GHOST DIRECTIONS AND INTERSECTION LATTICES IN SHALLOW RELU NETWORKS

Anonymous authors
Paper under double-blind review

ABSTRACT

The parameters that fit a given dataset with a shallow rectified linear unit (ReLU) network are far from unique. We study two global questions on finite data: when can two parameter settings with the same outputs be joined by a path that never changes those outputs, and where do several activation labels give algebraic explanations of the same output? We take an output-space view in which each activation cell is represented by its output-image closure, a linear *branch*. We compare the intersection of adjacent branches with the closure of their shared activation wall's image. Directions in the intersection but outside this wall-image closure are *ghost directions*. For a single wall, the dimension gap is at most one and is nonzero only when the branches coincide; a target outside the wall-image closure prevents a direct crossing at fixed output. Assembling these local incidences gives a global necessary condition for exact-fit paths, and refining activation cells by output-weight signs yields a finite graph that computes the exact fiber components by linear feasibility. On the multiplicity side, we study the maximal-branch intersection lattice: its dimension-labeled structure is generically constant at fixed uniform affine order type and width. Retaining all branches with activation-label weights stratifies outputs by their algebraic label multiplicity. For six planar inputs in general position and one hidden neuron, we classify the generic lattices across all 16 order types and determine their exceptional loci. Together the results separate shared outputs from feasible parameter transitions, bear on identifiability and mode connectivity, and supply finite, checkable examples.

1 INTRODUCTION

Two ReLU networks can agree on every training prediction while using different activation patterns, so the parameters that fit a given dataset are far from unique. This nonuniqueness has already been studied from several angles: parameter symmetries and functional equivalence identify important sources of it (Phuong & Lampert, 2020; Grigsby et al., 2023), and functional dimension quantifies the local output degrees of freedom of the realization map (Grigsby et al., 2025; Bona-Pellissier et al., 2024). These tools are adapted to local variation or equivalence as functions. A complementary question is *connectivity* on a finite batch: given two parameter settings that produce the same outputs, can they be joined by a path that never changes those outputs? We study this question for scalar-output, one-hidden-layer ReLU networks on a fixed batch $X = (x_1, \ldots, x_n)$. The evaluation map

$$E_X : \theta \longmapsto (f_\theta(x_1), \ldots, f_\theta(x_n))$$

defines an exact-fit parameter fiber $E_X^{-1}(y)$ for each target y, and we ask when two points of this fiber lie in the same path component.

Convex reformulations of shallow ReLU training provide a natural set of tools. They exploit activation patterns and positive homogeneity (Pilanci & Ergen, 2020) and have enabled detailed analyses of regularized optimal sets and how their connectivity depends on width (Mishkin & Pilanci, 2023; Kim et al., 2025). These analyses describe a regularized optimal set rather than the entire fiber of a prescribed output at fixed width. Sufficiently wide networks also admit connectivity guarantees for loss sublevel sets (Nguyen, 2019). To determine the exact components at an arbitrary width, we must account for two effects at once: transitions between activation cells and paths within a single cell.

We take an output-space viewpoint. Each activation cell is represented by the Zariski closure of the outputs it can produce, and we ask what this representation forgets. A direct transition

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.

between adjacent cells crosses their shared wall, whose image closure can be strictly smaller than the intersection of the two branches. Directions in this intersection but outside the wall-image closure are *ghost directions*. Such a gap can persist even when the two cells share the same output closure, so an unchanged output space can still hide a constraint on how the network passes from one cell to the next. This discrepancy between algebraic overlap and realizable wall geometry is the lens through which we study exact-fit connectivity.

Beyond connectivity, output closures raise a question of multiplicity: where does an output belong to several labeled branch closures? Such patterns have a combinatorial description through oriented matroids and zonotopes (Björner et al., 1999; Matena & Raffel, 2022). When the inputs are in fixed relative position (a fixed affine chirotope), these closures, which we call branches, keep fixed labels, but their intersections can still depend on the exact coordinates. The intersection lattice records all multiple intersections (Orlik & Terao, 2013; Björner & Ekedahl, 1997), including information absent from pairwise intersection dimensions. We study when this structure is stable as the inputs vary. A finer description keeps every branch, including the nonmaximal ones, and attaches activation-label weights to it; this additionally records how many algebraic explanations each output admits. These two levels of description distinguish the geometry of shared outputs from the multiplicity of the labels associated with them.

Our contributions can be summarized as follows.

1. The single-wall ghost dimension is at most one and is nonzero exactly when the two adjacent branches coincide strictly beyond the wall-image closure, so every ghost requires interaction between neurons (Theorem 3). At the global level, the wall-closure graph gives a necessary condition for exact-fit paths (Theorem 4), and a signed closed-cell graph computes the exact fiber components at fixed width by linear feasibility (Theorem 6).
2. For a fixed uniform affine chirotope and width, the maximal-branch intersection lattice has a single generic dimension-labeled type across all inputs of that chirotope, constant outside a proper algebraic discriminant (Theorem 8). Retaining every branch with its activation-label weights refines this lattice so that accumulating these weights counts the algebraic explanations of each output and gives every multiplicity superlevel set a unique irredundant representation (Theorem 9).
3. For $(n, d, m) = (6, 2, 1)$, we classify the 16 pairwise distinct generic lattices across all uniform planar order types and determine the exceptional loci where their structure changes (Theorems 13 and 14).

Implications. The ghost obstruction shows that matching output closures need not permit a direct transition across a shared wall. The graph constructions turn these local restrictions into statements about exact-fit connectivity, while the multiplicity stratification quantifies algebraic label redundancy at fixed width. These results bear on identifiability and mode connectivity for exact interpolators. Broader implications and future directions are discussed in Appendices F and G.2.

2 Background

2.1 Finite-data output arrangement

Let $X = (x_1, \dots, x_n) \in (\mathbb{R}^d)^n$, with $n, d \geq 1$, be a finite batch. Inputs are distinct unless stated otherwise. Write $\widetilde{x}_i = (x_i, 1)$ and $u_j = (w_j, b_j)$ to absorb the hidden bias, and stack the lifted rows $\widetilde{x}_i^\top$ into $\widetilde{X} \in \mathbb{R}^{n \times (d+1)}$. We study the scalar width-m network, with $m \geq 1$,

$$f_\theta(x) = \beta_0 + \sum_{j=1}^{m} \alpha_j \,\mathrm{ReLU}(u_j^\top \widetilde{x}), \qquad u_j \in \mathbb{R}^{d+1}. \tag{1}$$

The full parameter vector is $\theta = (\beta_0, (\alpha_j, u_j)_{j=1}^m) \in \mathbb{R}^{1+m(d+2)}$. A strict activation pattern $s = (s_i)_{i=1}^n \in \{0,1\}^n$ is realizable when its cone $C_s = \{u : (2s_i - 1)\widetilde{x}_i^\top u > 0 \text{ for every } i\}$ is nonempty; $s_i = 1$ indicates an active sample. Write $\mathcal{A}(X)$ for the realizable patterns, $D_s = \mathrm{diag}(s)$, and $L_s = \mathrm{im}(D_s \widetilde{X})$. Within this region, $\mathrm{ReLU}(\widetilde{X}u) = D_s \widetilde{X} u$.

For neuron patterns $s^{(j)} \in \mathcal{A}(X)$, write $a = (s^{(1)}, \dots, s^{(m)})$. Its parameter cell is $C_a = \{\theta : u_j \in C_{s^{(j)}} \text{ for every } j\}$, with output weights and bias unrestricted. Define

$$B_a(X) = \left[\mathbf{1}_n \; D_{s^{(1)}}\widetilde{X} \; \cdots \; D_{s^{(m)}}\widetilde{X}\right], \qquad V_a(X) = \operatorname{im} B_a(X). \tag{2}$$

Here $\mathbf{1}_n$ accounts for the output bias β_0, which shifts all sample outputs equally. Patterns may repeat because several neurons can be active on the same samples. The subspace V_a is unchanged by reordering a; we use multiset labels when neuron order is irrelevant.

Proposition 1 (Zariski closure of activation-cell images)**.** *If every pattern in a is strictly realizable, the Zariski closure of the corresponding parameter-cell image is $V_a(X)$. Hence the closure of the full evaluation image is the finite arrangement*

$$\mathcal{V}_X = \bigcup_{a \in \mathcal{A}(X)^m} V_a(X), \qquad \dim V_a(X) = \operatorname{rank} B_a(X). \tag{3}$$

The Zariski closure retains polynomial relations but discards activation inequalities, so V_a may contain outputs unrealizable in the cell. This gap underlies the wall analysis in Section 3. For univariate shallow networks, Bona-Pellissier et al. (2024, Proposition 15) give the activation-region image and cellwise rank matrix; we use the arbitrary-dimensional closure formulation as our baseline and prove it in Appendix A.1. Under affine general position, the largest branch attains the dimension bound $\min\{n, m(d+1)+1\}$ at every width, giving the finite-batch output dimension (Appendix A.2).

For a simple illustration, take $d = m = 1$, $x_1 < \cdots < x_n$, and $x = (x_1, \dots, x_n)^\top$. Each activation mask s selects a prefix or suffix (possibly empty or full), giving the branch

$$V_s = \operatorname{span}\{\mathbf{1}_n, s, s \odot x\}.$$

The input order fixes the available masks, while the coordinates determine the branch subspaces.

2.2 Related work

Output dimension and parameter identifiability. Functional and batch functional dimension quantify the local degrees of freedom of the realization map, so that output-preserving parameter changes contribute none (Grigsby et al., 2025; Bona-Pellissier et al., 2024). At the parameter level, identifiability then asks whether this redundancy is exhausted by known symmetries (Bona-Pellissier et al., 2022; Stock & Gribonval, 2023; Phuong & Lampert, 2020; Grillo & Montúfar, 2026), a question addressed through symmetry classifications and the geometry of realization fibers (Ramakrishnan, 2026; Flinth et al., 2026). The latter compare parameters representing the same function on the entire input domain; finite-batch agreement permits additional coincidences.

Solution sets and connectivity. Connectivity asks whether equivalent or optimal parameters can be joined by a continuous path. Empirical mode-connectivity studies construct paths with nearly constant loss (Garipov et al., 2018). For regularized shallow ReLU networks, positive homogeneity and activation-cone lifts give convex reformulations (Pilanci & Ergen, 2020) that make optimal sets accessible through polyhedral geometry and duality (Mishkin & Pilanci, 2023; Kim et al., 2025), revealing how connectivity varies with width and regularization. For unregularized networks, sufficient width conditions also ensure connected loss sublevel sets (Nguyen, 2019). These results establish connectedness in specified regimes; they do not enumerate the components of an arbitrary fixed-width exact-fit fiber.

