FIXED-INPUT STATE DEPENDENCE OF RESIDUAL TRANSMISSION IN SPIKING NEURAL NETWORKS

Anonymous authors
Paper under double-blind review

ABSTRACT

Residual connections in Spiking Neural Networks (SNNs) merge branch and shortcut signals before a post-add neuron, yet identical merged input need not yield identical transmitted spikes when the receiver retains membrane state. We isolate this dependence by fixing the realized post-add merge and replacing only the state carried from earlier transitions. For additive leaky integrate-and-fire (LIF) and parametric leaky integrate-and-fire (PLIF) dynamics, we derive an exact switching criterion. At a fixed merge, two retained states produce different spikes precisely when their state contributions straddle the firing margin. SpikeDF, a temporal residual SNN for video forgery detection, provides a binary case in which shortcut-only and branch-only origins yield the same unit-valued merged input but different retained states. Nearby cross-video state replacement switches 11.5–15.3% of deep residual outputs while changing mean output by at most $0.74\,\mathrm{pp}$; propagating only those changed spikes also alters detector scores. The same near-threshold concentration of switches appears in independently trained continuous-merge models on FaceForensics++ and post-add PLIF models on DVS128 Gesture. Across eight PLIF residual sites, pre-replacement boundary occupancy closely tracks the ordering of site-level switch rates, linking switching prevalence to firing-boundary geometry. Together, these results show that a fixed residual merge and nearly unchanged mean activity can coexist with threshold-structured changes in spike identity and downstream computation.

1 INTRODUCTION

Residual connections are central to deep Spiking Neural Networks (SNNs), and their computation is usually specified architecturally through branch representations and a chosen merge rule. A post-add spiking neuron, however, transforms that realized merge using membrane state retained from earlier transitions before emitting the signal seen downstream. The issue is therefore not simply that state enters a neuron recurrence, but whether a fixed residual merge fully specifies residual transmission, or whether distinct retained states can induce different spike-level outcomes under the same merged input.

Residual SNNs vary merge operators and shortcut representations to support deeper networks. Spike-Element-Wise ResNet (SEW-ResNet) modifies the residual operator to preserve identity mappings, while logical residual connections change merge algebra in the spike domain (Fang et al., 2021a; Shan et al., 2023). Membrane-based or densely additive designs alter what the shortcut carries and how information propagates across depth (Hu et al., 2024; Li et al., 2025). These designs establish how signals are represented, merged, and transported; our question instead conditions on the realized merge and asks what transmission changes remain when merged current is held fixed.

Neuron-model research separately establishes membrane state as part of spike generation. Parametric leaky integrate-and-fire (PLIF) neurons learn the membrane time constant, gated leaky integrate-and-fire (GLIF) models combine integration and reset dynamics, and recent studies modify initial membrane state, activity-dependent conductance, or longer-memory state laws (Fang et al., 2021b; Yao et al., 2022; Shen et al., 2024b; Bai et al., 2026; Ge et al., 2026). These models determine how temporal history enters subsequent spikes through decay, reset, conductance, or memory. The unresolved residual question is more conditional: after the merge has been realized, can state variation induce structured spike substitutions that are hidden by aggregate activity?

Reviewers: please read the Reviewer Guidelines (iclr.cc/Conferences/2027/ReviewerGuidelines) and the AI Policy for Reviewers (iclr.cc/Conferences/2027/AIPolicyForReviewers). If you used AI to expand, edit, or polish your review, please provide the input text to the LLM. Better still, consider skipping the LLM and submitting your original text: we, and the authors, are much more interested in your unedited thoughts than in what an LLM has to say. AI-assisted or not, you are putting your name and reputation behind your review: LLM-generated falsehoods, hallucinations or misrepresentations are subject to disciplinary action, which may include desk-rejecting all papers you have authored.

We therefore treat post-add residual transmission as the conditional mapping $S = \Gamma(M; R)$, where M is the realized merged current, R the retained membrane state, and S the transmitted spike. This formulation separates the architectural quantity being held fixed from the neuron variable being intervened on. It makes three mechanistic questions explicit: when changing only R alters S, how strongly that susceptibility depends on firing-boundary geometry and state pairing, and whether the resulting substitutions measurably affect downstream computation.

We answer these questions by fixing M and replacing only R. For additive leaky integrate-and-fire (LIF) and PLIF updates, a replacement switches the spike exactly when the two state contributions straddle the firing margin induced by the same merge. SpikeDF, a temporal residual SNN for video forgery detection, provides an exact binary instance: shortcut-only and branch-only residual origins can produce the same unit-valued merge while retaining different states.

These substantial spike-level changes nevertheless largely cancel at the population level. Nearby state replacement switches many residual outputs while barely changing mean activity, yet propagating only changed spikes alters detector scores; equal-count substitution shows that local susceptibility and downstream sensitivity need not rank sites identically. The same near-threshold concentration appears in continuous-merge FaceForensics++ models and a DVS128 Gesture setting with a different task, modality, and post-add neuron. Across eight sites, pre-replacement boundary occupancy tracks switch-rate ordering, linking susceptibility to firing-boundary geometry without donor states.

Our main contributions are summarized as follows:

- We formalize residual transmission by fixing merged input M and varying retained state R. For additive LIF and PLIF dynamics, switching follows an exact threshold crossing criterion, and prevalence depends on state pairing.
- In the exact binary equal input case of SpikeDF, nearby state replacement switches 11.5% to 15.3% of deep outputs while mean output changes by at most 0.74 pp. Propagating these changes and matching substitution counts distinguish local susceptibility from downstream sensitivity.
- Across discrete and continuous models with LIF and PLIF neurons, switching concentrates near threshold. Across eight DVS128 Gesture sites, boundary occupancy tracks switch rate ordering and provides a measure of site susceptibility without reference to donor choice.

2 Related Work

Residual Computation in SNNs. Residual SNNs vary branch representation, merge algebra, and post-merge transport. SEW-ResNet, OR residual connections, Multi Logical Residual Architecture (Multi-LRA), and XOResNet alter spike-domain merge operators, while activation-after-addition, Membrane-Shortcut ResNet (MS-ResNet), densely additive residual designs, and Spikingformer alter the representation presented to or propagated after the merge (Fang et al., 2021a; Shan et al., 2023; Shen et al., 2024a; Wu et al., 2026; Zheng et al., 2021a; Hu et al., 2024; Li et al., 2025; Zhou et al., 2023; 2026). Our analysis begins after the merge is formed and holds the resulting input M fixed. Table 1 summarizes three representative designs under a common merge-centered view.

Table 1: Representative residual SNN topologies under a common merge view.

Design	**Signals at merge**	**Post-merge representation**	
		Values	**Re-binarization**
Activation-after-add	shortcut + real-valued residual	real-valued	Immediate
SEW-ResNet (ADD)	spike + spike	$\{0, 1, 2\}$	None at merge
MS-ResNet	membrane shortcut + residual	real-valued	Before transform

Stateful Neuronal Dynamics. Neuron-dynamics work changes how retained state converts current input into spikes. PLIF learns the membrane time constant, Residual Membrane Potential SNN (RMP-SNN) changes reset, and GLIF parameterizes decay, integration, and reset (Fang et al., 2021b; Han et al., 2020; Yao et al., 2022). Two-compartment and complementary LIF models extend temporal state dynamics or improve gradient propagation, while other work changes initial membrane state, activity-dependent conductance, or longer-memory state laws (Ponghiran & Roy, 2022; Zhang et al.,

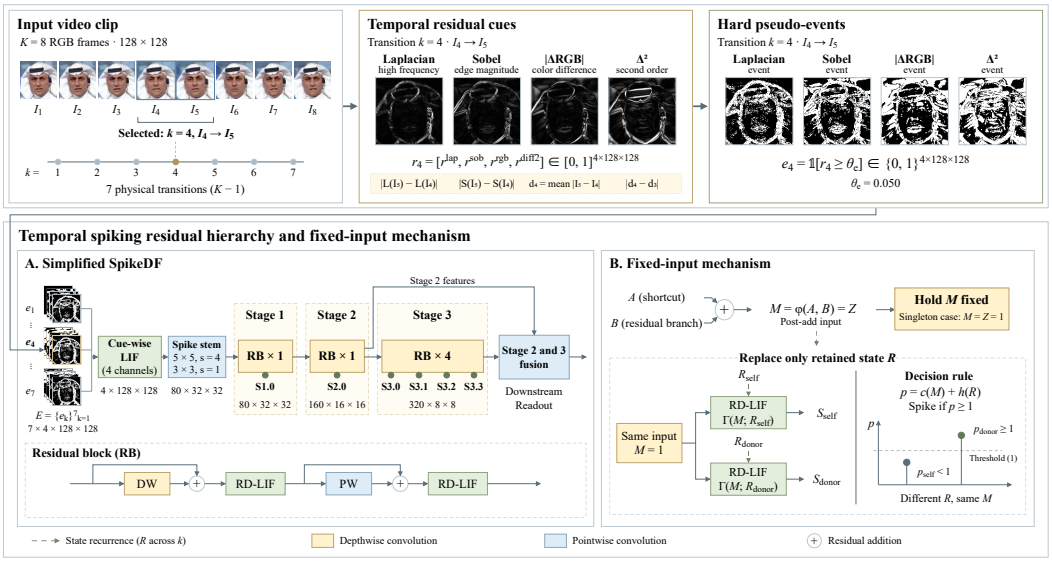


Figure 1: SpikeDF computation and fixed-input state replacement. Seven ordered transitions from an eight-frame clip enter the temporal residual network. At a selected post-add neuron, merge M is preserved and only retained state R is replaced before the output spike is produced.

2024; Huang et al., 2024; Shen et al., 2024b; Bai et al., 2026; Ge et al., 2026). Adversarial-robustness work identifies threshold-neighboring neurons as prone to state flips under small disturbances (Wang et al., 2026). Fixed-input state replacement instead keeps both the realized merge M and neuron update rule unchanged and exchanges only retained state R, isolating whether state alone changes post-add residual transmission.

Internal State Interventions. Causal mediation, interchange interventions, and causal tracing change internal representations or states to measure their functional role in model behavior (Vig et al., 2020; Geiger et al., 2021; Meng et al., 2022), while causal abstraction provides a formal language for such interventions (Geiger et al., 2025). Theory for SNNs has also identified causal regions in which output spike times vary smoothly with input spike times and network parameters (Dold & Petersen, 2025). We apply this intervention logic to retained membrane state after residual addition while holding the surrounding merge fixed.

Temporal Video Forgery Detection. Video forgery detection uses ordered temporal evidence across adjacent frames (Haliassos et al., 2021; Zheng et al., 2021b; Wang et al., 2023; Guo et al., 2025), and spike-driven temporal cues have also been used to detect generated videos (Jang et al., 2026). Adjacent-frame transitions provide ordered updates in which retained neuronal state evolves across the clip.

3 METHOD

SpikeDF processes $K = 8$ red, green, and blue (RGB) frames into $T = 7$ temporal transitions. The four-channel residual cue at transition k is denoted by r_k, its thresholded pseudo-event at $\theta_{\mathrm{e}} = 0.050$ by e_k, and the ordered event sequence by $E = \{e_k\}_{k=1}^{7}$.

