ER+EGS is only an approximation, as the true in-focus position does not necessarily coincide with the maximum of the event-rate function.

C.2. OLE'23 [3]: Sinc-Shaped Axial Response and Valley-Based Focusing in Microscopy

In the microscopy setting of OLE'23 [3], the axial wave field is modeled to vary with depth as a sinc function. During axial scanning, many events are triggered on both sides of the focal plane but almost none exactly at focus. As a result, the focus is determined by the *minimum* of a short-window event-count metric:

$$M(z) = \sum_{x_s,y_s} \sum_{t_s=t}^{t+\Delta t} [e_s], \qquad \hat{z} = \arg\min_z M(z), \tag{21}$$

where $[e_s]$ counts events at pixel (x_s, y_s) within $[t,\, t+\Delta t]$, and the known scan speed maps depth z to time t. The generation mechanism further implies that the magnitude of the depth derivative of log-brightness is maximized near both sides of focus, which explains the valley criterion:

$$\begin{aligned} z^\star &= \arg\max_z \left\| \frac{\partial L(x,y;z)}{\partial z} \right\| \\ &\approx \arg\max_z \left\| \frac{\partial \log |U(0,0;z)|^2}{\partial z} \right\|, \end{aligned} \tag{22}$$

with $U(0,0;z)$ the on-axis complex amplitude (sinc-shaped).

This valley-seeking criterion is well-justified under the microscope's axial-scanning setup with a static scene. However, in real-world environments with complex structures or motion, these assumptions may become unreliable when the microscope-specific axial assumptions no longer hold.

C.3. PBF [1]: Polarity-Based Fast Focusing

Before and after focus, the same structure produces opposite event polarities, so the positive event-rate sequence

Table 5. Comparison of event-camera autofocus methods. Our ESVA method uniquely leverages a physically grounded dual-peak–valley structure.

Method	Input	Focus Criterion	Physical Basis	Scene Assumptions	Limitations	Speed
ER+EGS [4]	Events only	Max event-rate peak	Weak (empirical extremum)	General; assumes single peak	True focus ≠ global max; noise-sensitive	Slow
OLE'23 [3]	Events (microscope)	Min short-window event count	Microscope-specific (sinc-like axial response)	Static scene and axial scanning configuration	May become unreliable outside static axial-scanning settings	Very Fast
PBF [1]	Events only	Opposite-polarity symmetry	Moderate (polarity inversion near focus)	Requires strong and balanced polarity responses	Sensitive to polarity imbalance; accuracy may degrade	Medium
ELP [2]	Events + Image	Event–Laplacian sign flip	Strong (image Laplacian combined with events)	Needs reliable APS/reconstruction images	Not event-only; depends on image quality	Very Slow
Ours (ESVA)	**Events only**	**Dual-peak–valley geometry**	**Strong; analytically derived from event-generation physics**	**Dynamic scenes; diverse motions and textures**	—	**Very Fast**

$P(v)$ and the negative sequence $N(v)$ are approximately symmetric about the true focus. After denoising and normalizing, the focus is obtained by minimizing a reverse-registration mean squared error:

$$\mathrm{MSE}(a) = \frac{1}{|I(a)|}\int_{I(a)} \big(N'(v) - P'(a-v)\big)^2\, dv,$$
$$v^\star = \tfrac{1}{2}\arg\min_a \mathrm{MSE}(a) + v_1, \tag{23}$$

where a is twice the symmetry center, $I(a)$ is the overlap interval under shift a, v_1 is the start of the original interval, and P', N' are the shifted sequences.

C.4. ELP [2]: One-Step Event-Driven Autofocus

This method links an event-based time-domain first-order change to an image-based spatial second-order Laplacian and defines the Event–Laplacian Product (ELP). Near focus, ELP flips sign from positive to negative, enabling one-step focus detection:

$$\mathrm{ELP}(t) = -\sum \big(\nabla^2 I(t) \cdot E(t)\big), \tag{24}$$

where $I(t)$ is the current grayscale image, ∇^2 is the Laplacian, $E(t)$ is the event frame accumulated over $[t-\Delta t,\, t]$, and p_i is event polarity. The focus is detected at the sign-change point:

$$t:\ \mathrm{ELP}(t-) > 0,\ \mathrm{ELP}(t+) < 0, \tag{25}$$

and the corresponding lens position is taken as the best focus.

By leveraging image-domain information (either from the measured intensity image or from an image reconstructed), this method achieves the most accurate focus estimation among current existing work, although its effectiveness still relies on having a reliable image representation rather than using events alone.

Summary. Table 5 provides a unified comparison of representative event-camera autofocus methods. The table highlights key differences in sensing inputs, focus criteria, and physical foundations. Prior approaches rely primarily on single-peak extremum search or polarity-based heuristics and lack a strong physical basis for their focus criteria. In contrast, our ESVA method uniquely leverages an analytically derived dual-peak–valley structure that arises directly from event-generation physics, enabling fast and robust autofocus across diverse scenes.

References

[1] Yuhan Bao, Lei Sun, Yuqin Ma, Diyang Gu, and Kaiwei Wang. Improving fast auto-focus with event polarity. *Optics Express*, 31(15):24025–24044, 2023. 4, 5

[2] Yuhan Bao, Shaohua Gao, Wenyong Li, and Kaiwei Wang. One-step event-driven high-speed autofocus. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*, pages 6222–6230, 2025. 5

[3] Zhou Ge, Haoyu Wei, Feng Xu, Yizhao Gao, Zhiqin Chu, Hayden K.-H. So, and Edmund Y. Lam. Millisecond autofocusing microscopy using neuromorphic event sensing. *Optics and Lasers in Engineering*, 160:107247, 2023. 4, 5

[4] Shijie Lin, Yinqiang Zhang, Lei Yu, Bin Zhou, Xiaowei Luo, and Jia Pan. Autofocus for event cameras. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*, pages 16344–16353, 2022. 4, 5