Activation geometry and intersection lattices. Activation patterns describe the network's piecewise-affine geometry; region counts and tropical methods bound the induced input partitions and decision boundaries (Montúfar et al., 2014; Serra et al., 2018; Hanin & Rolnick, 2019; Zhang et al., 2018; Brandenburg et al., 2024). Restricting the patterns to a finite batch yields a description by oriented matroids and zonotopes (Cover, 1965; Björner et al., 1999; Matena & Raffel, 2022), and for the associated linear cell closures arrangement theory supplies an intersection poset, or intersection lattice, recording all multiple intersections together with their dimensions (Orlik & Terao, 2013).

Relation to prior work. Unlike full-function symmetry classifications and width-based connectivity guarantees, we determine the exact components of a prescribed finite-batch fiber at fixed width, including disconnected cases, through a finite graph of linear feasibility tests. On the output side, we study how intersection lattices and algebraic label multiplicity depend on the input coordinates.

3 GHOST DIRECTIONS AND EXACT FIBER CONNECTIVITY

Two adjacent activation cells can have the same output closure, yet their shared boundary in parameter space may realize only part of the outputs the two closures have in common. Any output-preserving parameter path that crosses their common facet must pass through this shared boundary, so it is constrained by what the boundary realizes, not merely by the algebraic intersection of the two branches. We first compute the image closure of one shared boundary, then assemble these local incidences into a global obstruction. A signed closed-cell cover subsequently gives the exact fiber components through linear feasibility tests at fixed network width.

3.1 WALL-IMAGE CLOSURES AND GHOST DIRECTIONS

Consider two ordered activation cells C_a and C_b whose closures share a facet, and let $H_{a,b}$ be its relative interior. We call $H_{a,b}$ a *true wall* when crossing it flips exactly one activation: a single neuron j changes its state at a single sample i, while every other preactivation stays strictly signed. Equivalently, on $H_{a,b}$ the crossing vector satisfies $\widetilde{x}_i^\top u_j = 0$ and all remaining activation inequalities are strict, so $H_{a,b}$ has codimension one and the two labels differ by a single Hamming-one flip.

Write $a = (s^{(1)}, \dots, s^{(m)})$ and $b = (t^{(1)}, \dots, t^{(m)})$ for the pattern tuples of the two cells, and V_a, V_b for their output closures. The other neurons and the output bias contribute the same subspace U on both sides; the crossing neuron contributes $L_{a,b}$ on the wall:

$$U = \operatorname{span}\{\mathbf{1}_n\} + \sum_{k \neq j} L_{s^{(k)}}, \qquad L_{a,b} = D_{s^{(j)}} \widetilde{X}\big(\ker \widetilde{x}_i^\top\big),$$
$$V_a = U + L_{s^{(j)}}, \qquad V_b = U + L_{t^{(j)}}.$$

Here $L_{s^{(k)}} = \operatorname{im}(D_{s^{(k)}} \widetilde{X})$ as in Section 2, so U collects the bias and the noncrossing neurons; $L_{a,b}$ instead is the image of the crossing neuron restricted to the wall kernel $\ker \widetilde{x}_i^\top$. The definition of $L_{a,b}$ is independent of the chosen side because the masks differ only at the sample whose preactivation vanishes on the wall. This gives the following wall-image formula.

Proposition 2 (Wall image versus algebraic intersection)**.** *The shared-facet image has Zariski closure*

$$G_{a,b} := \overline{E_X(H_{a,b})}^{\mathrm{Zar}} = U + L_{a,b} \subseteq V_a \cap V_b.$$

The inclusion can be strict.

Proposition 2 shows that a common algebraic output need not be accessible through the shared wall. We call a vector in $(V_a \cap V_b) \setminus G_{a,b}$ a *ghost direction*, and measure this discrepancy by

$$\begin{aligned} \gamma &:= \dim(V_a \cap V_b) - \dim G_{a,b} \\ &= \dim V_a + \dim V_b - \dim(V_a + V_b) - \dim G_{a,b}. \end{aligned} \tag{4}$$

Since $G_{a,b} \subseteq V_a \cap V_b$ by Proposition 2, we have $\gamma \geq 0$, with $\gamma = 0$ exactly when the wall-image closure equals the algebraic overlap. Actual wall realizability still depends on the activation inequalities. For a true wall, this discrepancy admits a complete classification.

Theorem 3 (Complete single-wall ghost classification)**.** *For a true single wall, $\gamma \in \{0, 1\}$ and*

$$\gamma = 1 \quad \Longleftrightarrow \quad V_a = V_b \supsetneq G_{a,b}. \tag{5}$$

At width $m = 1$, every true wall has $V_a \cap V_b = G_{a,b}$; every ghost is therefore interaction-induced by the other neurons through U.

Theorem 3 shows that distinct adjacent branches have no ghosts: their intersection equals the wall-image closure. A ghost occurs when the branches coincide but the wall loses one output direction. Thus an unchanged output subspace can hide a restriction on crossing the wall, arising from interaction between neurons. Both proofs are in Appendix B.1.

A five-point ghost. Take $X = (0, 1, 2, 3, 4)$ and $m = 2$. The first neuron crosses its activation wall at $u_1 = (-1, 3)$, changing its pattern from 11100 to 11110, while the second keeps pattern 00111. For the adjacent labels $a = (11100, 00111)$ and $b = (11110, 00111)$, direct calculation gives

$$V_a = V_b = \mathbb{R}^5, \qquad G_{a,b} = \{y : -y_1 + y_2 + y_3 - 2y_4 + y_5 = 0\}.$$

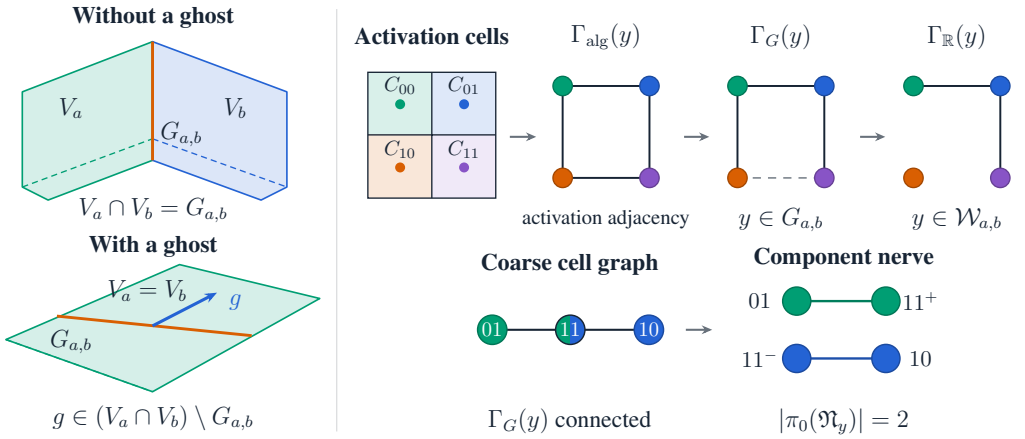

Figure 1: **From wall geometry to fiber connectivity. Left:** schematic output subspaces; the orange line is the wall-image closure $G_{a,b}$. Distinct branches meet on this line; coincident branches can also share a ghost direction outside it. **Upper right:** schematic activation cells generate an adjacency graph, then wall-closure and real-image tests remove edges (dashes mark excluded edges); real feasibility can also remove vertices. **Lower right:** in the two-sample example of Section 3.2, the coarse graph is connected, but splitting cell 11 by its two fiber components gives a disconnected component nerve. Superscripts denote output-weight signs.

Thus $\gamma = 1$. The target $y = (5, 3, 2, 3, 5)^\top$ is realized in both strict cells but lies outside $G_{a,b}$, so no path preserving this output can cross their shared wall. The parameter witnesses and subspace calculation are given in Appendix B.1.

3.2 From wall obstructions to exact fiber connectivity

A blocked wall excludes one crossing, but an exact-fit path could route through other cells. We therefore assemble all candidate transitions at a fixed output y into a graph. To distinguish algebraic membership from realizability, write the cell and wall images

$$\mathcal{Y}_a = E_X(C_a) \subseteq V_a, \qquad \mathcal{W}_{a,b} = E_X(H_{a,b}),$$

so $G_{a,b} = \overline{\mathcal{W}_{a,b}}^{\mathrm{Zar}}$. Three nested graphs record these transition tests:

- $\Gamma_{\mathrm{alg}}(y)$ has as vertices the labels a with $y \in V_a$, and as edges the Hamming-one activation adjacencies between such vertices;
- $\Gamma_G(y)$ keeps the same vertices but only the edges with $y \in G_{a,b}$;
- $\Gamma_{\mathbb{R}}(y)$ further requires $y \in \mathcal{Y}_a$ at vertices and $y \in \mathcal{W}_{a,b}$ at edges.

Thus $\Gamma_{\mathbb{R}}(y) \subseteq \Gamma_G(y) \subseteq \Gamma_{\mathrm{alg}}(y)$. Membership in $\Gamma_G(y)$ is decided by rank tests, exact over $\mathbb{Q}$ for rational X, y. Its disconnected components obstruct exact-fit paths between strict cells, even when a path follows boundaries or crosses several walls simultaneously.

Theorem 4 (Global wall-closure obstruction)**.** *For distinct inputs, let $\theta_0, \theta_1 \in \mathcal{F}_y := E_X^{-1}(y)$ lie in strict ordered cells C_{a_0}, C_{a_1}. If θ_0 and θ_1 belong to the same path component of $\mathcal{F}_y$, then a_0 and a_1 belong to the same connected component of $\Gamma_G(y)$. No transversality, genericity, or single-wall-crossing assumption is required.*

The obstruction also applies after quotienting ordered cells by neuron permutations. The proof is in Appendix B.2.

For the five-point ghost example, $\Gamma_{\mathrm{alg}}(y)$ has 12 vertices and four edges, connecting a and b. The target is a ghost direction at every edge, so $\Gamma_G(y)$ has no edges. The theorem rules out every exact-fit path between the witnesses, including routes through other cells (Appendix B.3).

Recovering exact connectivity. The graph gives a necessary condition, but constructing a parameter path also requires joining successive wall crossings within the same cell fiber. A cell fiber may contain several path components, so it cannot always be treated as one vertex. We therefore use the components of each closed-cell fiber:

$$\mathcal{F}_{a,y} := \mathcal{F}_y \cap \overline{C_a}^{\mathrm{Eucl}} = \bigsqcup_{\lambda \in \Pi_a(y)} \mathcal{F}^{\lambda}_{a,y},$$

where $\Pi_a(y)$ indexes the finitely many semialgebraic path components $\mathcal{F}^{\lambda}_{a,y}$; finiteness holds because $\mathcal{F}_{a,y}$ is semialgebraic (Bochnak et al., 2013). Define the *component-refined exact-fit nerve graph* $\mathfrak{N}_y$ to have these components as vertices, with an edge between two components exactly when the corresponding closed sets $\mathcal{F}^{\lambda}_{a,y}$ intersect; such an intersection may lie on a face of arbitrary codimension. Constructing this nerve requires the true within-cell path components; Section 3.3 replaces that task with linear feasibility using output-weight signs. The proposition below is a structural identity for arbitrary finite n, d, m, with π_0 denoting the set of path components.