In Figure 1, A and B are residual branch outputs, $M = \phi(A, B)$ is their post-add merge (Z in SpikeDF), R is retained membrane state, $S = \Gamma(M; R)$ is the output spike, and $p = c(M) + h(R)$ is the pre-threshold quantity, where c and h denote merge and state contributions. R_{self} and R_{donor} denote the recorded and substituted states, and a singleton residual merge has only one active branch. We use residual-decision LIF (RD-LIF) to denote the placement of a conventional one-compartment LIF immediately after a residual sum; sites use stage-position labels S1.0, S2.0, and S3.0–S3.3.

3.1 State-Conditioned Residual Transmission

We now index the quantities in Figure 1 by a scalar output position j and residual site ℓ and allow the merge rule to depend on residual topology $\mathcal{T}_\ell$:

$$M_{j,\ell} = \phi_{\mathcal{T}_\ell}(A_{j,\ell}, B_{j,\ell}), \qquad S_{j,\ell} = \Gamma_\ell(M_{j,\ell}; R_{j,\ell}). \tag{1}$$

Here $M_{j,\ell}$ is the neuron input, $S_{j,\ell} \in \{0,1\}$, and $R_{j,\ell}$ is the membrane state retained immediately before the current LIF/PLIF update. Our comparisons condition on M and vary R, independent of which branch configuration produced that merge.

At physical transition k, the two branch signals first form M_k. The neuron combines this input with membrane state carried from earlier transitions, emits the current spike, and then updates the state available to transition $k+1$. Each neuron therefore carries its own state trajectory across transitions.

For the additive LIF and PLIF updates analyzed here, the normalized pre-threshold state separates into merge and state contributions,

$$p = c(M) + h(R), \qquad S = \mathbb{I}[p \geq 1], \qquad m(M) = 1 - c(M), \tag{2}$$

where $\mathbb{I}[\cdot]$ denotes the indicator function and $m(M)$ is the signed threshold margin set by the fixed merge. Positive $m(M)$ denotes a merge contribution below threshold, whereas negative $m(M)$ denotes one already above threshold. Holding M fixed therefore holds $c(M)$ and $m(M)$ fixed, so state replacement changes only $h(R)$.

For two states R_a and R_b at fixed merge M, let $h_a = h(R_a)$ and $h_b = h(R_b)$. Their outputs differ exactly when the fixed threshold margin lies between the two state contributions,

$$S(M, R_a) \neq S(M, R_b) \quad \Longleftrightarrow \quad \min(h_a, h_b) < m(M) \leq \max(h_a, h_b). \tag{3}$$

Switching occurs when (h_a, h_b) straddles the fixed boundary $m(M)$. Under a specified pairing distribution over R_a and R_b, averaging this crossing event gives

$$q_{\mathrm{sw}} = \Pr[\min(h_a, h_b) < m(M) \leq \max(h_a, h_b)]\,. \tag{4}$$

Equation (3) is pointwise, whereas Eq. (4) is a population quantity and therefore also depends on how original and donor states are paired around the fixed boundary. For a conditioning context X, let q_{s} and q_{d} be the probabilities that original and donor state contributions fall below the fixed margin. If J is the conditional probability that both do so, define $C(X) = J - q_{\mathrm{s}} q_{\mathrm{d}}$. Then

$$q_{\mathrm{sw}}(X) = q_{\mathrm{s}}(1 - q_{\mathrm{d}}) + (1 - q_{\mathrm{s}}) q_{\mathrm{d}} - 2C(X).$$

If original and donor states are conditionally independent given X and share the marginal $q_X = q_{\mathrm{s}} = q_{\mathrm{d}}$, this reduces to $2q_X(1-q_X)$; pooling unresolved context before forming the independent reference introduces the nonnegative gap $2\,\mathrm{Var}(q_X)$. Boundary occupancy therefore captures pre-replacement susceptibility but does not by itself determine switch prevalence. Appendix A.2 gives the derivation and boundary bounds. The same pairwise crossing condition applies when the merge moves from a discrete to a continuous range.

3.2 Fixed-Input State Replacement

For an output position j, $R_{j,\ell}$ is the self state and $R_{v,\ell}$ is a donor state recorded at position v. We keep merge $M_{j,\ell}$ unchanged and replace only the self state:

$$S^{\mathrm{self}}_{j,\ell} = \Gamma_\ell(M_{j,\ell}; R_{j,\ell}), \qquad S^{j\leftarrow v}_{j,\ell} = \Gamma_\ell(M_{j,\ell}; R_{v,\ell}), \qquad \Delta_{jv,\ell} = \mathbb{I}[S^{\mathrm{self}}_{j,\ell} \neq S^{j\leftarrow v}_{j,\ell}]. \tag{5}$$

A nonzero $\Delta_{jv,\ell}$ marks a local spike change at fixed M. The counterfactual output is computed from the original merge $M_{j,\ell}$ and donor state $R_{v,\ell}$.

State replacement changes R for one neuron update while keeping M and neuron parameters fixed. For downstream score analysis, only spikes that changed under replacement are substituted at the selected site, after which the frozen network resumes from the recorded recurrent trajectory. State removal and scaling, which modify subsequent trajectories, are analyzed separately in Appendix C.3.

To decompose q_{sw} by switch direction, define $q^+ = \Pr(S^{\mathrm{self}} = 0, S^{j\leftarrow v} = 1)$ and $q^- = \Pr(S^{\mathrm{self}} = 1, S^{j\leftarrow v} = 0)$. Then

$$q_{\mathrm{sw}} = q^+ + q^-, \qquad \Delta_{\mathrm{mean}} = \mathbb{E}[S^{j\leftarrow v} - S^{\mathrm{self}}] = q^+ - q^-. \tag{6}$$

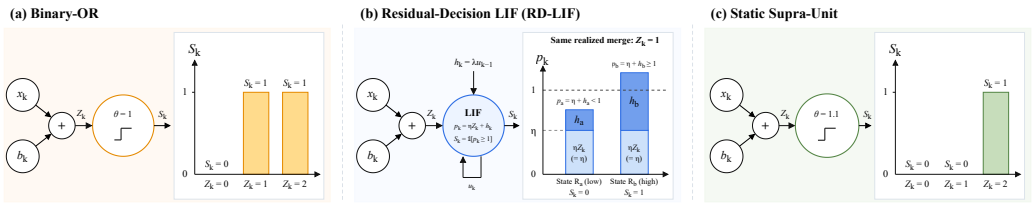


Figure 2: Transmission at fixed $Z_k = 1$. Binary-OR accepts and Static Supra-Unit rejects $Z_k = 1$. RD-LIF combines this merge with state contribution h_k, so different states can place $p_k = \eta Z_k + h_k$ on opposite sides of threshold.

Here q_{sw} counts positions whose binary output changes, whereas Δ_{mean} is the signed change in mean output. Hence $\Delta_{\mathrm{mean}} = 0$ is compatible with $q_{\mathrm{sw}} > 0$ whenever $q^+ = q^- > 0$.

For continuous summaries we also use unsigned threshold distance $d(M) = |m(M)| = |1 - c(M)|$. The signed margin in Eq. (3) determines each crossing, while $d(M)$ groups outputs by distance from threshold. We further define boundary occupancy before state replacement as $B_\ell(\epsilon) = \Pr_\ell[d(M) < \epsilon]$, the share of outputs within distance ϵ of the boundary. Because B_ℓ depends only on M and neuron parameters, it can be computed before donor replacement. Appendix A.2 derives the corresponding population decomposition and bounds.

3.3 The Z=1 Equal-Input Case in SpikeDF

The overview notation Z becomes exact in SpikeDF because both operands at each residual addition are binary at evaluation. At physical transition k, let $x_k \in \{0, 1\}$ be the shortcut spike and $b_k \in \{0, 1\}$ the transformed branch spike. With DropPath disabled at evaluation,

$$Z_k = x_k + b_k \in \{0, 1, 2\}. \tag{7}$$

The three merge values correspond to silent agreement, $Z = 1$ disagreement, and active agreement. Shortcut-only $(1, 0)$ and branch-only $(0, 1)$ both realize $Z = 1$ and therefore present the same scalar current to the neuron. The $Z = 1$ subset consequently compares different states at fixed input.

SpikeDF uses RD-LIF, a one-compartment LIF placed immediately after each residual sum. Let $u_k = V_k/V_{\mathrm{r}}$ be the normalized post-reset membrane, $\lambda = \exp(-1/\tau)$ the retention factor per transition, $\eta = g/V_{\mathrm{r}}$ the ratio of input gain to threshold, and κ the subtractive reset scale. Its normalized forward recurrence is

$$p_k = \lambda u_{k-1} + \eta Z_k, \qquad S_k = \mathbb{I}[p_k \geq 1], \qquad u_k = p_k - \kappa S_k. \tag{8}$$

Thus Eq. (2) specializes to $c(Z_k) = \eta Z_k$ and $h(R) = \lambda u_{k-1}$. A common positive rescaling of membrane, input gain, and firing threshold leaves the forward spike output unchanged, so current strength enters through the ratio $\eta = g/V_{\mathrm{r}}$. The unnormalized recurrence, its relation to standard LIF parameterizations, and the surrogate gradient training rule are given in Appendices A.1 and D.2.

When $1/2 < \eta < 1$, current alone places $Z = 1$ below threshold while $Z = 2$ is already above threshold. Within $h_k = \lambda u_{k-1} \in [1 - 2\eta, 1)$, where the $Z = 0$ and $Z = 2$ outputs remain unchanged, crossing $h_k = 1 - \eta$ changes only the $Z = 1$ output. We use this parameter range for the analysis. Figure 2 contrasts it with two memoryless reference rules on the same residual alphabet: Binary-OR accepts every $Z = 1$ input, Static Supra-Unit rejects every $Z = 1$ input, and RD-LIF transmits according to whether state moves the same merge across threshold.

At evaluation, each RD-LIF maintains its own membrane state and immediately re-binarizes the local residual sum. The $\{0, 1, 2\}$ merge therefore exists only between binary addition and the next spike output, after which the transformed branch again receives a binary spike. SpikeDF applies this ordering after both depthwise and pointwise residual additions and at multiscale fusion. Residual block equations are given in Appendix D.2.

3.4 Temporal Sequence and Analysis Sites

Given a clip $\mathcal{X} = \{I_i\}_{i=1}^{K}$ of $K = 8$ ordered RGB frames, SpikeDF forms $T = K - 1 = 7$ physical frame transitions. We reserve i for frame index, k for transition index, and ℓ for network depth. Every

Table 2: Matched post-add controls on FaceForensics++ validation. Recurrence denotes state carried across transitions; late activity is the mean active-state density over Stage 3 and fusion.

Blocks	Residual rule	Recurrence	Dynamics	Late activity	Task AUC
4	Static Supra-Unit	no	–	$0.149 \pm 0.023\%$	82.23 ± 1.65
4	Fixed-Stateful LIF	yes	fixed	$11.02 \pm 2.26\%$	90.95 ± 0.23
4	RD-LIF	yes	learned	$10.47 \pm 0.96\%$	91.08 ± 0.37
4	Binary-OR	no	–	$64.25 \pm 0.85\%$	92.27 ± 0.61
6	Static Supra-Unit	no	–	$0.089 \pm 0.023\%$	73.97 ± 1.71
6	Fixed-Stateful LIF	yes	fixed	$4.92 \pm 0.75\%$	91.31 ± 0.65
6	RD-LIF	yes	learned	$4.80 \pm 0.14\%$	90.76 ± 0.20
6	Binary-OR	no	–	$85.31 \pm 0.46\%$	92.52 ± 0.89

clip resets membrane state to zero, and recurrent state is carried only along the physical transition axis k within that clip.

The overview cues r_k are constructed from four adjacent-frame residuals: Laplacian, Sobel, absolute RGB difference, and a second-order difference. After normalization and concatenation, thresholding gives the binary pseudo-event tensor $e_k = \mathbb{I}[r_k \geq \theta_{\mathrm{e}}]$ with $\theta_{\mathrm{e}} = 0.050$. These synchronized pseudo-events form the temporally ordered input to Stage 1. State replacement is applied later, after residual addition. Cue definitions, tensor shapes, and input representation comparisons are given in Appendices D.2 and C.2.

The temporal backbone has three residual stages: Stages 1 and 2 contain one block each, while Stage 3 contains the remaining blocks. The four- and six-block variants are denoted SpikeDF-C and SpikeDF-S. Projected Stage 2 spikes are combined with Stage 3 spikes and re-binarized by a dedicated neuron before pooling. Spatially averaged spikes pass through a spiking linear projection, and learned softmax weights aggregate the seven transition features. A separate continuous branch summarizes pre-threshold residual magnitude and enters only at final fusion before a multilayer perceptron, outside the binary residual network. Appendix D.2 gives stage widths, fusion, readout, and optimization details, and Appendix C.1 reports the matched control with this auxiliary path removed.

4 Experiments

4.1 Experimental Protocol

The experiments measure spike switching at fixed input, cancellation in mean output, downstream score changes, and threshold dependence with continuous merges. The main SpikeDF models are trained on FaceForensics++ (FF++) under c23 compression (Rossler et al., 2019). We report the video-level area under the receiver operating characteristic curve (AUC) for task performance. Mechanism measurements are switch rate, signed change in mean output, and downstream score change. Unless stated otherwise, $\pm$ denotes sample standard deviation across independently trained models, and paired differences are computed within model before averaging. Differences between percentage quantities are reported as percentage-point differences using pp.

Preselected held-out operating points and complete task metrics are reported in Appendix C.1. Fixed-input analyses use FF++ validation data. The local replacement cohort contains one deterministic first clip from 128 unique validation videos (64 real and 16 from each manipulation family), with donor states drawn from different videos. Model seed ($n = 3$) is the replication unit. Appendix D specifies splits, donor assignment, aggregation, and analysis protocols.

Recurrent and Memoryless Controls. A recurrent LIF with fixed parameters (Fixed-Stateful LIF) closely follows the learned RD-LIF activity profile, whereas the memoryless endpoints produce much sparser or denser late activity (Table 2).

4.2 Local Switching Under Fixed-Input Replacement

The crossing condition in Eq. (3) predicts output switches whenever state replacement straddles the $Z = 1$ margin. Before substitution, each recorded post-add output is reconstructed from the frozen

Table 3: Nearby-donor robustness at fixed residual input. Random uses different-video donors; nearest uses the rank-1 match in state-summary space. Retained is the nearest-to-random switch-rate ratio.

Site	Random switch (%)	Nearest switch (%)	Retained (%)	Self output (%)	Nearest output (%)	Δ output (pp)
S1.0	38.09 ± 0.93	33.46 ± 0.83	87.8	57.09 ± 0.32	49.92 ± 1.31	-7.16
S2.0	27.63 ± 2.15	24.25 ± 1.94	87.8	47.18 ± 0.56	44.90 ± 0.30	-2.27
S3.0	14.03 ± 1.75	13.33 ± 1.69	95.1	49.39 ± 0.95	49.35 ± 0.99	-0.04
S3.1	13.67 ± 0.84	13.26 ± 0.89	97.0	50.09 ± 1.75	50.32 ± 1.89	$+0.23$
S3.2	15.58 ± 0.92	15.26 ± 0.75	98.0	37.08 ± 1.52	37.83 ± 1.37	$+0.74$
S3.3	11.78 ± 0.56	11.46 ± 0.60	97.2	23.77 ± 2.81	23.54 ± 2.89	-0.22

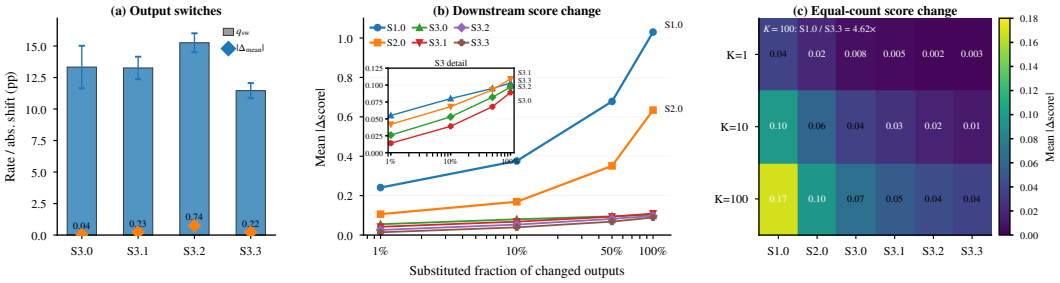


Figure 3: Output switches and downstream score changes. (a) Fixed-input switch rate and absolute change in mean output at deep Stage 3 sites. (b) Score change with increasing substitution fraction. (c) Equal-count score changes for $K = 1$, 10, and 100 across sites. Error bars denote sample standard deviation across three checkpoints.

neuron quantities and retained state. Across 1,769,472 sampled output–donor pairs spanning six sites and three checkpoints, replay reconstructs every sampled spike and every observed switch indicator agrees with Eq. (3) at the specified numerical tolerance.

To limit donor mismatch, the primary comparison selects the nearest donor from a different video in a summary space of $Z = 1$ activity, retained state, current, and threshold distance. Table 3 compares this matched intervention with unrestricted random donors at all six sites.

Nearest matching preserves most random-donor switching at every site (Table 3). Across the four deep sites, 11.46–15.26% of outputs still switch while signed changes in mean output remain within 0.74 pp, consistent with near cancellation between $0 \to 1$ and $1 \to 0$ events. A matched independent-marginal reference systematically overestimates random-donor switching, showing that population prevalence depends on original–donor pairing beyond marginal boundary occupancy (Appendix B.2).

4.3 Downstream Consequences of Output Switching

For each $Z = 1$ position, the counterfactual spike is computed from the original merged input and the donor's retained state. We substitute only outputs changed by fixed-input state replacement at the selected site and then resume the frozen network from its unmodified downstream computation. Re-inserting the recorded spikes changes detector scores by at most 10^{-6}, providing a propagation sanity check. Fractional substitution increases the fraction of changed outputs propagated, whereas equal-count substitution fixes the number K across sites.

Table 4: Equal-count downstream score changes. The same number K of changed outputs is substituted at each site. Entries report mean $|\Delta\text{score}|$ ($\pm$ sample standard deviation across three checkpoints).

Site	$K = 1$	$K = 10$	$K = 100$
S1.0	0.0372 ± 0.0018	0.0974 ± 0.0047	0.1677 ± 0.0037
S2.0	0.0203 ± 0.0008	0.0587 ± 0.0017	0.0952 ± 0.0027
S3.0	0.0076 ± 0.0018	0.0357 ± 0.0059	0.0658 ± 0.0044
S3.3	0.0028 ± 0.0002	0.0130 ± 0.0015	0.0363 ± 0.0032

Score change grows with the substituted fraction, while equal-count substitution reveals marked site dependence: at $K = 100$, the S1.0 score change is $4.62\times$ that at S3.3.

4.4 Threshold-Dependent Switching Across Residual Models

Independently trained activation-after-addition SpikingResNet19 models on FF++ and post-add PLIF models on DVS128 Gesture extend the fixed-input analysis to continuous merges and to a distinct task, input modality, and post-add neuron. The DVS128 Gesture network has eight residual sites, labeled B1–B8 from shallow to deep. Both model families are evaluated on their native tasks before intervention; per-seed video AUC and test accuracy are reported in Table B.1 (Appendix B).

In independently trained activation-after-addition FF++ models, more than 99.997% of merges are noninteger, yet state replacement produces far more switches near threshold than far from it. Across all eight DVS128 Gesture sites, switch rates near threshold are in the tens of percent, while rates far from threshold remain below 0.35%.

Outputs near threshold form a minority before replacement but contain most switches in both experiments. Across B1–B8, boundary occupancy $B_\ell(0.40)$ has Spearman rank correlation $\rho_s = 0.976$ with switch rate. Restricting continuous-FF++ donors to the same label or manipulation changes near-boundary switch rate by at most 0.65 pp relative to unrestricted donors across the three sites; DVS128 Gesture replay has zero spike mismatches and zero charge reconstruction error. Full statistics are in Appendices B.4 and B.5.

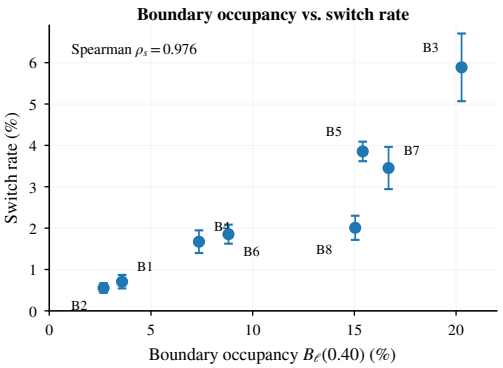


Figure 4: Boundary occupancy and switch rate. Each point denotes one DVS128 Gesture PLIF site. The axes show $B_\ell(0.40)$ and switch rate under state replacement using a donor from another sample of the same class.

Table 5: Threshold-distance switching across residual models. Near/far use $d \le 0.10/d \ge 0.40$; boundary uses $d < 0.40$, with columns giving its shares of all outputs and all switches. DVS128 Gesture rows are prespecified shallow, middle, and deep sites.

Experiment	Site	Near switch (%)	Far switch (%)	Boundary outputs (%)	Boundary switches (%)
FF++ continuous LIF	Stage 1	36.80 ± 1.64	0.107 ± 0.013	6.72	94.31
	Stage 2	36.05 ± 0.53	0.186 ± 0.051	7.79	91.92
	Stage 3	37.83 ± 0.99	0.105 ± 0.020	6.24	94.04
DVS128 Gesture PLIF	B1	36.07 ± 0.80	0.0818 ± 0.0227	3.58	88.67
	B4	34.62 ± 1.16	0.1099 ± 0.0272	7.36	93.92
	B8	28.62 ± 0.66	0.0286 ± 0.0158	15.04	98.85

4.5 Fixed-Input Switching Under Distribution Shift

We evaluate the same three FF++-trained SpikeDF-S checkpoints on Celeb-DF-v2 under a different forgery distribution without adaptation. Table 6 compares fixed-input switching at the matched residual sites with the corresponding FF++ measurements.

Table 6: Fixed-input switching on Celeb-DF-v2. Δ switch is paired to FF++ by checkpoint; mean gap is the absolute self–cross output difference (pp). Origin columns use fixed $Z = 1$ input.