Proposition 5 (Exact global component nerve)**.** *The inclusions $\mathcal{F}^{\lambda}_{a,y} \hookrightarrow \mathcal{F}_y$ induce a bijection*

$$\pi_0(\mathcal{F}_y) \cong \pi_0(\mathfrak{N}_y).$$

The union of the $\mathcal{F}^{\lambda}_{a,y}$ lying in one connected component of $\mathfrak{N}_y$ is exactly one path component of $\mathcal{F}_y$.

Appendix B.2 proves this finite closed-cover result. The following example shows why one vertex per activation label is insufficient: the coarse graph $\Gamma_G(y)$ can be connected while the fiber is not.

Example 1 (A connected coarse graph with a disconnected fiber)**.** Take $X = (0, 1)$, $y = (0, 1)$, and $m = 1$. The coarse graph $\Gamma_G(y)$ is connected, yet $\mathcal{F}_y$ has exactly two path components. Indeed both $f_+(x) = \mathrm{ReLU}(x)$ and $f_-(x) = 1 - \mathrm{ReLU}(1 - x)$ interpolate y but lie in different components, because the output weight α cannot change sign along an exact-fit path. The component-refined nerve $\mathfrak{N}_y$ separates the two.

Appendix B.2 gives the parameterizations and extends the example to any two distinct inputs.

3.3 A COMPUTABLE GRAPH FOR EXACT FIBER COMPONENTS

The nerve $\mathfrak{N}_y$ of Section 3.2 counts fiber components exactly, but constructing it would require decomposing each closed-cell fiber. We instead use a finer cover with path-connected pieces described by linear feasibility. Fix a *weak sign* $\varepsilon_j \in \{-1, 1\}$ for each output weight α_j, imposing $\varepsilon_j \alpha_j \geq 0$; the inequality is weak, so $\alpha_j = 0$ is allowed and lies in both sign choices. Once the signs and the activation pattern of each neuron are fixed, positive homogeneity turns the exact-fit condition into linear constraints in normalized variables. The resulting feasible set is a convex polyhedron; path connectivity in the original parameters, including zero weights, follows from explicit paths. For $s \in \mathcal{A}(X)$, write

$$K_s = \{u : (2s_i - 1)\widetilde{x}_i^{\top} u \geq 0 \text{ for every } i\}.$$

Refine an ordered activation label $a = (s^{(1)}, \ldots, s^{(m)})$ by $\varepsilon \in \{-1, 1\}^m$, and denote the signed label by $\sigma = (a, \varepsilon)$. Write $v = (v_1, \ldots, v_m)$ for the absorbed vectors $v_j = |\alpha_j| u_j$, and define

$$\mathcal{F}_{\sigma,y} = \mathcal{F}_y \cap \{\varepsilon_j \alpha_j \geq 0,\ u_j \in K_{s^{(j)}} \text{ for every } j\}, \tag{6}$$

$$\mathcal{P}_\sigma(y) = \left\{ (\beta_0, v) : v_j \in K_{s^{(j)}}, \quad \beta_0 \mathbf{1}_n + \sum_{j=1}^{m} \varepsilon_j D_{s^{(j)}} \widetilde{X} v_j = y \right\}. \tag{7}$$

This substitution maps $\mathcal{F}_{\sigma,y}$ onto the convex polyhedron $\mathcal{P}_\sigma(y)$. When $\alpha_j = 0$ the substitution sends the neuron to $v_j = 0$. The construction below retains these zero-weight cases when connecting pieces in the original parameter space.

Let $\Gamma_\pm(y)$ be the graph whose vertices are the signed labels σ with $\mathcal{F}_{\sigma,y} \neq \varnothing$, and whose edges join two vertices whenever their pieces intersect. These edges include connections through zero output weights and through simultaneous activation changes; they are not restricted to Hamming-one adjacency between activation cells and also account for intersections of higher codimension.

Theorem 6 (Exact fiber components via linear feasibility). *For every finite real batch X, including repeated or dependent inputs, target y, and width $m \geq 1$, $\mathcal{F}_{\sigma,y}$ is nonempty exactly when $\mathcal{P}_\sigma(y)$ is feasible, and every nonempty $\mathcal{F}_{\sigma,y}$ is path connected. Pairwise intersections are also decided by linear feasibility, and*

$$\pi_0(\mathcal{F}_y) \cong \pi_0(\Gamma_\pm(y)). \tag{8}$$

This counts components in the original ordered parameter space, without general-position or width assumptions beyond $m \geq 1$. For rational X, y, the graph is computable by exact rational linear feasibility; an empty graph certifies an empty fiber.

The explicit edge tests, zero-weight path construction, and complexity bounds are in Appendix B.3. With $r = |\mathcal{A}(X)|$ activation patterns, $\Gamma_\pm(y)$ has at most $(2r)^m$ vertices, so it computes the components at fixed width but can grow exponentially. In the five-point ghost example, $\Gamma_G(y)$ already separates the witnesses by rank tests; $\Gamma_\pm(y)$ has eight isolated vertices and gives the exact component count by feasibility tests.

Corollary 7 (Width transition for affinely independent batches). *If $\operatorname{rank} \widetilde{X} = n$ (hence $n \leq d+1$) and y is nonconstant, then*

$$|\pi_0(\mathcal{F}_y)| = \begin{cases} 2, & m = 1, \\ 1, & m \geq 2. \end{cases}$$

For any finite X, a constant-target fiber is contractible at every width $m \geq 1$.

Affine independence is essential: the one-dimensional batch $X = (0, 1, 2)$ with $y = (0, 1, 0)$ has six components at width two. Appendix B.3 gives the proof of the corollary and the counterexample, together with the exact graph counts.

4 Intersection lattices and label multiplicity

Section 3 compared branch intersections with shared-wall image closures to study parameter paths. Here we study the intersections themselves: each algebraic branch V_a of Section 2 represents the output closure of an activation label, and pairwise intersection dimensions do not determine where three or more branches meet. We first ask when the maximal-branch intersection structure is generically constant as the inputs vary, then retain all distinct branches and their label weights to count algebraic explanations of each output.

4.1 The maximal-branch lattice

Assume $n \geq d+1$ and X is in affine general position. Its affine chirotope (Björner et al., 1999), which records the orientation of every $(d+1)$-subset of lifted inputs, is

$$\chi_X(I) = \operatorname{sign} \det \widetilde{X}_I, \qquad |I| = d+1.$$

Here $\widetilde{X}_I$ lists the rows indexed by I in increasing order. For a realizable uniform chirotope χ, let $\mathcal{R}_\chi$ be the open semialgebraic set on which all these determinant signs equal χ. We call $\mathcal{R}_\chi$ a realization cell; it need not be connected. Oriented-matroid theory gives $\mathcal{A}(X) = \mathcal{A}_\chi$ for every $X \in \mathcal{R}_\chi$: input batches with the same chirotope realize the same strict activation patterns (Björner et al., 1999). Their branch labels therefore stay fixed, while the subspaces $V_a(X)$ can vary with the coordinates.

Let $a \in \mathcal{A}(X)^m$ be an ordered width-m branch label, with branch subspace $V_a = \operatorname{im} B_a(X)$ from Section 2. Two branches may be equal or one may be contained in another. Discard duplicates and branches strictly contained in another, and write $\mathcal{M}_X$ for the remaining maximal subspaces. The full, dimension-labeled intersection lattice, in the sense of subspace-arrangement theory (Björner & Ekedahl, 1997), is

$$\Lambda_X = \{\mathbb{R}^n\} \cup \left\{ \bigcap_{V \in J} V : \varnothing \neq J \subseteq \mathcal{M}_X \right\}. \tag{9}$$

The lattice includes every intersection of maximal branches, with $\mathbb{R}^n$ as the empty intersection. Equal subspaces occur once, ordered by inclusion and labeled by dimension. The join of two flats is the intersection of all lattice flats containing both. We obtain the following generic constancy result.

Theorem 8 (Generic lattice constancy). *Fix n, d, m and a realizable uniform affine chirotope χ. There is a nonzero real polynomial $\Delta_{\chi,m}$ such that for all $X, Y \in \mathcal{R}_\chi \setminus Z(\Delta_{\chi,m})$:*

1. *Λ_X and Λ_Y are isomorphic as dimension-labeled posets;*
2. *dimensions, equality, inclusion, and maximality of all labeled branches, including nonmaximal ones, are preserved; the isomorphism in (1) matches maximal branches with the same labels; and*
3. *the graph obtained from Hamming-one activation adjacency by retaining labels of maximal branches and identifying labels with equal branch subspaces is the same for X and Y.*

The complement of $Z(\Delta_{\chi,m})$ is Euclidean open dense in $\mathcal{R}_\chi$, and the exceptional set has Lebesgue measure zero.

In words, once the chirotope is fixed, the way the branches sit inside one another is the same for almost every input batch: nudging a generic X moves the output subspaces $V_a(X)$ but keeps their intersection dimensions and containments intact, so the labeled lattice can only change on the measure-zero locus $Z(\Delta_{\chi,m})$. The proof is in Appendix C.1.

4.2 The weighted all-branch lattice

The maximal lattice Λ_X records branch incidence but discards nonmaximal branches and the number of activation labels sharing a subspace. For any finite batch X, index branches up to neuron permutation by the set $\mathcal{A}_X^{(m)}$ of width-m multisets of patterns from $\mathcal{A}(X)$. Write $\mathsf{A}_X = \{V_a : a \in \mathcal{A}_X^{(m)}\}$ for the distinct branches and assign each $V \in \mathsf{A}_X$ the weight

$$w(V) = \#\{a \in \mathcal{A}_X^{(m)} : V_a = V\}.$$

Closing the distinct branches A_X under intersections, including the empty intersection $\mathbb{R}^n$, gives Λ_X^{lab}. We order its flats by inclusion and retain the branch weights w; Appendix C.3 gives the formal poset description.

Label multiplicity. For $y \in \mathbb{R}^n$, define its algebraic activation-explanation set and multiplicity by

$$\mathcal{B}_{\mathrm{lab}}(y) = \{a \in \mathcal{A}_X^{(m)} : y \in V_a\}, \qquad b_{\mathrm{lab}}(y) = \#\mathcal{B}_{\mathrm{lab}}(y). \tag{10}$$

Thus $b_{\mathrm{lab}}(y)$ counts multiset labels whose branch closures contain y, rather than parameter settings realizing y. Recovering these counts requires both nonmaximal branches and label weights.