Site	Celeb-DF switch (%)	Δ switch (pp)	Mean output gap (pp)	Shortcut-only switch (%)	Branch-only switch (%)
S1.0	37.41 ± 1.00	-1.02 ± 0.06	10.81	37.91 ± 1.04	36.88 ± 1.72
S3.1	13.97 ± 1.40	$+0.28 \pm 0.54$	0.12	16.06 ± 0.99	12.33 ± 1.81
S3.3	11.96 ± 0.30	$+0.10 \pm 0.19$	0.24	15.79 ± 1.28	6.65 ± 1.03

Switching at the two deep sites remains close to FF++, whereas S1.0 shifts more (Table 6). Across the deep sites, the mean output gap is at most $0.24\,\mathrm{pp}$; the S1.0 switch-rate change is $-1.02 \pm 0.06\,\mathrm{pp}$. At fixed $Z = 1$, shortcut-only and branch-only cases also show different switch fractions at the deeper sites, where the two origins have different state distributions. Full task metrics and state statistics are reported in Appendix C.4.

Trajectory-Level State Controls. Fixed-input replacement isolates local state dependence, while complementary recursive interventions test the role of retained state over complete trajectories. The effect is strongly site dependent: removing Stage 3 state alone changes video AUC by only $0.19\,\mathrm{pp}$, whereas removing Stage 3 together with fusion lowers AUC by $44.88\,\mathrm{pp}$; interventions at earlier stages are also substantial. These controls address trajectory-level dependence rather than fixed-input counterfactuals, separating local susceptibility from broader functional dependence on recurrent state. The complete sweep, including cross-video history and all stage-wise interventions, is reported in Appendix C.3.

5 Discussion

Fixed-input replacement separates merge semantics from transmission semantics. A residual merge fixes the current presented to the post-add neuron, whereas the transmitted spike also depends on retained state. For residual blocks with a stateful post-add update, branch signals and merge algebra therefore do not alone specify the propagated spike. The experiments further distinguish local susceptibility to fixed-input state replacement, downstream sensitivity to a controlled set of changed spikes, and trajectory-level dependence on recurrent state. These quantities need not rank residual sites identically, and opposing $0 \to 1$ and $1 \to 0$ switches can additionally cancel in the mean.

Population switch prevalence is likewise not determined by threshold proximity alone. Boundary occupancy captures pre-replacement susceptibility, whereas the paired retained-state distribution determines how often that boundary is crossed; independent matched marginals can overestimate switching when unresolved context remains heterogeneous. This separation persists across discrete and continuous LIF/PLIF models and for frozen models under distribution shift. The analysis assumes additive post-add updates with pre-threshold state $c(M) + h(R)$; other orderings require analogous decompositions, while ordinary inference follows the data- and dynamics-induced joint distribution of M and R.

6 Conclusion

Residual addition determines the input to a stateful post-add neuron but not necessarily its transmitted spike. Across discrete and continuous residual models, retained state most often changes outputs near threshold and can alter downstream scores despite small mean-output changes. Population switching also depends on state pairing around the firing boundary, while local susceptibility, downstream sensitivity, and trajectory-level dependence need not rank sites identically. Residual SNN analysis should therefore distinguish merge semantics from the state-dependent transmission that follows.

AI USE STATEMENT

In this work, we used generative AI tools for language editing. We reviewed the AI-assisted work, verified the reported values, and take responsibility for the final content of the submission.

ETHICS STATEMENT

We use the established FaceForensics++, Celeb-DF-v2, and DVS128 Gesture benchmarks and collect no new human-subject data. Face-forgery detectors can produce false positives and should not be used as sole evidence for identity, authorship, or misconduct decisions. Demographic fairness and deployment calibration remain unevaluated and would be necessary before high-stakes use.

REPRODUCIBILITY STATEMENT

Section 4.1 defines the common experimental protocol, and Sections 4.2–4.5 report the principal mechanism evaluations. Appendix B contains the detailed validation and cross-model evidence, Appendix C the control and distribution-shift analyses, and Appendix D the evaluation map and complete reproduction protocols.

REFERENCES

Qianyi Bai, Wang Wang, and Qiang Yu. A brain-inspired gating mechanism unlocks robust computation in spiking neural networks. In *International Conference on Learning Representations*, volume 2026, pp. 156097–156129, 2026.

Dominik Dold and Philipp Christian Petersen. Causal pieces: analysing and improving spiking neural networks piece by piece. *arXiv preprint arXiv:2504.14015*, 2025.

Wei Fang, Zhaofei Yu, Yanqi Chen, Tiejun Huang, Timothée Masquelier, and Yonghong Tian. Deep residual learning in spiking neural networks. volume 34, pp. 21056–21069, 2021a.

Wei Fang, Zhaofei Yu, Yanqi Chen, Timothée Masquelier, Tiejun Huang, and Yonghong Tian. Incorporating learnable membrane time constant to enhance learning of spiking neural networks. In *2021 IEEE/CVF International Conference on Computer Vision (ICCV)*, pp. 2641–2651. IEEE, 2021b.

Chengjie Ge, Yufeng Peng, Zihao Li, Qiyu Kang, Xueyang Fu, Xuhao Li, Qixin Zhang, Ren Junhao, and Zheng-Jun Zha. Fractional-order spiking neural network. In *International Conference on Learning Representations*, volume 2026, pp. 79651–79696, 2026.

Atticus Geiger, Hanson Lu, Thomas Icard, and Christopher Potts. Causal abstractions of neural networks. volume 34, pp. 9574–9586, 2021.

Atticus Geiger, Duligur Ibeling, Amir Zur, Maheep Chaudhary, Sonakshi Chauhan, Jing Huang, Aryaman Arora, Zhengxuan Wu, Noah Goodman, Christopher Potts, et al. Causal abstraction: A theoretical foundation for mechanistic interpretability. *Journal of Machine Learning Research*, 26 (83):1–64, 2025.

Zonghui Guo, Yingjie Liu, Jie Zhang, Haiyong Zheng, and Shiguang Shan. Face forgery video detection via temporal forgery cue unraveling. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, pp. 7396–7405, 2025.

Alexandros Haliassos, Konstantinos Vougioukas, Stavros Petridis, and Maja Pantic. Lips don't lie: A generalisable and robust approach to face forgery detection. In *2021 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*, pp. 5037–5047. IEEE, 2021.

Bing Han, Gopalakrishnan Srinivasan, and Kaushik Roy. Rmp-snn: Residual membrane potential neuron for enabling deeper high-accuracy and low-latency spiking neural network. In *2020 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*, pp. 13555–13564. IEEE, 2020.

Yifan Hu, Lei Deng, Yujie Wu, Man Yao, and Guoqi Li. Advancing spiking neural networks toward deep residual learning. *IEEE transactions on neural networks and learning systems*, 36(2): 2353–2367, 2024.

Yulong Huang, Xiaopeng Lin, Hongwei Ren, Haotian Fu, Yue Zhou, Zunchang Liu, Biao Pan, and Bojun Cheng. Clif: Complementary leaky integrate-and-fire neuron for spiking neural networks. 2024.

Minsuk Jang, Yujin Yang, Hee-Seon Kim, Minseok Son, Younghun Kim, and Changick Kim. Detecting ai-generated videos with spiking neural networks. *arXiv preprint arXiv:2605.05895*, 2026.

Yudong Li, Yunlin Lei, and Xu Yang. Rethinking residual connection in training large-scale spiking neural networks. *Neurocomputing*, 616:128950, 2025.

Kevin Meng, David Bau, Alex Andonian, and Yonatan Belinkov. Locating and editing factual associations in gpt. volume 35, pp. 17359–17372, 2022.

Emre O Neftci, Hesham Mostafa, and Friedemann Zenke. Surrogate gradient learning in spiking neural networks: Bringing the power of gradient-based optimization to spiking neural networks. *IEEE Signal Processing Magazine*, 36(6):51–63, 2019.

Wachirawit Ponghiran and Kaushik Roy. Spiking neural networks with improved inherent recurrence dynamics for sequential learning. In *Proceedings of the AAAI conference on artificial intelligence*, volume 36, pp. 8001–8008, 2022.

Andreas Rossler, Davide Cozzolino, Luisa Verdoliva, Christian Riess, Justus Thies, and Matthias Nießner. Faceforensics++: Learning to detect manipulated facial images. In *Proceedings of the IEEE/CVF international conference on computer vision*, pp. 1–11, 2019.

Yimeng Shan, Xuerui Qiu, Rui-jie Zhu, Ruike Li, Meng Wang, and Haicheng Qu. Or residual connection achieving comparable accuracy to add residual connection in deep residual spiking neural networks. *arXiv preprint arXiv*, 2311, 2023.

Hangchi Shen, Huamin Wang, Yuqi Ma, Long Li, Shukai Duan, and Shiping Wen. Multi-lra: Multi logical residual architecture for spiking neural networks. *Information Sciences*, 660:120136, 2024a.

Hangchi Shen, Qian Zheng, Huamin Wang, and Gang Pan. Rethinking the membrane dynamics and optimization objectives of spiking neural networks. volume 37, pp. 92697–92720, 2024b.

Jesse Vig, Sebastian Gehrmann, Yonatan Belinkov, Sharon Qian, Daniel Nevo, Yaron Singer, and Stuart Shieber. Investigating gender bias in language models using causal mediation analysis. volume 33, pp. 12388–12401, 2020.

Shuai Wang, Malu Zhang, Yulin Jiang, Dehao Zhang, Ammar Belatreche, Yu Liang, Yimeng Shan, Zijian Zhou, Yang Yang, and Haizhou Li. Robust spiking neural networks against adversarial attacks. 2026.

Zhendong Wang, Jianmin Bao, Wengang Zhou, Weilun Wang, and Houqiang Li. Altfreezing for more general video face forgery detection. In *2023 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*, pp. 4129–4138. IEEE, 2023.

Jianfang Wu, Junsong Wang, Qiang Yu, and Zhiwei Gao. Xoresnet: Exclusive-or meta-residuals facilitate deep spiking neural networks learning. *Neurocomputing*, pp. 134933, 2026.

Xingting Yao, Fanrong Li, Zitao Mo, and Jian Cheng. Glif: A unified gated leaky integrate-and-fire neuron for spiking neural networks. volume 35, pp. 32160–32171, 2022.

Shimin Zhang, Qu Yang, Chenxiang Ma, Jibin Wu, Haizhou Li, and Kay Chen Tan. Tc-lif: A two-compartment spiking neuron model for long-term sequential modelling. In *Proceedings of the AAAI conference on artificial intelligence*, volume 38, pp. 16838–16847, 2024.

Hanle Zheng, Yujie Wu, Lei Deng, Yifan Hu, and Guoqi Li. Going deeper with directly-trained larger spiking neural networks. In *Proceedings of the AAAI conference on artificial intelligence*, volume 35, pp. 11062–11070, 2021a.

Yinglin Zheng, Jianmin Bao, Dong Chen, Ming Zeng, and Fang Wen. Exploring temporal coherence for more general video face forgery detection. In *2021 IEEE/CVF International Conference on Computer Vision (ICCV)*, pp. 15024–15034. IEEE, 2021b.