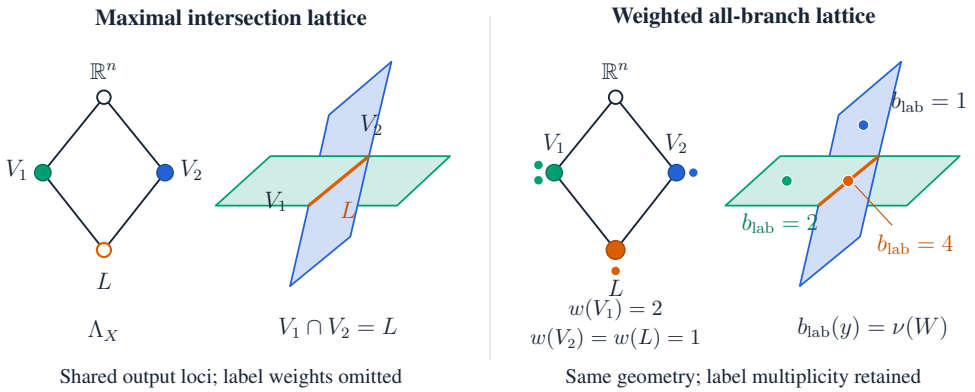

Figure 2: **Intersection geometry and label multiplicity.** An illustrative two-plane arrangement connects Sections 4.1 and 4.2. **Left:** the maximal lattice includes $L = V_1 \cap V_2$ as an intersection flat (hollow node). **Right:** L is also retained as a branch (filled node), and small markers indicate label weights $w(V_1) = 2$, $w(V_2) = w(L) = 1$. The geometry is the same, while the weights give multiplicities 2 on $V_1 \setminus L$, 1 on $V_2 \setminus L$, and 4 on L. Thus weights and branch designations add information even when retaining nonmaximal branches adds no new intersection flats.

For $W \in \Lambda_X^{\mathrm{lab}}$, let

$$\nu(W) = \sum_{\substack{V \in \mathsf{A}_X \\ V \supseteq W}} w(V), \qquad W^\circ = W \setminus \bigcup_{U \in \Lambda_X^{\mathrm{lab}},\, U \subsetneq W} U. \tag{11}$$

Here W° is the part of W outside all strictly smaller flats. These nonempty strata partition $\mathbb{R}^n$. For $y \in W^\circ$, the branches containing y are exactly those containing W; their weights sum to $\nu(W)$. The following theorem gives the resulting multiplicity stratification; its proof is in Appendix C.3.

Theorem 9 (Multiplicity stratification)**.** *On each stratum W°, the multiplicity is constant: $b_{\mathrm{lab}}(y) = \nu(W)$ for all $y \in W^\circ$. For an integer $k \geq 1$, set $\Omega_{\geq k} = \{y : b_{\mathrm{lab}}(y) \geq k\}$ and let M_k be the inclusion-maximal flats with $\nu(W) \geq k$. Then*

$$\Omega_{\geq k} = \bigcup_{\substack{W \in \Lambda_X^{\mathrm{lab}} \\ \nu(W) \geq k}} W = \bigcup_{W \in \mathsf{M}_k} W. \tag{12}$$

The last union is the unique irredundant representation by flats of Λ_X^{lab}.

Each threshold is therefore determined by its maximal qualifying flats. These counts concern branch closures; strict realizability also requires the activation inequalities. For the five-point ghost target, $b_{\mathrm{lab}}(y) = 6$, but only four multiset labels are strictly realizable. The remaining two labels have unique absorbed solutions incompatible with their activation masks (Appendix B.3).

4.3 Intersection structure for six planar inputs

For six planar points in affine general position and one hidden neuron, there are 32 activation patterns. Theorem 13 in Appendix D.1 shows that the 16 uniform order types yield 16 pairwise nonisomorphic generic maximal-branch lattices. Theorem 14 in Appendix D.3 identifies the exceptional loci within each realization cell. Within an order type, additional coordinate relations determine departures from its generic lattice. The examples below distinguish changes in higher-order intersections from the activation explanations contributed by nonmaximal branches.

For the maximal-branch lattice, the path in Appendix D.5 stays within one realization cell and exhibits a change in a three-branch intersection dimension while all branch dimensions and labeled pairwise intersection dimensions remain fixed at the transition. The triple-intersection dimension increases from one to two. Pairwise dimensions therefore do not determine the full lattice. Genericity here is relative to the path, which lies in the ambient exceptional locus.

For the weighted all-branch lattice of Section 4.2, all 32 labeled branches contain the constant-output line, so by Theorem 9 every output on this line has multiplicity 32. At generic configurations, these branches are distinct and all weights equal 1; keeping only maximal branches lowers the count to 18–21, depending on the order type. Retaining nonmaximal branches therefore recovers activation explanations even on this common output line. Across the 16 generic lattices, the intermediate superlevel sets vary, and retaining nonmaximal branches adds strata in 11 cases.

5 Conclusion

We studied how activation patterns organize the outputs and exact-fit parameter fibers of shallow ReLU networks on finite data. Our central finding is that algebraic branch intersections can contain ghost directions absent even from shared-wall image closures, revealing a gap between output overlap and parameter connectivity. The wall graph gives a necessary condition for exact-fit paths, and the signed closed-cell graph computes the exact fiber components at fixed width by linear feasibility. On the output side, we established generic constancy of the intersection lattice and showed how branch weights determine algebraic label multiplicity. In the six-point planar case $(n, d, m) = (6, 2, 1)$, we classified the generic lattices of all 16 uniform order types and determined where their incidence structures change. These results provide a geometric basis for analyzing identifiability and mode connectivity in interpolating networks at fixed width.

Reproducibility Statement

Appendix G describes the computational methods, arithmetic conventions, software environment, and coverage of the reported calculations. These details explain how geometric objects and feasibility certificates are computed and checked, supporting independent reproduction of the results.

Ethics Statement

This work is theoretical and computational. It uses only synthetic point configurations and mathematical software, and involves no human subjects, personal data, scraped content, or deployed systems. Broader implications and future directions are discussed in Appendices F and G.2.

AI Use Statement

Generative AI assisted with manuscript editing, code development, organization of experiment records, image generation, consistency checks, and assistance with proof revision. The authors are responsible for the proofs, claims, code, and reported results.

References

Oswin Aichholzer, Franz Aurenhammer, and Hannes Krasser. Enumerating order types for small sets with applications. In *Proceedings of the seventeenth annual symposium on Computational geometry*, pp. 11–18, 2001.

Anders Björner and Torsten Ekedahl. Subspace arrangements over finite fields: cohomological and enumerative aspects. *advances in mathematics*, 129(2):159–187, 1997.

Anders Björner, Michel Las Vergnas, Bernd Sturmfels, Neil White, and Günter M. Ziegler. *Oriented matroids*. Number 46 in Encyclopedia of Mathematics and its Applications. Cambridge University Press, second edition, 1999.

Jacek Bochnak, Michel Coste, and Marie-Françoise Roy. *Real algebraic geometry*. Springer Science & Business Media, 2013.

Joachim Bona-Pellissier, François Malgouyres, and François Bachoc. Local identifiability of deep relu neural networks: the theory. *Advances in neural information processing systems*, 35:27549–27562, 2022.

Joachim Bona-Pellissier, François Malgouyres, and François Bachoc. Geometry-induced regularization in deep relu neural networks. *arXiv preprint arXiv:2402.08269*, 2024.

Stephen Boyd and Lieven Vandenberghe. *Convex Optimization*. Cambridge University Press, Cambridge, UK, 2004. ISBN 9780521833783.

Marie-Charlotte Brandenburg, Georg Loho, and Guido Montúfar. The real tropical geometry of neural networks. *arXiv preprint arXiv:2403.11871*, 2024.

Wenlin Chen and Hong Ge. Neural characteristic activation analysis and geometric parameterization for relu networks. *Advances in Neural Information Processing Systems*, 37:97562–97586, 2024.

Thomas M Cover. Geometrical and statistical properties of systems of linear inequalities with applications in pattern recognition. *IEEE transactions on electronic computers*, EC-14(3):326–334, 1965.

Axel Flinth, Stefano Mereta, and Michele Pernice. On the fibers and semi-algebraicity of relu neuromanifolds. *arXiv preprint arXiv:2606.02826*, 2026.

Timur Garipov, Pavel Izmailov, Dmitrii Podoprikhin, Dmitry Vetrov, and Andrew G Wilson. Loss surfaces, mode connectivity, and fast ensembling of dnns. *Advances in neural information processing systems*, 31, 2018.

Elisenda Grigsby, Kathryn Lindsey, and David Rolnick. Hidden symmetries of relu networks. In *International Conference on Machine Learning*, pp. 11734–11760. PMLR, 2023.

J Elisenda Grigsby, Kathryn Lindsey, Robert Meyerhoff, and Chenxi Wu. Functional dimension of feedforward relu neural networks. *Advances in Mathematics*, 482:110636, 2025.

Moritz Grillo and Guido Montúfar. Most relu networks admit identifiable parameters. *arXiv preprint arXiv:2605.03601*, 2026.

Boris Hanin and David Rolnick. Complexity of linear regions in deep networks. In *International Conference on Machine Learning*, pp. 2596–2604. PMLR, 2019.

Arthur Jacot, Franck Gabriel, and Clément Hongler. Neural tangent kernel: Convergence and generalization in neural networks. *Advances in neural information processing systems*, 31, 2018.

Sungyoon Kim, Aaron Mishkin, and Mert Pilanci. Exploring the loss landscape of regularized neural networks via convex duality. In *International Conference on Learning Representations*, volume 2025, pp. 7760–7813, 2025.

Michael S Matena and Colin Raffel. A combinatorial perspective on the optimization of shallow relu networks. *Advances in Neural Information Processing Systems*, 35:22187–22198, 2022.

Aaron Mishkin and Mert Pilanci. Optimal sets and solution paths of relu networks. In *International Conference on Machine Learning*, pp. 24888–24924. PMLR, 2023.

Guido Montúfar, Razvan Pascanu, Kyunghyun Cho, and Yoshua Bengio. On the number of linear regions of deep neural networks. *Advances in neural information processing systems*, 27, 2014.

Quynh Nguyen. On connected sublevel sets in deep learning. In *International conference on machine learning*, pp. 4790–4799. PMLR, 2019.

Peter Orlik and Hiroaki Terao. *Arrangements of hyperplanes*. Springer Science & Business Media, 2013.

Mary Phuong and Christoph H Lampert. Functional vs. parametric equivalence of relu networks. In *International conference on learning representations*, 2020.

Mert Pilanci and Tolga Ergen. Neural networks are convex regularizers: Exact polynomial-time convex optimization formulations for two-layer networks. In *International Conference on Machine Learning*, pp. 7695–7705. PMLR, 2020.

Pranavkrishnan Ramakrishnan. A complete symmetry classification of shallow relu networks. *arXiv preprint arXiv:2604.14037*, 2026.

Thiago Serra, Christian Tjandraatmadja, and Srikumar Ramalingam. Bounding and counting linear regions of deep neural networks. In *International conference on machine learning*, pp. 4558–4566. PMLR, 2018.

Pierre Stock and Rémi Gribonval. An embedding of relu networks and an analysis of their identifiability. *Constructive Approximation*, 57(2):853–899, 2023.

Michelle L Wachs. Poset topology: tools and applications. *arXiv preprint math/0602226*, 2006.

Liwen Zhang, Gregory Naitzat, and Lek-Heng Lim. Tropical geometry of deep neural networks. In *International Conference on Machine Learning*, pp. 5824–5832. PMLR, 2018.

A Output arrangement and batch dimension

A.1 The finite-data output arrangement

Unless otherwise specified, closures of output sets are taken in the real Zariski topology on $\mathbb{R}^n$; parameter-cell closures are Euclidean.

Proof of Proposition 1. For a realizable strict pattern s, the cone C_s from Section 2 is nonempty and Euclidean open in $\mathbb{R}^{d+1}$. On the parameter cell C_a, Eq. (1) evaluates as

$$E_X(\theta) = \beta_0 \mathbf{1}_n + \sum_{j=1}^{m} \alpha_j D_{s^{(j)}} \widetilde{X} u_j,$$

so the image is contained in V_a. To prove Zariski density in V_a, it suffices to restrict to $\alpha_1 = \cdots = \alpha_m = 1$. The resulting linear map from $\mathbb{R} \times (\mathbb{R}^{d+1})^m$ to V_a is surjective and hence open. It sends the nonempty open set $\mathbb{R} \times C_{s^{(1)}} \times \cdots \times C_{s^{(m)}}$ to a relatively open subset of V_a contained in $E_X(C_a)$. Such a subset is Zariski dense in V_a, proving $\overline{E_X(C_a)}^{\mathrm{Zar}} = V_a$.

Every boundary parameter is a limit of parameters in an incident strict cell. By continuity of E_X, its output belongs to the corresponding V_a, which is Euclidean closed. Thus the full image is contained in $\bigcup_a V_a$. This finite union is Zariski closed, and each V_a is a cell-image closure, proving Eq. (3). □

A.2 Output dimension under affine general position

We now compute the largest branch dimension,

$$d_X(m, d) = \max_{a \in \mathcal{A}(X)^m} \operatorname{rank} B_a(X).$$

By Proposition 1, this is also the dimension of the full output-image closure $\mathcal{V}_X$. The matrix size gives $d_X(m, d) \le \min\{n, m(d+1)+1\}$. This bound is attained for every batch in affine general position, independently of its order type.

Proposition 10 (Output dimension under affine general position)**.** *If every subset of at most $d+1$ points of X is affinely independent, then for every $m \ge 1$,*

$$d_X(m, d) = \min\{n, m(d+1)+1\}. \tag{13}$$

Proof. The upper bound follows from the n rows and $m(d+1)+1$ columns of B_a. First let $n = m(d+1)+1$. Choose a linear functional ℓ whose values on the inputs x_i are distinct and relabel them in increasing projection order. Every prefix and suffix in this order is strictly separated by an affine hyperplane normal to ℓ. We choose one suffix and nested prefixes so that a column relation can be eliminated block by block, starting from the last $d+1$ samples. Select the suffix

$$s_i^* = \mathbf{1}\{i \ge 2\}$$

and the $m-1$ prefixes

$$p_i^{(r)} = \mathbf{1}\{i \le (d+1)r + 1\}, \qquad r = 1, \dots, m-1.$$

In a linear relation among the columns of the resulting B_a, write the coefficient of $\mathbf{1}_n$ as c, the affine form attached to the suffix as $\phi_0(x)$, and those attached to the prefixes as $\phi_r(x)$. On the final $d+1$ samples, only $c + \phi_0(x)$ is present. Those points are affinely independent, so $c + \phi_0 \equiv 0$. If $m > 1$, substituting this identity forces ϕ_{m-1} to vanish on the preceding block of $d+1$ affinely independent samples, hence identically. Continuing backward eliminates all prefix forms. At the first sample the suffix is inactive, so $c = 0$, and then $\phi_0 = 0$. The critical square branch matrix is invertible.

If $n < m(d+1)+1$, extend X to the critical size while preserving affine general position: at each step, choose a new point outside the finitely many proper affine subspaces spanned by at most d existing points. Construct the invertible matrix for this larger batch, then retain the rows of the original samples. Any subset of its rows is linearly independent, and the restricted activation patterns remain realizable, giving rank n. If $n > m(d+1)+1$, order the full batch by ℓ and apply the critical construction to its first $m(d+1)+1$ samples. Using the same separating hyperplanes on the remaining samples preserves the invertible submatrix, giving rank $m(d+1)+1$. □

Under the same affine-general-position assumption, for $n \geq 2$ a full-dimensional branch closure exists exactly when

$$m \geq m_{\mathrm{sat}} := \left\lceil \frac{n-1}{d+1} \right\rceil.$$

This is a threshold for closure dimension: the realizable image contains a nonempty open subset of $\mathbb{R}^n$. Branch incidence, label multiplicity, and gluing in parameter space can still distinguish batches with the same maximal dimension.

B Wall geometry and exact fiber connectivity

B.1 Wall images and ghost directions

Proof of Proposition 2 and Theorem 3. On the common parameter facet in Section 3, the crossing vector u_j lies in $\ker \widetilde{x}_i^\top$ and satisfies strict inequalities for all other samples. This set is nonempty and relatively open in the kernel; the remaining hidden vectors lie in their strict cones. Fix all output weights to one. With these weights fixed, evaluation on the ambient wall subspace is a linear map onto $U + L_{a,b}$. The open-mapping argument of Appendix A.1 therefore gives a nonempty relatively open subset of $U + L_{a,b}$ inside the facet image. Since the entire facet image is contained in this subspace, its Zariski closure is $G_{a,b} = U + L_{a,b}$. Moreover, $L_{a,b} \subseteq L_{s^{(j)}} \cap L_{t^{(j)}}$, so $G_{a,b} \subseteq V_a \cap V_b$, proving Proposition 2. The subspace dimension formula gives Eq. (4).

The wall kernel has codimension one, so $\dim(L_{s^{(j)}}/L_{a,b}) \leq 1$ and likewise for $L_{t^{(j)}}$. Adding U and then taking the quotient by $G_{a,b}$ cannot increase these dimensions. Thus $V_a/G_{a,b}$ and $V_b/G_{a,b}$ are subspaces of $\mathbb{R}^n/G_{a,b}$ of dimension at most one, giving $\gamma \leq 1$ through the identity

$$(V_a/G_{a,b}) \cap (V_b/G_{a,b}) = (V_a \cap V_b)/G_{a,b}.$$

If their intersection is nonzero, they coincide as a line, so $V_a = V_b \supsetneq G_{a,b}$. Conversely, if $V_a = V_b \supsetneq G_{a,b}$, their common quotient is one-dimensional, giving $\gamma = 1$.

It remains to prove the width-one claim. Write $s = s^{(1)}$, $t = t^{(1)}$, and assume $s_i = 0, t_i = 1$. Take any $y \in V_a \cap V_b$ and write $y = c\mathbf{1}_n + D_s\widetilde{X}z = c'\mathbf{1}_n + D_t\widetilde{X}z'$. If another sample $k \neq i$ is inactive on both sides, row k gives $c = c'$ and row i gives $\widetilde{x}_i^\top z' = 0$. Hence $D_t\widetilde{X}z' = D_s\widetilde{X}z'$, so $y \in G_{a,b}$. If no such k exists, t is all-active and s is inactive only at i. Since $\mathbf{1}_n = \widetilde{X}e_{d+1}$, the all-active branch is $V_b = \operatorname{im}\widetilde{X}$. For any z, set $c = \widetilde{x}_i^\top z$ and $h = z - ce_{d+1}$, where e_{d+1} is the last coordinate vector. Then $h \in \ker \widetilde{x}_i^\top$ and $\widetilde{X}z = c\mathbf{1}_n + D_s\widetilde{X}h$, so the all-active branch already lies in $G_{a,b}$. Swapping the two sides covers deactivation and proves Theorem 3. □

Parameter witnesses for the five-point ghost example. For $X = (0, 1, 2, 3, 4)$, let $a = (11100, 00111)$ and $b = (11110, 00111)$. With parameters listed as $(\beta_0; \alpha_1, u_1; \alpha_2, u_2)$, take

$$\theta_a = \big(0; 2, (-1, 5/2); 2, (1, -3/2)\big), \qquad \theta_b = \big(-2; 2, (-1, 7/2); 3, (1, -5/3)\big).$$

Their strict activation labels are a and b, respectively, and direct evaluation gives $E_X(\theta_a) = E_X(\theta_b) = (5, 3, 2, 3, 5)^\top$. At the wall point $u_1 = (-1, 3)$, only sample $x_4 = 3$ has zero pre-activation; a small change of the first neuron's bias realizes either adjacent pattern. The output bias and the second neuron contribute

$$U = \operatorname{span}\{\mathbf{1}_5, (0, 0, 1, 1, 1)^\top, (0, 0, 2, 3, 4)^\top\}, \qquad L_{a,b} = \operatorname{span}\{(3, 2, 1, 0, 0)^\top\}.$$

These four generators are independent and annihilated by $q = (-1, 1, 1, -2, 1)^\top$, hence $G_{a,b} = U + L_{a,b} = \ker q^\top$. The column $(1, 1, 1, 0, 0)^\top$ belongs to L_{11100} and has $q^\top$-value 1; the column $(1, 1, 1, 1, 0)^\top$ belongs to L_{11110} and has value -1. Each branch therefore extends $G_{a,b}$ to $\mathbb{R}^5$. Finally, $q^\top E_X(\theta_a) = -1$, certifying that the shared target lies outside the wall-image closure. No path preserving this target can cross the common wall.

Exact wall-image computations. For 6,008 adjacent-cell wall cases, we compute the branch intersection $V_a \cap V_b$ and the true-wall image closure $G_{a,b}$ over $\mathbb{Q}$. Their dimensions differ in 158 cases, each by one, and every case agrees with Theorem 3. Appendix G.1 records the common computational methods and coverage.

B.2 GLOBAL EXACT-FIT FIBERS

Proof of Theorem 4. The ReLU graph is semialgebraic, hence so is $\mathcal{F}_y = E_X^{-1}(y)$. Two points in the same path component of $\mathcal{F}_y$ admit a semialgebraic path. Pull back the activation-hyperplane stratification along it. Semialgebraic cell decomposition (Bochnak et al., 2013) partitions the interval into finitely many points and open intervals with constant preactivation signs.