Chenlin Zhou, Liutao Yu, Zhaokun Zhou, Zhengyu Ma, Han Zhang, Huihui Zhou, and Yonghong Tian. Spikingformer: Spike-driven residual learning for transformer-based spiking neural network. *arXiv e-prints*, pp. arXiv–2304, 2023.

Chenlin Zhou, Liutao Yu, Zhaokun Zhou, Han Zhang, Jiaqi Wang, Huihui Zhou, Zhengyu Ma, and Yonghong Tian. Spikingformer: A key foundation model for spiking neural networks. In *Proceedings of the AAAI Conference on Artificial Intelligence*, volume 40, pp. 2236–2244, 2026.

A Theory of Fixed-Input State Dependence

Unrolling the post-add recurrence isolates the retained state contribution in the crossing condition at fixed input. The pairwise threshold relation then induces a population switch rate through the joint distribution of original and donor states, with SpikeDF's binary $Z = 1$ case and the continuous-input LIF as two concrete realizations.

A.1 Retained State in Post-Add Dynamics

RD-LIF is a one-compartment LIF placed immediately after residual addition. In unnormalized variables, its forward recurrence at physical transition k is

$$\begin{aligned} H_k &= \lambda V_{k-1} + gZ_k, \\ S_k &= \mathbb{I}[H_k \geq V_{\mathrm{r}}], \\ V_k &= H_k - \kappa V_{\mathrm{r}} S_k, \end{aligned} \tag{9}$$

where H_k is the pre-threshold membrane, V_k is the post-reset state retained for the next transition, $\lambda = \exp(-1/\tau)$ for a unit transition step, g is input gain, V_{r} is firing threshold, and κ scales subtractive reset. Equation (9) is the standard exponential-leak, subtractive-reset LIF recurrence. RD-LIF refers to its placement after residual addition.

With the normalized variables in Eq. (8), unrolling the recurrence over n transitions gives

$$u_k = \lambda^n u_{k-n} + \sum_{j=0}^{n-1} \lambda^j \big(\eta Z_{k-j} - \kappa S_{k-j}\big). \tag{10}$$

A residual input j transitions in the past is weighted by $\lambda^j = \exp(-j/\tau)$ before subsequent resets. The retained state is therefore a function of both earlier residual merges and earlier spikes and resets. At $\kappa = 1$, the recurrence uses standard reset-by-subtraction and carries threshold overshoot forward (Han et al., 2020). Learned κ changes the amount of state retained after a spike.

A.2 Population Switching and Boundary Bounds

Section 3.1 states the population decomposition that separates marginal boundary occupancy from original–donor pairing. Here we derive that relation and give two bounds that connect switching to threshold proximity.

For a conditioning context X, let $A_{\mathrm{s}} = \mathbb{I}[h_{\mathrm{s}} < m]$, $A_{\mathrm{d}} = \mathbb{I}[h_{\mathrm{d}} < m]$, $q_{\mathrm{s}} = \Pr(A_{\mathrm{s}} = 1 \mid X)$, $q_{\mathrm{d}} = \Pr(A_{\mathrm{d}} = 1 \mid X)$, and $J = \Pr(A_{\mathrm{s}} = 1, A_{\mathrm{d}} = 1 \mid X)$. Since switching is the exclusive-or event $A_{\mathrm{s}} \neq A_{\mathrm{d}}$,

$$q_{\mathrm{sw}}(X) = q_{\mathrm{s}} + q_{\mathrm{d}} - 2J. \tag{11}$$

Define the conditional dependence term $C(X) = J - q_{\mathrm{s}} q_{\mathrm{d}}$. Then

$$q_{\mathrm{sw}}(X) = q_{\mathrm{s}}(1 - q_{\mathrm{d}}) + (1 - q_{\mathrm{s}}) q_{\mathrm{d}} - 2C(X). \tag{12}$$

Averaging over X recovers q_{sw} in Eq. (4). Conditional independence gives $C(X) = 0$; with the shared conditional marginal $q_X = \Pr(A = 1 \mid X)$, the expression reduces to $2q_X(1 - q_X)$.

Pooling over unresolved context can inflate that independent-draw reference. If original and donor states are conditionally independent given X and share this marginal, with $\bar{q} = \mathbb{E}[q_X]$, then

$$2\bar{q}(1-\bar{q}) - \mathbb{E}[2q_X(1-q_X)] = 2\,\mathrm{Var}(q_X) \geq 0. \tag{13}$$

Thus a current-binned independent reference can exceed the observed switch rate when finer context remains heterogeneous.

Equation (3) also yields two boundary bounds. Every switch satisfies $|m| \leq H_{\max} = \max(|h_{\mathrm{s}}|, |h_{\mathrm{d}}|)$, so for any $\varepsilon > 0$,

$$q_{\mathrm{sw}} \leq \Pr(|m| \leq \varepsilon) + \Pr(H_{\max} > \varepsilon). \tag{14}$$

Alternatively, let $D = h_{\mathrm{s}} - m$ and $\Delta h = h_{\mathrm{d}} - h_{\mathrm{s}}$. Away from threshold ties, switching is equivalent to

$$D(D + \Delta h) < 0, \tag{15}$$

which requires $|D| \leq |\Delta h|$ and therefore

$$q_{\mathrm{sw}} \leq \Pr(|D| \leq \varepsilon) + \Pr(|\Delta h| > \varepsilon). \tag{16}$$

Equation (14) is centered on the fixed current contribution, whereas Eq. (16) is centered on the original output relative to threshold.

A.3 Binary Residual Preimages at Z=1

In SpikeDF, both residual operands are binary at evaluation. For the additive rule $\phi_+(x, b) = x + b$,

$$\phi_+^{-1}(0) = \{(0,0)\}, \qquad \phi_+^{-1}(1) = \{(1,0),(0,1)\}, \qquad \phi_+^{-1}(2) = \{(1,1)\}. \tag{17}$$

The two disagreement origins therefore present the same scalar current $Z = 1$ to the neuron. The local alphabet exists only between residual addition and the next spike output. The transformed branch then receives a binary spike.

The $Z = 1$ selection range follows from the normalized RD-LIF rule in Eq. (8). Write $V_{\mathrm{r}} = (1+\delta)g$ with $0 < \delta < 1$, equivalently $1/2 < \eta = g/V_{\mathrm{r}} < 1$. At zero retained state, $Z = 1$ is below threshold while $Z = 2$ is above threshold. The $Z = 1$ margin is $\delta/(1+\delta)$. For the Fixed-Stateful parameters $\tau = 2$, $V_{\mathrm{r}} = 1.1$, $g = 1$, and unit transition step, $\lambda = e^{-1/2}$ and $Z = 1$ fires when

$$V_{k-1} \geq \frac{V_{\mathrm{r}} - g}{\lambda} = \frac{0.1}{e^{-1/2}} \approx 0.165. \tag{18}$$

Thus the fixed current contribution places $Z = 1$ near, but below, threshold and leaves retained state able to change the output spike.

Within the range where $Z = 0$ remains off and $Z = 2$ remains on, let $\pi_1 = \Pr(Z = 1)$, $\pi_2 = \Pr(Z = 2)$, and $q = \Pr(h \geq 1 - \eta \mid Z = 1)$. The binary active state density then satisfies the corollary

$$\rho_{\mathrm{Static}} = \pi_2 \;\leq\; \rho_{\mathrm{RD\text{-}LIF}} = \pi_2 + q\pi_1 \;\leq\; \rho_{\mathrm{OR}} = \pi_2 + \pi_1. \tag{19}$$

This inequality applies while the $Z = 0$ and $Z = 2$ endpoint outputs remain unchanged and describes how retained state selects the $Z = 1$ spike between the Static Supra-Unit and Binary-OR endpoints.

A.4 Continuous-Input LIF Dynamics

The same fixed-input crossing condition applies when the merged input is continuous. The continuous FF++ models use an activation-after-addition residual ordering with a standard LIF that applies input decay and hard reset. For merged input M_k,

$$H_k = V_{k-1} + \frac{1}{\tau}(M_k - V_{k-1}) = \left(1 - \frac{1}{\tau}\right) V_{k-1} + \frac{1}{\tau} M_k, \tag{20}$$

$$S_k = \mathbb{I}[H_k \geq 1], \qquad V_k = H_k(1 - S_k). \tag{21}$$

Hence Eq. (2) applies with $c(M) = M/\tau$ and $h(R) = (1 - 1/\tau)R$. The same crossing condition therefore applies when the observed merged input M is continuous.

A.5 DVS128 Gesture PLIF Dynamics

The DVS128 Gesture experiment uses a PLIF with hard reset whose input is not attenuated by the leak update. For merged input M_k and learned time constant τ,

$$H_k = \left(1 - \frac{1}{\tau}\right) V_{k-1} + M_k, \tag{22}$$

$$S_k = \mathbb{I}[H_k \geq V_{\mathrm{r}}], \qquad V_k = H_k(1 - S_k). \tag{23}$$

After normalization by V_{r}, Eq. (2) applies with $c(M) = M/V_{\mathrm{r}}$ and $h(R) = (1 - 1/\tau)R/V_{\mathrm{r}}$. Fixed-input donor replacement therefore obeys the same crossing condition in Eq. (3). Only the neuron parameterization differs from the input-decay LIF in Appendix A.4.

B Fixed-Input State Dependence Across Residual Models

Table B.1 reports best validation video AUC for the independently trained continuous-merge FF++ models and official-test accuracy for DVS128 Gesture after validation-based checkpoint selection.

Donor proximity, propagation depth, merge continuity, and task modality vary across the experiments in this appendix, while post-add input remains fixed within each state-replacement comparison.

Table B.1: Task performance of the cross-setting models.

Setting	Metric	Seed	Value (%)
FF++ continuous	Video AUC	2026	84.86
		2027	87.92
		2028	86.72
		Mean	86.50 ± 1.54
DVS128 Gesture	Accuracy	2026	94.10
		2027	95.49
		2028	96.18
		2029	96.88
		2030	95.49
		Mean	95.63 ± 1.03

B.1 Replay Validation and Donor Proximity

Section 4.2 reports the six-site replay and random-versus-nearest donor summary. At S3.3, the normalized $Z = 1$ current contribution is $c(Z = 1) = 0.89609 \pm 0.00054$, leaving threshold margin $m = 0.10391 \pm 0.00054$.

The SpikeDF replacement cohort contains 128 unique FF++ validation videos and one deterministic first evaluation clip per video. Model seed is the replication unit, while output count gives within-model measurement coverage. Cohort composition and assignment rules are specified in Appendix D.

Nearby matching represents each video by $Z = 1$ frequency and mean output, retained state moments and quantiles, the fraction of positive retained state, normalized current statistics, and distance to threshold. The donor is the nearest different video in this summary space.

B.2 Population Switching and Pairing Dependence

For any sampled output–donor population, q_{sw} counts positions whose binary spike changes, whereas Δ_{mean} records the net signed change. At the deep SpikeDF sites, opposing $0 \to 1$ and $1 \to 0$ changes nearly cancel in the mean while a substantial fraction of individual spikes change.

Table B.2: Independent-marginal reference for random-donor switching.