The path image on each open interval lies in a fixed arrangement face, possibly of positive codimension. Choose a strict chamber incident to that face; its closure contains the entire path segment and both endpoints. At each shared endpoint, the chambers chosen for the two neighboring intervals are incident to the face containing that point. In a sufficiently small ball around the point, only hyperplanes containing this face are present. A generic polygonal path between interior points of the two chosen chambers crosses these hyperplanes one at a time, giving an adjacency walk through incident chambers. This auxiliary path determines the chamber sequence; the wall-closure memberships follow from the original fiber path's shared endpoint. Distinct inputs give distinct activation hyperplanes, so each crossed facet is a true single wall. For every adjacency a, b in this sequence, that endpoint lies in the Euclidean closure of the shared facet $H_{a,b}$. Continuity of E_X gives

$$y \in \overline{E_X(H_{a,b})}^{\mathrm{Eucl}} \subseteq \overline{E_X(H_{a,b})}^{\mathrm{Zar}} = G_{a,b}.$$

The same point lies in the closure of each incident chamber, so $y \in V_a$ at every intermediate vertex. Replacing every higher-codimension event by this finite local walk and concatenating proves connectivity in $\Gamma_G(y)$, including paths that follow activation boundaries for whole intervals. Passing from ordered cells to pattern multisets is a graph quotient and can only merge components. □

Proof of Proposition 5. The finitely many closed chamber fibers $\mathcal{F}_{a,y}$ cover $\mathcal{F}_y$. A closed semialgebraic set has finitely many connected components, each closed and semialgebraically path connected (Bochnak et al., 2013); these are the $\mathcal{F}_{a,y}^{\lambda}$. If two vertices of $\mathfrak{N}_y$ share an edge, concatenate paths in the two pieces through a point of their intersection. Induction along a graph walk shows that the union associated with one graph component is path connected.

Different graph-component unions are disjoint: an intersection would itself create an edge. Each is a finite union of closed subsets of $\mathcal{F}_y$, and its complement is the finite union associated with the other graph components. Thus it is both open and closed relative to $\mathcal{F}_y$. It is therefore exactly one connected, hence path-connected, component of the semialgebraic fiber. □

Proof for Example 1. More generally, let $x_1 \neq x_2$ and $m = 1$. For a constant target, $\mathcal{F}_y$ is contractible. For a nonconstant target, it has exactly two path components, distinguished by $\operatorname{sgn}(\alpha)$, although $\Gamma_G(y)$ is connected. We prove this general statement below.

Because $x_1 \neq x_2$, the affine evaluation map

$$T : \mathbb{R}^{d+1} \to \mathbb{R}^2, \qquad u = (w, b) \mapsto (z_1, z_2) = (w^\top x_1 + b, w^\top x_2 + b)$$

is surjective and has kernel dimension $d - 1$. Put $\delta = y_2 - y_1$. The exact-fit equations imply

$$\delta = \alpha\big(\mathrm{ReLU}(z_2) - \mathrm{ReLU}(z_1)\big), \qquad \beta_0 = y_1 - \alpha\,\mathrm{ReLU}(z_1). \tag{14}$$

If $\delta \neq 0$, then $\alpha \neq 0$, so its sign is constant on every fiber path. Both signs occur. Let $\kappa = \operatorname{sgn}(\delta)$. Choose $T(u) = (-\kappa, \kappa)$ and $\alpha = |\delta|$ for a positive-weight witness; choose $T(u) = (\kappa, -\kappa)$ and $\alpha = -|\delta|$ for a negative-weight witness, with β_0 determined by Eq. (14).

To show that each sign sector is path connected, choose a linear right inverse R of T. When $r = \delta/\alpha > 0$, the solutions of $\mathrm{ReLU}(z_2) - \mathrm{ReLU}(z_1) = r$ are parameterized by

$$z_1 \in \mathbb{R}, \qquad z_2 = r + \max\{z_1, 0\}.$$

They lift to all network parameters in this sector through

$$u = R(z_1, z_2) + v, \qquad v \in \ker T, \qquad \beta_0 = y_1 - \alpha\,\mathrm{ReLU}(z_1).$$

These formulas continuously parameterize the sector by (α, z_1, v), with α in a fixed sign half-axis. For $r < 0$, interchange the two samples. Each parameter domain is the product of a half-axis, $\mathbb{R}$, and $\ker T$, hence path connected. There are therefore exactly two fiber components.

If $\delta = 0$, contractibility follows from the constant-target deformation retraction proved for arbitrary batches and widths in Appendix B.3.

Finally, for a nonconstant two-sample target the branch closures for patterns 01, 10, and 11 are all $\mathbb{R}^2$. The shared-wall closures on 01–11 and 11–10 are also $\mathbb{R}^2$, so $\Gamma_G(y)$ is connected. For $\delta > 0$, the parameterizations above give four path-connected pieces of the closed-cell fibers, with the only edges of $\mathfrak{N}_y$ being

$$(01, +) \longleftrightarrow (11, +), \qquad (11, -) \longleftrightarrow (10, -),$$

where the signs refer to α. The two edges correspond to $z_1 = 0$ and $z_2 = 0$, respectively; the 00 closed-cell fiber is empty. For $\delta < 0$, the signs reverse. Thus the two components of the 11 closed-cell fiber separate the connected coarse graph into the two components of $\mathfrak{N}_y$. □

Computed two-sample fibers. Exact computations for 24 rational width-one cases in input dimensions one and two give eight connected constant-target fibers and 16 two-component nonconstant-target fibers (Table 3), in agreement with the preceding proof.

B.3 Signed fiber pieces and the width transition

The two-sample example separates a cell fiber by its output-weight sign. For arbitrary batches and widths, fixing all weak output-weight signs gives the path-connected cover used in Theorem 6.

Linear tests for intersections. For signed labels $\sigma = (a, \varepsilon)$ and $\tau = (b, \eta)$, write $a = (s^{(1)}, \dots, s^{(m)})$ and $b = (t^{(1)}, \dots, t^{(m)})$. Their pieces meet precisely when the following system is feasible:

$$\begin{cases} v_j \in K_{s^{(j)}} \cap K_{t^{(j)}}, & \varepsilon_j = \eta_j, \\ v_j = 0, & \varepsilon_j \neq \eta_j, \end{cases} \qquad j = 1, \dots, m,$$
$$\beta_0 \mathbf{1}_n + \sum_{j:\varepsilon_j = \eta_j} \varepsilon_j D_{s^{(j)}} \widetilde{X} v_j = y. \tag{15}$$

Denote the feasible polyhedron of this system by $\mathcal{P}_{\sigma,\tau}(y)$. The constraints allow simultaneous activation changes and zero output weights, including intersections on faces of higher codimension.

Proof of Theorem 6. The finite central arrangement $\widetilde{x}_i^\top u = 0$ has nonzero defining forms, since each lifted row has last coordinate one. Its chamber closures cover $\mathbb{R}^{d+1}$ even when rows repeat or are dependent. For each realizable s, we have $K_s = \overline{C_s}^{\mathrm{Eucl}}$: fixing $u_* \in C_s$, every $u \in K_s$ is approached by $u + \epsilon u_* \in C_s$ as $\epsilon \downarrow 0$. Consequently the sets $\mathcal{F}_{\sigma,y}$ form a finite closed semialgebraic cover of $\mathcal{F}_y$, including representations supported entirely on activation boundaries.

Positive homogeneity gives the surjection $\mathcal{F}_{\sigma,y} \to \mathcal{P}_\sigma(y)$ by $v_j = |\alpha_j| u_j$. The canonical lift $\alpha_j = \varepsilon_j$, $u_j = v_j$ is a right inverse. To join a parameter point to this lift, first handle every zero output weight: shrink its hidden vector to zero at zero weight, then move the weight to ε_j with the hidden vector zero. For each initially nonzero weight, use

$$\rho_j(t) = (1 - t)|\alpha_j| + t > 0, \qquad \alpha_j(t) = \varepsilon_j \rho_j(t), \qquad u_j(t) = \frac{|\alpha_j|}{\rho_j(t)} u_j.$$

These paths preserve the output and all weak sign constraints. Their endpoint is the canonical lift of the original normalized point. The lift of a straight segment in the convex polyhedron then joins any two such endpoints. Thus every nonempty $\mathcal{F}_{\sigma,y}$ is path connected.

Suppose a parameter point belongs to both $\mathcal{F}_{\sigma,y}$ and $\mathcal{F}_{\tau,y}$. Opposite assigned signs force $\alpha_j = 0$, hence $v_j = 0$. At equal signs, v_j belongs to both activation cones. This proves necessity of Eq. (15). Conversely, lift a feasible edge solution by

$$(\alpha_j, u_j) = \begin{cases} (\varepsilon_j, v_j), & \varepsilon_j = \eta_j, \\ (0, 0), & \varepsilon_j \neq \eta_j. \end{cases}$$

The zero vector lies in every cone. At equal signs, differing mask bits force the corresponding preactivation to be zero, so $D_{s^{(j)}} \widetilde{X} v_j = D_{t^{(j)}} \widetilde{X} v_j$. The lifted point therefore lies in both original signed pieces and has output y.

The finite closed-cover argument in Appendix B.2 now applies: paths concatenate through intersection witnesses within each graph component, while distinct graph-component unions are disjoint and relatively open and closed in the fiber. This proves Eq. (8). Labels containing the same endpoint are adjacent, so its graph component is unambiguous. Neuron order is retained throughout. □

Complexity and exact certificates. If $r = |\mathcal{A}(X)|$, there are at most $N = (2r)^m$ signed labels. After activation-pattern enumeration, a direct algorithm needs at most $N + \binom{N}{2}$ feasibility tests, each with $1 + m(d+1)$ variables and $O(m(n+d)+n)$ inequalities after doubling equalities. The graph is finite but can be exponentially large. By comparison, the wall-closure graph starts from r^m ordered labels and uses rank tests only for branch membership and Hamming-one adjacencies, with at most $mnr^m/2$ candidate edges. It can certify disconnection without constructing the signed graph, as in the five-point example below.

For rational X, y, enumerate strict activation patterns by testing every binary mask with normalized margins $(2s_i - 1)\widetilde{x}_i^\top u \geq 1$. Each resulting vertex or edge system has rational coefficients. A feasible system $Mz \leq h$ has a rational witness; by Farkas' lemma (Boyd & Vandenberghe, 2004, Section 5.8), an infeasible one admits rational multipliers

$$\lambda \geq 0, \qquad \lambda^\top M = 0, \qquad \lambda^\top h = -1.$$

Rational arithmetic verifies these certificates. Our complete graph certificate records every mask and vertex decision and every pair of retained vertices, including rejected edges.

Global connectivity of the five-point ghost example. Return to $X = (0, 1, 2, 3, 4)$, $y = (5, 3, 2, 3, 5)^\top$, and $m = 2$ from Appendix B.1. There are ten strict patterns, namely the prefixes and suffixes, including the all-zero and all-one patterns. Of the 100 ordered pairs, exactly 12 satisfy $y \in V_a$: the four pairs displayed below, $(00011, 01111)$, $(11000, 11110)$, and their neuron swaps. Each of these 12 branch matrices is nonsingular. Their only Hamming-one edges are

$$(00111, 11110) \longleftrightarrow (00111, 11100) \longleftrightarrow (01111, 11100)$$

and the two neuron-swapped edges. On each edge, the branch rank is five, the wall-image rank is four, and adjoining y to the wall generators raises their rank to five. Thus each edge has a ghost gap of one and is excluded from $\Gamma_G(y)$. In particular, the witnesses θ_a, θ_b lie in different fiber components by Theorem 4.

To count all components, solve the signed systems of Eq. (7). For $\varepsilon = (+, +)$, the four realizable pairs have the following unique normalized solutions:

a	β_0	v_1	v_2
$(00011, 11000)$	2	$(2, -5)$	$(-2, 3)$
$(00111, 11100)$	0	$(2, -3)$	$(-2, 5)$
$(00111, 11110)$	-2	$(3, -5)$	$(-2, 7)$
$(01111, 11100)$	-2	$(2, -1)$	$(-3, 7)$

All preactivations are strictly signed according to their masks, and changing either output-weight sign makes the corresponding cone constraints infeasible. For the other two pairs, the unique absorbed vector $\alpha_2 u_2$ is $(-1, -1)$ or $(1, -5)$, respectively. Its preactivations have constant sign on X, incompatible with the mixed mask 01111 or 11110 for either nonzero sign of α_2; zero weight cannot realize either nonzero vector. These pairs therefore have no feasible signed piece.

Including neuron swaps, the 400 candidate signed labels yield eight nonempty pieces. Each lies strictly inside its ordered activation cell with positive output weights, so the pieces are pairwise disjoint. Theorem 6 makes each piece path connected, proving that the original ordered fiber has exactly eight components. Up to neuron permutation, the same enumeration gives six algebraic labels but only four strictly realizable labels. Hence $b_{\mathrm{lab}}(y) = 6$, illustrating the distinction in Section 4.2.

Proof of Corollary 7. Let $T = \widetilde{X}$ be surjective and choose a linear right inverse R. For one unit with preactivations $z = Tu$, the path

$$z(t) = (1 - t)z + t\,\mathrm{ReLU}(z), \qquad u(t) = u + R\big(z(t) - z\big)$$

leaves positive coordinates unchanged and raises negative coordinates to zero. It preserves the ReLU output, stays in the original closed activation cone, and ends in the all-active closed cone. Applying this path to all units joins each feasible signed piece to the all-active piece with the same signs.

Every all-active sign tuple is feasible. Use one positive-sign neuron with $\beta_0 \le \min_i y_i$, or one negative-sign neuron with $\beta_0 \ge \max_i y_i$; surjectivity realizes the required nonnegative preactivations, and all other units may be zero. For $m \ge 2$, two all-active sign tuples differing at neuron j share a point with $\alpha_j = u_j = 0$, using another neuron to fit y. The subgraph on all-active vertices contains a connected sign hypercube reached by every feasible vertex. At width one, a nonconstant target forbids $\alpha = 0$, separating the two signs; each sign sector is connected through its all-active piece. This proves the claim.

Finally, for $y = c\mathbf{1}_n$ and arbitrary X, keep all hidden vectors fixed and set

$$\beta_0(t) = c + (1-t)(\beta_0 - c), \qquad \alpha_j(t) = (1-t)\alpha_j.$$

The output remains $c\mathbf{1}_n$. This is a deformation retraction onto $\{\beta_0 = c, \alpha = 0\} \cong \mathbb{R}^{m(d+1)}$, which is contractible. □

A six-component fiber with dependent inputs. Affine independence is essential. For the collinear batch $X = (0, 1, 2)$ and tent target $y = (0, 1, 0)$ at width $m = 2$, neither α_j nor the scalar input slope w_j can vanish: that neuron would contribute only a constant, leaving a monotone one-neuron function. Thus the ordered sign tuple $(\operatorname{sgn}\alpha_1, \operatorname{sgn} w_1, \operatorname{sgn}\alpha_2, \operatorname{sgn} w_2)$ is constant along every path in the fiber. The three fits

$$\begin{aligned} f_1(x) &= \mathrm{ReLU}(x+1) - 2\,\mathrm{ReLU}(x-1) - 1, \\ f_2(x) &= \mathrm{ReLU}(3-x) - 2\,\mathrm{ReLU}(1-x) - 1, \\ f_3(x) &= 1 - \mathrm{ReLU}(1-x) - \mathrm{ReLU}(x-1) \end{aligned}$$

have sign tuples $(+,+,-,+)$, $(+,-,-,-)$, and $(-,-,-,+)$, respectively. Together with their neuron-swapped versions, they give six distinct tuples, so the fiber has at least six path components. The exact graph computation below shows that it has exactly six.

Computed signed-fiber components. For the tent data, the linear tests above yield 32 nonempty pieces from 144 candidate signed labels and 56 edges. The resulting graph has six components of sizes $4, 4, 4, 4, 8, 8$, giving the upper bound by Theorem 6. For the affinely independent triangle $X = ((0,0), (1,0), (0,1))$ with $y = (0, 1, 2)$ at the same width, the graph is connected, as predicted by Corollary 7. Table 3 summarizes eleven rational examples, including batches with repeated inputs.

C Proofs for intersection lattices and label multiplicity

C.1 Proof of generic lattice constancy

Matching matrices. We first express multiple-intersection dimensions in terms of matrix ranks. Each branch matrix has $p = 1 + m(d+1)$ columns. For labels $a_1, \dots, a_k$ with $k \ge 2$, define

$$C_{a_1,\dots,a_k}(X) = \begin{bmatrix} B_{a_1} & -B_{a_2} & 0 & \cdots & 0 \\ B_{a_1} & 0 & -B_{a_3} & \cdots & 0 \\ \vdots & & & \ddots & \vdots \\ B_{a_1} & 0 & 0 & \cdots & -B_{a_k} \end{bmatrix}. \tag{16}$$

Its kernel consists of coefficient tuples with a common output in all k branches. For $|J| \ge 2$, write C_J using a fixed ordering of the labels. Singleton intersections are handled directly by B_a.

Lemma 11 (Intersection–rank identity)**.** *For $k \ge 2$,*

$$\dim \bigcap_{j=1}^{k} V_{a_j}(X) = \sum_{j=1}^{k} \operatorname{rank} B_{a_j}(X) - \operatorname{rank} C_{a_1,\dots,a_k}(X). \tag{17}$$

Proof. Write $B_j = B_{a_j}$. The matching matrix represents

$$F : (\mathbb{R}^p)^k \to (\mathbb{R}^n)^{k-1}, \quad (c_1, \dots, c_k) \mapsto (B_1c_1 - B_2c_2, \dots, B_1c_1 - B_kc_k).$$

Mapping a kernel tuple to its common output gives a surjection $\ker F \to \cap_j \operatorname{im} B_j$. Its kernel is $\oplus_j \ker B_j$. Therefore

$$\dim \ker F = kp - \sum_j \operatorname{rank} B_j + \dim \bigcap_j \operatorname{im} B_j,$$
$$\dim \ker F = kp - \operatorname{rank} C_{a_1,\dots,a_k}.$$

Equating the two expressions proves Eq. (17). □

To recover incidence, we retain the intersection dimension for each labeled subfamily, rather than only a dimension histogram. The polynomial entries of the branch and matching matrices let us fix all these dimensions simultaneously away from the zero set of finitely many minors.

Proof of Theorem 8. The finite label set $\mathcal{C} = \mathcal{A}_\chi^m$ is fixed throughout $\mathcal{R}_\chi$. For $J \subseteq \mathcal{C}$, define

$$I_J(X) = \bigcap_{a \in J} V_a(X), \qquad I_\varnothing(X) = \mathbb{R}^n, \qquad d_J(X) = \dim I_J(X).$$

Collect every $B_a(X)$ and every $C_J(X)$ with $|J| \geq 2$. With labels fixed, these are polynomial matrices on the ambient space $(\mathbb{R}^d)^n$. For each such matrix $P(X)$, let ρ_P be its maximum rank on this space. If $\rho_P > 0$, choose a ρ_P-minor whose determinant δ_P is a nonzero polynomial; if $\rho_P = 0$, set $\delta_P = 1$. Define

$$\Delta_{\chi,m} = \prod_P \delta_P.$$

The polynomial ring is an integral domain, so the finite product is nonzero. Since $\mathcal{R}_\chi$ is a nonempty open set defined by strict determinant signs, this product cannot vanish identically there. Its zero set has empty interior and Lebesgue measure zero, so $\mathcal{R}_\chi \setminus Z(\Delta_{\chi,m})$ is Euclidean open dense in $\mathcal{R}_\chi$. On this set every matrix has rank ρ_P. Lemma 11 therefore fixes $d_J(X)$ for $|J| \geq 2$; branch ranks fix singleton dimensions, and $d_\varnothing(X) = n$.

At any fixed configuration, suppressing X in the notation, we have

$$I_J = I_K \iff d_J = d_K = d_{J\cup K}, \qquad I_J \subseteq I_K \iff d_J = d_{J\cup K}. \tag{18}$$

Since $I_J \cap I_K = I_{J\cup K}$, the dimension criteria in Eq. (18) follow from $\dim(A \cap B) = \dim A$ iff $A \subseteq B$. Thus for any $X, Y \in \mathcal{R}_\chi \setminus Z(\Delta_{\chi,m})$, the map $I_J(X) \mapsto I_J(Y)$ is well defined, preserves dimensions and inclusion, and has the inverse obtained by exchanging X and Y. It is therefore an isomorphism of the posets of all branch intersections, including $\mathbb{R}^n$.

Applying Eq. (18) to singleton subsets fixes branch equality, inclusion, and maximality. Restricting the isomorphism to intersections of maximal branches gives $\Lambda_X \cong \Lambda_Y$, matching branches with the same labels. Finally, Hamming-one adjacency depends only on $\mathcal{A}_\chi$. Retaining the fixed set of maximal-branch labels and identifying the fixed branch-equality classes therefore gives the same graph for X and Y. □

C.2 Controlling the lattice with small subfamilies

The construction in Appendix C.1 uses every label subset. We now bound the intersection order needed to control the entire lattice. Keep χ fixed and index branches by width-m multisets: permuting neurons only permutes column blocks of B_a, leaving branch subspaces and matching ranks unchanged. Here J indexes multiset labels, with a fixed ordered representative for each label used to form B_a and C_J. Let M be the number of these labels and $p = 1 + m(d+1)$.