Site	Empirical (%)	Indep. ref. (%)	Gap (pp)
S1.0	38.24	48.79	10.55
S2.0	27.80	49.59	21.79
S3.0	13.98	49.52	35.54
S3.1	13.61	49.76	36.16
S3.2	15.54	46.52	30.97
S3.3	11.84	35.79	23.95

The population decomposition in Section 3.1, derived in Appendix A.2, predicts that a scalar-current independent-draw reference can overestimate empirical switching when unresolved context remains heterogeneous. Table B.2 compares the estimate $2q(c)(1 - q(c))$ with observed original–donor pairs, where $q(c)$ is the empirical below-margin probability within current bin c; gaps are reported in pp and averaged over three checkpoints.

B.3 Downstream Consequences of Output Switching

Section 4.3 defines the spike-substitution intervention and reports the headline fractional and equal-count results. Thresholded prediction flips provide the corresponding task-level changes. Table B.3 reports the complete propagation statistics; local switch rates and donor robustness are given in Table 3.

Table B.3: Downstream score changes from identified output switches. (a) Fractional substitution, (b) equal-count substitution, and (c) prediction flips under fractional substitution. Values are mean ± sample standard deviation across three checkpoints after within-checkpoint subset averaging.

(a) Controlled-fraction propagation, mean $|\Delta\text{score}|$

Site	1%	10%	50%	All	All $\lvert\Delta p\rvert$ (pp)
S1.0	0.241 ± 0.007	0.376 ± 0.017	0.678 ± 0.021	1.030 ± 0.087	15.58 ± 1.61
S2.0	0.106 ± 0.005	0.169 ± 0.003	0.351 ± 0.032	0.634 ± 0.113	9.12 ± 1.77
S3.0	0.055 ± 0.004	0.080 ± 0.004	0.095 ± 0.008	0.103 ± 0.007	1.37 ± 0.22
S3.1	0.042 ± 0.004	0.068 ± 0.004	0.093 ± 0.007	0.109 ± 0.008	1.44 ± 0.25
S3.2	0.026 ± 0.003	0.053 ± 0.005	0.082 ± 0.004	0.097 ± 0.006	1.36 ± 0.15
S3.3	0.014 ± 0.002	0.039 ± 0.006	0.068 ± 0.002	0.089 ± 0.006	1.21 ± 0.17

(b) Equal-count propagation

Site	$K = 1$	$K = 10$	$K = 100$
S1.0	0.0372 ± 0.0018	0.0974 ± 0.0047	0.1677 ± 0.0037
S2.0	0.0203 ± 0.0008	0.0587 ± 0.0017	0.0952 ± 0.0027
S3.0	0.0076 ± 0.0018	0.0357 ± 0.0059	0.0658 ± 0.0044
S3.1	0.0054 ± 0.0008	0.0262 ± 0.0039	0.0528 ± 0.0064
S3.2	0.0024 ± 0.0005	0.0181 ± 0.0022	0.0424 ± 0.0049
S3.3	0.0028 ± 0.0002	0.0130 ± 0.0015	0.0363 ± 0.0032

(c) Prediction flips under fractional substitution (%)

Site	1%	10%	50%	All
S1.0	4.17 ± 1.04	5.10 ± 0.59	11.09 ± 2.04	19.53 ± 1.56
S2.0	1.35 ± 0.24	2.24 ± 0.24	4.38 ± 1.89	10.16 ± 2.82
S3.0	0.94 ± 0.72	0.89 ± 0.86	1.51 ± 0.59	1.04 ± 0.45
S3.1	0.99 ± 0.18	0.83 ± 0.39	0.99 ± 0.59	1.30 ± 1.19
S3.2	0.89 ± 0.36	0.94 ± 0.00	1.30 ± 0.33	1.30 ± 0.45
S3.3	0.10 ± 0.09	0.73 ± 0.59	1.09 ± 1.09	1.56 ± 1.56

B.4 Threshold-Dependent Switching with Continuous Merges

The activation-after-addition models replace SpikeDF's discrete $Z = 1$ merge with continuous post-add inputs while retaining a stateful neuron. Their LIF update is given in Appendix A.4. Architecture, optimization, and subset details are given in Appendix D.

Table B.4 gives the full stagewise fixed-input summary for continuous merges underlying the representative FF++ rows in Table 5.

Table B.4: Continuous-merge fixed-input switching on FF++. Near and far denote $d = |1 - c| \leq 0.10$ and $d \geq 0.40$. Boundary and switch shares use $d < 0.40$. Values are mean ± sample standard deviation across three models.

Site	All switches (%)	Near switches (%)	Far switches (%)	Boundary share (%)	Switch share (%)
Stage 1	1.76 ± 0.15	36.80 ± 1.64	0.107 ± 0.013	6.72	94.31
Stage 2	2.12 ± 0.27	36.05 ± 0.53	0.186 ± 0.051	7.79	91.92
Stage 3	1.65 ± 0.35	37.83 ± 0.99	0.105 ± 0.020	6.24	94.04

Across the 128-video analysis subset, 99.998–99.999% of transformed branch values are nonbinary and 99.998–99.999% of merged inputs are noninteger. Fixed-input reconstruction reproduces every

output spike across all nine model–site cells. State replacement is then applied directly at the observed continuous merge M.

Let $d = |m| = |1 - c|$ denote unsigned threshold distance, with signed margin $m = 1 - c$ as in the crossing condition. The predefined near and far regions are $d \leq 0.10$ and $d \geq 0.40$. Figure B.1 reports the full profile, in which switch rate decreases as d increases at all three stages.

Donors with the same label or manipulation produce switch rates near the boundary that are close to those from unrestricted donors drawn from different videos. Relative to unrestricted donors, the mean difference is at most 0.65 pp across the three sites. Donor-map repetitions are averaged within model, while training seed remains the replication unit.

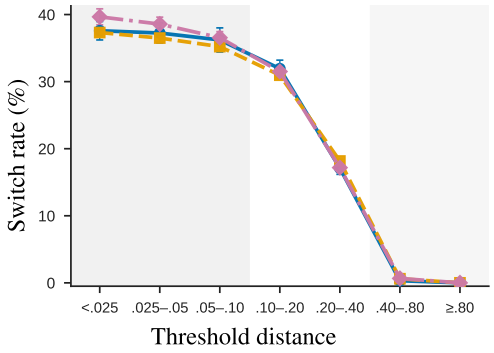


Figure B.1: Switch rate versus threshold distance for the continuous-merge FF++ models.

B.5 Threshold-Dependent Switching on DVS128 Gesture

The DVS128 Gesture experiment changes task, input modality, architecture instance, and post-add neuron instance while retaining stateful post-add computation. Five independently trained eight-block post-add PLIF models reach $95.63 \pm 1.03\%$ accuracy on the official 288-sample test set. Each residual block terminates in a stateful PLIF neuron, giving analysis sites B1–B8. The PLIF update is given in Appendix A.5, and training and donor protocols are specified in Appendix D.

The observed post-add input M is held fixed and only the membrane state before the current step is replaced. Donors come from a different sample of the same gesture class, with three deterministic donor maps averaged within each trained model. B1, B4, and B8 are the prespecified shallow, middle, and deepest representative sites, and the near and far regions match the continuous FF++ analysis.

Table B.5: DVS128 Gesture fixed-input switching. Donors are same-class different-sample states. Near and far denote $d = |1 - c| \leq 0.10$ and $d \geq 0.40$, and boundary and switch shares use $d < 0.40$. Values are mean $\pm$ sample standard deviation across five models after within-model averaging of three donor maps.

Site	All switches (%)	Near switches (%)	Far switches (%)	Boundary share (%)	Switch share (%)	Δ activity (pp)
B1	0.705 ± 0.163	36.07 ± 0.80	0.0818 ± 0.0227	3.58	88.67	-0.132 ± 0.026
B2	0.554 ± 0.117	34.68 ± 0.94	0.0773 ± 0.0236	2.67	86.34	-0.138 ± 0.030
B3	5.887 ± 0.818	37.01 ± 0.44	0.2742 ± 0.0150	20.27	96.22	-0.566 ± 0.073
B4	1.672 ± 0.274	34.62 ± 1.16	0.1099 ± 0.0272	7.36	93.92	$+0.193 \pm 0.151$
B5	3.854 ± 0.236	35.52 ± 1.38	0.3402 ± 0.0724	15.41	92.52	-0.846 ± 0.157
B6	1.854 ± 0.230	33.52 ± 0.66	0.1210 ± 0.0279	8.81	94.10	-0.428 ± 0.129
B7	3.453 ± 0.510	32.93 ± 1.03	0.1826 ± 0.0855	16.68	95.75	-1.101 ± 0.217
B8	2.008 ± 0.292	28.62 ± 0.66	0.0286 ± 0.0158	15.04	98.85	-0.547 ± 0.059

Across all eight sites, switching for outputs near threshold is 28.6–37.0%, whereas switching far from threshold is only 0.029–0.340%. The $d < 0.40$ boundary region contains 2.67–20.27% of outputs but accounts for 86.34–98.85% of all switches. At the deepest site B8, 15.04% of outputs in this region account for 98.85% of switches. Many output switches can again coexist with a much smaller signed activity shift: B3 switches 5.89% of outputs while mean post-add activity changes by only -0.57 pp.

Donors from the same class exhibit the separation between outputs near and far from threshold at every site. Across five models, eight sites, and three donor maps per model, float32 replay produces zero spike mismatches and zero charge reconstruction error. Disagreements with the threshold crossing condition outside the float32 threshold envelope are also zero. All eight sites follow the

same local crossing rule, while their output and state distributions differ. Across B1–B8, boundary occupancy for $d < 0.40$ and switch rate have $\rho_s = 0.976$.

C Control Analyses and Distribution Shift

The analyses in this appendix vary neuron parameter learning, input and auxiliary pathways, retained state perturbations across the network, and the evaluation distribution.

C.1 Recurrence Versus Learned Neuronal Parameters

For controls that expose multivalued residual states, active state density is

$$\rho(X) = \mathbb{E}[\mathbb{I}(X > 0)], \qquad \rho_{\text{late}} = \tfrac{1}{2}\left(\rho_{\text{Stage3}} + \rho_{\text{Fusion}}\right). \tag{24}$$

For binary states, ρ equals the mean spike value, whereas for multivalued additive propagation it records the fraction of nonzero entries.

We use AUC for task discrimination. Table C.1 reports the four preselected FF++ operating points together with a residual-threshold sweep from one seed. The panel averaged across seeds gives the operating point used for analysis. The sweep locates the sharp activity transition relative to the $Z = 1$ current.

Table C.1: FF++ operating points and threshold sensitivity. (a) Four preselected held-out configurations. (b) Single-seed threshold sweep around the singleton current.

(a) Preselected operating points

Model	Blocks	Test AUC	Late act. (%)
Compact, $\theta_e = .075$	4	85.94 ± 0.46	12.06 ± 1.44
SpikeDF-C, $\theta_e = .050$	4	90.13 ± 1.14	10.45 ± 0.99
Sparse, $\theta_e = .075$	6	86.56 ± 0.78	4.88 ± 0.47
SpikeDF-S, $\theta_e = .050$	6	89.78 ± 0.74	4.78 ± 0.15

(b) Threshold sweep

Model	Val. AUC	Late act. (%)
Fixed $V_r = 1.00$	88.60	31.00
Fixed $V_r = 1.05$	88.91	4.46
Fixed $V_r = 1.10$	88.51	5.10
All dyn. fixed	88.89	5.28
No rate reg.	88.40	5.51

Table C.2 adds a multivalued SEW-ADD-style residual rule. Table C.3 reports complementary task metrics for representative residual rules. Pair-macro AUC is the unweighted mean of the four real-versus-manipulation AUCs in FF++, and EER denotes equal error rate.

Table C.2: Additional multivalued residual-rule control on FF++ validation. Δ AUC is paired to learned RD-LIF in pp. Values are mean $\pm$ sample standard deviation over three seeds.

Blocks	Residual rule	Video AUC	Δ AUC (pp)	Late activity (%)
4	SEW-ADD-style	92.44 ± 0.58	$+1.37 \pm 0.60$	58.75 ± 0.86
6	SEW-ADD-style	92.77 ± 0.46	$+2.00 \pm 0.59$	75.18 ± 0.25

Table C.3: Complementary task metrics for representative residual rules on FF++ validation. Pair-macro AUC is the unweighted mean of the four real-versus-manipulation AUCs. Values are mean $\pm$ sample standard deviation over three seeds.

Blocks	Model	Clip AUC	Pair-macro AUC	Video EER
4	SpikeDF	86.22 ± 0.31	88.73 ± 0.43	17.59 ± 0.70
4	Binary-OR	88.37 ± 0.56	91.19 ± 0.21	15.95 ± 1.92
4	SEW-ADD-style	88.73 ± 0.52	91.39 ± 0.19	14.88 ± 1.45
6	SpikeDF	85.67 ± 0.31	88.15 ± 0.41	17.65 ± 2.04
6	Binary-OR	88.35 ± 0.92	91.10 ± 0.75	16.49 ± 1.42
6	SEW-ADD-style	89.11 ± 0.29	91.66 ± 0.20	14.55 ± 1.93

Table C.4 resolves residual activity across Stage 3 and fusion. Fixed-Stateful follows the learned RD-LIF profile, Static approaches zero activity, and the OR or additive relaxations are much denser.

Table C.5: Matched temporal-input controls on FF++ validation. The Artificial Neural Network backbone is fixed while temporal input is hard pseudo-events or continuous pre-threshold residual magnitude. Δ AUC is the paired change from hard to continuous input in pp. Values are mean $\pm$ sample standard deviation over three seeds. Parameter counts are in millions.

Blocks (M)	Temporal input	Video AUC	Δ AUC (pp)	Clip AUC	Pair-macro AUC	Video EER
4 (2.60)	Hard pseudo-events	93.02 ± 0.37	–	88.48 ± 0.60	90.55 ± 0.52	14.58 ± 0.52
4 (2.60)	Continuous residuals	94.27 ± 0.21	$+1.25 \pm 0.46$	90.70 ± 0.23	92.97 ± 0.18	13.30 ± 0.54
6 (3.85)	Hard pseudo-events	92.78 ± 0.56	–	88.58 ± 0.67	90.95 ± 0.45	14.88 ± 0.59
6 (3.85)	Continuous residuals	94.39 ± 0.25	$+1.61 \pm 0.63$	90.78 ± 0.48	92.82 ± 0.41	13.75 ± 1.08

Table C.4: Residual active-state density on FF++ validation. For binary rules, ρ is firing rate; for SEW-ADD-style propagation it is the nonzero fraction. Values are mean $\pm$ sample standard deviation over three seeds.

Blocks	Residual path	Stage 3 (%)	Fusion (%)
4	Static Supra-Unit	0.219 ± 0.043	0.0788 ± 0.0037
4	Fixed-Stateful LIF	11.69 ± 2.09	10.35 ± 2.44
4	RD-LIF	11.35 ± 0.91	9.59 ± 1.03
4	Binary-OR	60.68 ± 0.80	67.82 ± 0.91
4	SEW-ADD-style	54.53 ± 0.83	62.96 ± 0.91
6	Static Supra-Unit	0.128 ± 0.036	0.0495 ± 0.0104
6	Fixed-Stateful LIF	4.73 ± 1.06	5.12 ± 0.52
6	RD-LIF	4.52 ± 0.22	5.08 ± 0.23
6	Binary-OR	83.89 ± 0.53	86.74 ± 0.40
6	SEW-ADD-style	72.55 ± 0.31	77.80 ± 0.21

C.2 Input Representation and Auxiliary Pathway

The detector contains thresholded temporal input and a continuous residual-statistics path at final fusion. Matched variants change these two inputs while preserving the post-add residual backbone.

The matched Artificial Neural Network controls retain the same temporal backbone but differ in whether the input is thresholded or preserves pre-threshold residual magnitude. Table C.5 reports video and clip metrics in one matched comparison.

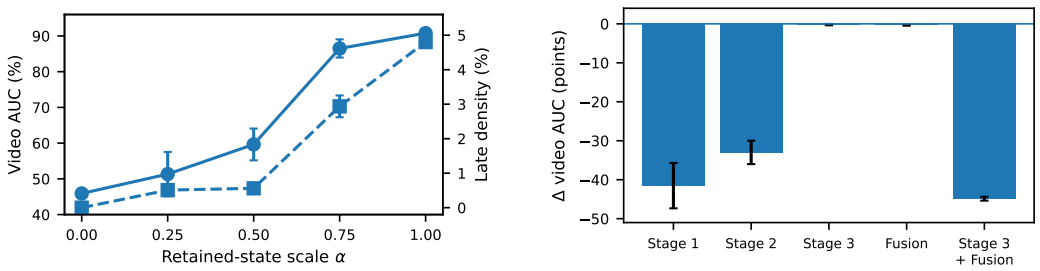
Table C.6: Six-block residual-rule control after removal of the auxiliary residual-statistics pathway. Trainable parameter counts are shown in millions. Values are mean $\pm$ sample standard deviation over three seeds.

Residual rule (M)	Video AUC	Clip AUC	Pair-macro AUC	Video EER	Late activity (%)
Static Supra-Unit (3.830)	75.78 ± 1.50	72.64 ± 0.62	75.63 ± 0.24	31.16 ± 1.70	0.068 ± 0.031
Fixed-Stateful LIF (3.830)	90.84 ± 0.10	85.41 ± 0.16	87.88 ± 0.67	17.86 ± 1.01	5.17 ± 0.32
RD-LIF (3.846)	90.47 ± 0.56	85.32 ± 0.39	87.84 ± 0.39	16.99 ± 1.00	4.95 ± 0.15

Table C.6 reports the same residual rule comparison after removing the auxiliary continuous residual statistics pathway. Removing the final continuous residual statistics projection and retraining the six-block Static, Fixed-Stateful, and learned RD-LIF variants preserves the recurrent–memoryless separation: relative to Static, the paired video AUC changes are 15.06 ± 1.57 pp for Fixed-Stateful and 14.69 ± 1.24 pp for learned RD-LIF, while their direct paired difference is 0.36 ± 0.54 pp.

C.3 Local and Network-Wide Retained State Controls

Interventions on state source, state amplitude, and stage-local state removal are applied recursively and therefore change later inputs and retained states. Their task effects characterize dependence on retained state across the resulting network trajectories.



(a) Retained state amplitude

(b) State removal by stage

Figure C.1: Retained state controls across the network on FF++ validation. (a) Video AUC and late active state density under state scaling α. (b) Paired AUC change under state removal by stage. These controls alter later trajectories recursively; error bars denote sample standard deviation across three models.

Section 4.5 reports representative recursive controls. Table C.7 gives the complete sweep over state source and temporal order, retained-state amplitude, and removal location. Zero history and $\alpha = 0$ define the corresponding zero-state endpoints.

Table C.7: Retained-state controls across the network on FF++ validation. Control families vary state source or order, state amplitude, and state removal by stage. These controls are recursive. Δ AUC is paired to the unmodified checkpoint in pp. Values are mean $\pm$ sample standard deviation across checkpoints.

Control family	Condition	Video AUC	Δ AUC (pp)	Late density (%)
	Observed self-history	90.78 ± 0.16	–	4.80 ± 0.14
	Permuted transitions	90.06 ± 0.61	-0.72 ± 0.46	4.84 ± 0.11
Source / order	Reversed transitions	89.34 ± 0.42	-1.44 ± 0.30	4.94 ± 0.03
	Cross-video history	75.90 ± 0.68	-14.88 ± 0.53	4.48 ± 0.37
	Zero history	45.90 ± 0.59	-44.88 ± 0.50	0.00 ± 0.00
	$\alpha = 0.00$	45.90 ± 0.59	–	0.00 ± 0.00
	$\alpha = 0.25$	51.33 ± 6.21	–	0.51 ± 0.12
Amplitude	$\alpha = 0.50$	59.63 ± 4.46	–	0.56 ± 0.11
	$\alpha = 0.75$	86.50 ± 2.55	–	2.94 ± 0.32
	$\alpha = 1.00$	90.78 ± 0.16	–	4.80 ± 0.14
	Stage 1	49.25 ± 5.87	-41.53 ± 5.82	2.75 ± 0.33
	Stage 2	57.79 ± 2.84	-33.00 ± 3.00	2.54 ± 0.52
Stage removal	Stage 3	90.59 ± 0.11	-0.19 ± 0.18	1.63 ± 0.14
	Fusion	90.51 ± 0.08	-0.27 ± 0.20	3.04 ± 0.11
	Stage 3 + fusion	45.90 ± 0.59	-44.88 ± 0.50	0.00 ± 0.00

The amplitude control multiplies every retained post-add state by α immediately before the next update. Before recursive effects accumulate, the local spike is $S_\alpha = \mathbb{I}[c + \alpha h \geq 1]$ and, for $h > 0$, the current–state pair crosses threshold at $\alpha^\star = m/h$. Because scaling is applied throughout the network, later spikes, inputs, and states also change. The resulting task curve characterizes recursive network dependence across the full trajectory.

The stage control sets retained state to zero within one part of the residual network. Removing Stage 1 or Stage 2 state produces large AUC losses. Removing Stage 3 or fusion alone changes AUC only slightly. Removing both eliminates the late recurrent path and produces a large loss. The non-additive pattern reflects dependence across stages as trajectories change recursively.

C.4 Distribution Shift on Celeb-DF-v2

Celeb-DF-v2 places the frozen FF++ models under a different forgery distribution. Task discrimination and fixed-input measurements use the same frozen checkpoints at matched residual sites.

Task Discrimination Without Adaptation. The three FF++-trained SpikeDF-S models are evaluated on the complete official 518-video test split without fine-tuning and without model selection

based on external labels. Each video is aggregated from four eight-frame clips. We additionally report average precision (AP) and EER. Appendix D specifies class counts and handling of short videos.

Table C.8: Frozen zero-shot evaluation on the official Celeb-DF-v2 test set. FF++-trained SpikeDF-S checkpoints are used without fine-tuning. AUC, AP, and EER values are percentages.

Checkpoint	Video AUC	Video AP	Video EER	Clip AUC
2026	73.79	82.58	31.47	67.83
2027	73.20	81.04	32.04	67.78
2028	75.09	83.08	30.89	68.78
Mean	74.03 ± 0.97	82.24 ± 1.06	31.47 ± 0.58	68.13 ± 0.57

Mean video AUC is $74.03 \pm 0.97\%$ and average precision is $82.24 \pm 1.06\%$ for the frozen FF++ checkpoints. The fixed-input analysis uses the same checkpoints at matched residual sites.

Fixed-Input Switching at Matched Residual Blocks. The fixed-input analysis uses a deterministic 128-video subset at S1.0, S3.1, and S3.3. The same three frozen checkpoints and three state assignments from different videos per checkpoint are used as on FF++. Subset composition is specified in Appendix D.

Table C.9 combines fixed-$Z = 1$ replacement on Celeb-DF-v2, paired changes relative to FF++, and switching conditioned on origin at the three matched sites. At the two deep blocks, state from another video switches roughly 12–14% of $Z = 1$ outputs while self–cross mean output gaps remain at most $0.24\,\mathrm{pp}$. Paired changes in switch rate relative to FF++ are below $0.3\,\mathrm{pp}$. The two $Z = 1$ origins have identical merged current within each site, so the different switch fractions reflect different distributions of state for the two origins.

Table C.9: Fixed-input switching under Celeb-DF-v2 shift. Panels report fixed-$Z = 1$ replacement, retained-state/current statistics, paired changes from FF++, and switching conditioned on origin. Percentage-valued differences are in pp. Values are mean $\pm$ sample standard deviation across three frozen checkpoints unless deterministic.

(a) Fixed-input replacement

Site	Self output (%)	Cross output (%)	Switch (%)
S1.0	56.58 ± 0.85	45.77 ± 2.24	37.41 ± 1.00
S3.1	50.69 ± 1.88	50.57 ± 2.08	13.97 ± 1.40
S3.3	23.43 ± 2.48	23.19 ± 2.46	11.96 ± 0.30

(b) Retained-state and current statistics

Site	$\Pr(h > 0)$ (%)	Current term c	State term h
S1.0	64.00 ± 1.05	0.9146 ± 0.0014	0.2351 ± 0.0076
S3.1	54.32 ± 1.73	0.9060 ± 0.0012	0.2274 ± 0.0092
S3.3	23.58 ± 2.52	0.8961 ± 0.0006	0.1180 ± 0.0117

(c) Paired shift from FF++

Site	Δ switch (pp)	FF++ gap (pp)	Celeb gap (pp)
S1.0	-1.02 ± 0.06	11.41	10.81
S3.1	$+0.28 \pm 0.54$	0.02	0.12
S3.3	$+0.10 \pm 0.19$	0.29	0.24

(d) Origin-conditioned switching

Site	Shortcut (%)	Branch (%)	Gap (pp)
S1.0	37.91 ± 1.04	36.88 ± 1.72	$+1.03$
S3.1	16.06 ± 0.99	12.33 ± 1.81	$+3.73$
S3.3	15.79 ± 1.28	6.65 ± 1.03	$+9.13$

D Reproducibility and Evaluation Protocols

Table D.1 summarizes the data splits, replication units, and analysis subsets used across experiments. The remaining subsections specify optimization and assignment rules for each model and task.

D.1 Evaluation Map and Replication Units

Table D.1: Evaluation map and replication units. Donor maps and subset selections are averaged within model, while the model seed remains the replication unit.

Experiment	Split / subset	Replication unit	Use
SpikeDF operating points	FF++ held-out test	model seed ($n = 3$)	Task performance and late residual activity
Residual rule controls	FF++ validation	model seed ($n = 3$)	Recurrent vs. memoryless post-add transmission
Fixed-$Z = 1$ replacement	FF++ validation; 128 videos	model seed ($n = 3$)	Six-site replacement and downstream score changes
Continuous merge	FF++ validation; 128 videos	model seed ($n = 3$)	Three post-add LIF sites with continuous M
Distribution shift	Celeb-DF-v2 test; 128-video subset	frozen model seed ($n = 3$)	External evaluation and three-site fixed-input analysis
DVS128 Gesture experiment	DVS128 Gesture official test	training seed ($n = 5$)	Eight post-add PLIF sites

The held-out FF++ test evaluates the preselected SpikeDF operating points, while SpikeDF state replacement analyses use validation data. Frozen FF++-trained models are evaluated on Celeb-DF-v2. The continuous FF++ and DVS128 Gesture models are trained independently for their respective fixed-input analyses.

D.2 SpikeDF Architecture, Readout, and Optimization

Each input clip is resized to 128×128 pixels. Adjacent frame cues comprise absolute RGB difference, Laplacian-response difference, Sobel edge-magnitude difference, and a second-order residual difference. Repeated padding transitions for short videos are masked before temporal thresholding and residual statistics aggregation. After normalizing each cue and concatenating them into $r_k \in [0, 1]^{C \times H_{\mathrm{img}} \times W_{\mathrm{img}}}$, a binary pseudo-event tensor $e_k = \mathbb{I}[r_k \geq \theta_{\mathrm{e}}]$ uses $\theta_{\mathrm{e}} = 0.050$ in the final models.

The temporal backbone uses stage widths 80, 160, and 320 and depths $(1, 1, L - 2)$ for L residual blocks. Each block contains a depthwise spike-convolution branch and a two-layer pointwise spike multilayer perceptron. Both residual additions are immediately re-binarized by independent post-add neurons. Let $s_{k,\ell}$ denote the binary input to block ℓ, DP_ℓ denote DropPath, $\mathcal{D}_\ell$ denote the depthwise branch, and $\mathcal{P}_{\ell,1}, \mathcal{P}_{\ell,2}$ denote the pointwise branches. The block computes

$$a^{\mathrm{dw}}_{k,\ell} = s_{k,\ell} + \mathrm{DP}_\ell(\mathcal{D}_\ell(s_{k,\ell})), \tag{25}$$

$$q_{k,\ell} = \mathcal{L}_{V_{\mathrm{r}}}(a^{\mathrm{dw}}_{k,\ell}; V^{\mathrm{dw}}_{k-1,\ell}), \tag{26}$$

$$a^{\mathrm{pw}}_{k,\ell} = q_{k,\ell} + \mathrm{DP}_\ell(\mathcal{P}_{\ell,2}(\mathcal{P}_{\ell,1}(q_{k,\ell}))), \tag{27}$$

$$s_{k,\ell+1} = \mathcal{L}_{V_{\mathrm{r}}}(a^{\mathrm{pw}}_{k,\ell}; V^{\mathrm{pw}}_{k-1,\ell}). \tag{28}$$

At evaluation, DropPath is disabled, so each local sum has the $\{0, 1, 2\}$ alphabet and is immediately re-binarized before the next transformed branch. Stage 2 features are projected to 320 channels, resized by nearest-neighbor interpolation, added to Stage 3 spikes, and re-binarized before pooling.

A spiking linear layer produces a 320-dimensional feature at each physical transition, and learned softmax weights aggregate the seven transition features. The auxiliary branch summarizes pre-threshold residual magnitude by channel-wise mean, standard deviation, and late-minus-early trend before final fusion. The classifier applies layer normalization followed by a multilayer perceptron with a 576-dimensional hidden layer.

The forward RD-LIF threshold remains hard during training. For $x = H - V_{\mathrm{r}}$, backpropagation uses the sigmoid surrogate $\sigma_\alpha(x) = [1 + \exp(-\alpha x)]^{-1}$ with

$$\frac{\partial S}{\partial H} \approx \sigma'_\alpha(x) = \alpha\, \sigma_\alpha(x)\big[1 - \sigma_\alpha(x)\big], \tag{29}$$

following standard surrogate gradient training (Neftci et al., 2019). The learned configuration initializes $\tau = 2.0$, standard neuron threshold at 1.0, residual neuron threshold at 1.10, $\kappa = 1.0$, input gain at 1.0, and surrogate slope at 4.0. Every clip starts from zero membrane.

Training uses 30 epochs with AdamW, effective batch size 240, learning rate 1.125×10^{-3}, weight decay 0.05, three warm-up epochs, gradient clipping at 1.0, and label smoothing 0.05. Matched control families use the same schedule. The four preselected operating points are evaluated on the held-out FF++ test. Residual-state analyses use validation data.

D.3 FF++ Fixed-Input Assignment and Propagation Protocol

The six-site fixed-input, nearby donor, and switched-output propagation experiments use the same 128 unique FF++ validation videos: 64 real and 16 from each manipulation family, with one deterministic first evaluation clip per video. Assignments across videos exclude pairs from the same video. Matching by nearby state is performed within each checkpoint and site and always selects a different video. Repeated donor maps are averaged within model before variation is summarized across the three model seeds.

Fractional and fixed-count downstream analyses use the same output cohort and donor map. Five deterministic subset selections are averaged within checkpoint for the fractional analysis. The downstream analysis substitutes only outputs that changed under fixed-input state replacement and then resumes the frozen network from the recorded recurrent trajectory. Later membrane states are not recursively recomputed.

D.4 Continuous-Merge FF++ Protocol

The continuous FF++ models are trained separately with seeds 2026–2028 using the activation-after-addition SpikingResNet19 ordering, stage widths 80, 160, and 320, block depths $(3, 3, 2)$, and a direct linear binary classifier. Their post-add LIF uses $\tau = 2$, threshold 1, zero reset potential, hard reset, and a rectangular surrogate during training.

Models are evaluated on a fixed 128-video FF++ validation subset balanced between real and manipulated videos at three post-add sites. Donors with the same label or manipulation always come from a different source video, as do unrestricted donors. Fixed-input reconstruction has zero spike mismatches across all nine model–site cells. The maximum normalized charge error is below 8×10^{-7} over approximately 6.17 billion valid post-add updates.

D.5 Celeb-DF-v2 Evaluation Protocol

Celeb-DF-v2 is used without adaptation. Task performance uses the complete official 518-video test split: 108 Celeb-real and 70 YouTube-real videos form the 178-video real class, and 340 Celeb-synthesis videos form the fake class. One official real video has only three available preprocessed face frames and is retained with the repeated-padding transition mask used for short FF++ clips.

The fixed-input analysis uses a deterministic 128-video subset drawn from the 517 videos with at least eight available preprocessed face frames. It contains 39 Celeb-real, 25 YouTube-real, and 64 Celeb-synthesis videos and is evaluated at S1.0, S3.1, and S3.3. The same three frozen SpikeDF-S checkpoints are used for task and mechanism evaluation.

D.6 DVS128 Gesture Protocol

DVS128 Gesture uses seeds 2026–2030, the official 288-sample test, and $T = 40$ split-by-number frames at 128×128. A class-stratified 120-sample validation subset (seed 1701) leaves 1,056 of the 1,176 training samples for optimization. The eight-block model uses base width 64 and post-add PLIF sites B1–B8; the PLIF has hard reset, learnable τ initialized at $4/3$, threshold $V_{\mathrm{r}} = 1$, and a rectangular surrogate (Eq. (23)).

Training uses 300 Adam epochs, cosine learning rate $10^{-3} \to 10^{-6}$, batch size 4 with four-step accumulation (effective 16), and gradient clipping at 5. The best-validation checkpoint is tested; three deterministic same-class donor maps are averaged per seed ($n = 5$). Float32 replay has zero spike/charge error and no crossing-condition disagreement outside the threshold envelope.