Lemma 12 (Small-subfamily control)**.** *Set*

$$h = \min\{n-1, p, M\}.$$

Every intersection $I_J(X)$ equals $I_K(X)$ for some $K \subseteq J$ with $|K| \leq h$, allowing $K = \varnothing$ for the ambient space. If all branch ranks and all matching ranks for $2 \leq |J| \leq h$ retain their generic values, then all labeled intersection dimensions and incidence relations are generic.

Proof. All branches contain $W = \mathrm{span}\{\mathbf{1}_n\}$. In $\mathbb{R}^n/W$, begin with the empty intersection and add branches from J whenever they strictly decrease the current intersection. The process stops at I_J/W: otherwise some branch in J would decrease it further. Each step lowers dimension by at least one, so at most $n-1$ branches are selected. After the first step, the quotient dimension is at most $p-1$, giving at most p steps in total. At most M labels can be selected, proving the first claim.

For the second claim, work first over $\mathbb{R}(X)$, the rational function field in the input coordinates. Ranks over this field equal the generic ranks used in Appendix C.1. The same argument gives, for each $|J| \geq 2$, a subset $K \subseteq J$ with $|K| \leq h$ and $I_K = I_J$ over this field. Let d_J^{gen} denote their common generic dimension. At a configuration where the specified ranks are generic, the dimension of I_K is fixed, so

$$\dim I_J(X) \leq \dim I_K(X) = d_J^{\mathrm{gen}}.$$

Conversely, branch ranks are unchanged and $\mathrm{rank}\, C_J(X)$ cannot exceed its function-field rank. Lemma 11 therefore gives $\dim I_J(X) \geq d_J^{\mathrm{gen}}$. Empty and singleton intersection dimensions are already fixed. Thus all dimensions are generic; Eq. (18) then fixes the incidence relations. □

Size and degree bounds. For a uniform rank-$(d+1)$ affine chirotope, Cover's counting formula (Cover, 1965) and multiset counting give

$$r = 2\sum_{j=0}^{d} \binom{n-1}{j}, \qquad M = \binom{r+m-1}{m}.$$

The polynomial $\Delta_{\chi,m}$ in Theorem 8 can therefore be chosen using only branch matrices and matching matrices through order h, with at most $\sum_{k=1}^{h} \binom{M}{k}$ selected-minor factors. At most md columns of each branch matrix depend on input coordinates, and their entries are affine-linear. A determinant uses one entry per column, so its degree is bounded by both its order and the number of such columns. For a k-branch matching matrix, these bounds are $(k-1)n$, kp, and kmd. Consequently,

$$\deg \Delta_{\chi,m} \leq M \min\{n, p, md\} + \sum_{k=2}^{h} \binom{M}{k} \min\{(k-1)n, kp, kmd\}. \tag{19}$$

C.3 Proof of the multiplicity stratification

The weighted all-branch lattice of Section 4.2 is formally the embedded poset

$$\mathfrak{L}_X^{\mathrm{lab}} = (\Lambda_X^{\mathrm{lab}}, \preceq, \iota, \mathsf{A}_X, w),$$

where $U \preceq W$ means $U \subseteq W$, A_X marks the branch elements, and $\iota(W) = W \subseteq \mathbb{R}^n$ records their positions in output space. The weighted poset determines the combinatorial multiplicity filtration; the embedding specifies the corresponding output loci.

Proof of Theorem 9. Fix any finite batch X. A finite union of proper linear subspaces cannot cover a flat W, so every W° is nonempty. For any $y \in \mathbb{R}^n$, define

$$W_y = \bigcap_{\substack{V \in \mathsf{A}_X \\ y \in V}} V,$$

with the empty intersection equal to $\mathbb{R}^n$. This is the smallest lattice flat containing y, hence $y \in W_y^\circ$. If $y \in W^\circ$, then $W_y \subseteq W$, and proper inclusion would contradict the definition of W°. Thus $W = W_y$, proving that the strata partition $\mathbb{R}^n$.

For $y \in W^\circ$, we have $W = W_y$, so the branches containing y are exactly those containing W. Summing their weights gives

$$b_{\mathrm{lab}}(y) = \sum_{\substack{V \in \mathsf{A}_X \\ V \supseteq W}} w(V) = \nu(W),$$

proving the stratum formula. Since $U \subseteq W$ implies $\nu(U) \geq \nu(W)$, every $y \in W$ satisfies $b_{\mathrm{lab}}(y) = \nu(W_y) \geq \nu(W)$. Thus each flat with $\nu(W) \geq k$ lies in $\Omega_{\geq k}$; conversely, every $y \in \Omega_{\geq k}$ belongs to the qualifying flat W_y. This proves Eq. (12).

Removing flats contained in other qualifying flats leaves M_k with the same union. To prove uniqueness, consider any finite representation of $\Omega_{\geq k}$ by lattice flats. Each $W \in \mathsf{M}_k$ must lie in one of its flats T, since the intersections with those flats cover W. By the same finite-union argument, T lies in some $W' \in \mathsf{M}_k$. Maximality and $W \subseteq T \subseteq W'$ force $W = T = W'$. Thus every representation contains M_k; all additional flats are redundant, proving uniqueness. □

For a fixed realizable uniform chirotope χ, Appendix C.1 fixes the branch-equality classes and all labeled intersection relations on $\mathcal{R}_\chi \setminus Z(\Delta_{\chi,m})$, so the multiset-label counts $w(V)$ are fixed as well. Consequently, the induced isomorphism preserves ν, each M_k, and the combinatorial structure of the filtration $\{\Omega_{\geq k}\}_k$.

D Six-point lattices: generic classification and exceptional loci

For six planar inputs and one hidden neuron, every uniform order type has 32 strict activation patterns and maximal output dimension four (Proposition 10). These shared counts leave open how the branches intersect and how many activation labels explain a given output. This case makes the structures of Section 4 explicit: we classify the generic maximal-branch lattices, compute label multiplicity, and determine which coordinate choices change the generic incidence structure within a fixed order type.

Up to point relabeling and global reflection, the six-point case has 16 uniform planar order types (Aichholzer et al., 2001). Table 1 summarizes their generic lattices, multiplicities, and exceptional loci; throughout this appendix, $(n, d, m) = (6, 2, 1)$.

D.1 Generic maximal-branch lattices

We first ask whether the generic lattice distinguishes the 16 input order types. An invertible affine change of input coordinates preserves the corresponding branch subspaces, so we fix a noncollinear triple and write the remaining six coordinates as (u, v, r, s, w, z). Over $\mathbb{Q}(u, v, r, s, w, z)$, we construct the branch matrices B_a, retain the distinct maximal branches, and close them under intersection. This gives the following classification.

Theorem 13 (Complete generic lattice classification)**.** *For $(n, d, m) = (6, 2, 1)$, the 16 uniform planar order types produce 16 pairwise nonisomorphic generic dimension-labeled maximal-branch intersection posets. Their full characteristic-polynomial coefficient tuples are also pairwise distinct. Every function-field type is matched by an exact rational realization in the same open chirotope cell.*

Proof by exact computation. Function-field ranks determine intersection dimensions and containments, with reduced row spaces of basis transposes identifying equal subspaces. The rank argument of Appendix C.1, applied in each affine gauge, shows that these ranks and incidence relations hold on an open dense subset of the realization cell. For each of the 16 types, reconstruction at an exact rational configuration in that cell matches the function-field lattice with activation labels fixed, checking all flat dimensions and Hasse covers. The same check is performed for the all-branch lattice.

For each resulting poset, compute the Möbius function from its Hasse covers. The pairs $(|\Lambda_X|, \mu)$ in Table 1 are distinct for all 16 types. Both entries are poset invariants, so these pairs already prove that the generic posets are pairwise nonisomorphic. With the inclusion order of Section 4.1, the characteristic polynomial (Björner & Ekedahl, 1997) is

$$P_{\Lambda_X}(\xi) = \sum_{W \in \Lambda_X} \mu_{\Lambda_X}(W, \mathbb{R}^6)\, \xi^{\dim W}.$$

Exact comparison of all $\binom{16}{2} = 120$ pairs confirms that the coefficient tuples are pairwise distinct, verifying the theorem's additional assertion. □

The Möbius number alone takes only 14 distinct values. Dimension counts also lose incidence information: types ot6_06 and ot6_15 have identical dimension histograms and total cover counts, but different cover counts between individual dimension levels. The lower block of Table 1 gives this comparison; the upper block also records the multiplicities and exceptional-factor counts developed below.

Table 1: **Six-point order types: lattices, multiplicities, and exceptional loci. Upper block:** census IDs are grouped by hull size, with types 06 and 15 shaded. h counts hull vertices, $b = |\mathcal{M}_X|$ maximal branches, and $|\Lambda_X|$ intersection flats, including $\mathbb{R}^6$. The Möbius number is $\mu = \mu_{\Lambda_X}(\mathrm{span}\{\mathbf{1}_6\}, \mathbb{R}^6)$. The Betti vector $\widetilde{\beta}_{\mathbb{F}_2} = (\dim_{\mathbb{F}_2} \widetilde{H}_i(\mathcal{O}(P_X); \mathbb{F}_2))_{i=0}^{2}$ uses the order complex $\mathcal{O}(P_X)$, whose simplices are chains in $P_X = \Lambda_X \setminus \{\mathrm{span}\{\mathbf{1}_6\}, \mathbb{R}^6\}$ (Wachs, 2006). q counts intrinsic irreducible hypersurfaces in Eq. (20), $\Delta_{\mathrm{all}} = |\Lambda_X^{\mathrm{lab}} \setminus \Lambda_X|$ counts the additional flats, and $g_2 = |\mathsf{M}_2|$ counts the irredundant generators of $\Omega_{\geq 2}$ (Theorem 9). **Lower block:** a Hasse cover $r \to s$ joins flats of dimensions r and s with no intervening flat. Types 06 and 15 both have dimension histogram $(1{:}1, 2{:}83, 3{:}33, 4{:}18, 6{:}1)$ and 365 covers; bold counts distinguish their incidence profiles.

Generic structure across 16 order types

	Geometry			Topology		Factors	All branches	
Type	h	b	$\lvert\Lambda_X\rvert$	μ	$\widetilde{\beta}_{\mathbb{F}_2}$	q	Δ_{all}	g_2
ot6_01	6	18	117	-24	$[0, 24, 0]$	15	0	43
ot6_02	5	18	126	-28	$[0, 28, 0]$	15	3	49
ot6_03		19	130	-23	$[0, 25, 2]$	17	0	48
ot6_09		20	134	-14	$[0, 20, 6]$	20	0	46
ot6_04	4	19	138	-26	$[0, 28, 2]$	11	2	52
ot6_05		20	142	-19	$[0, 25, 6]$	9	0	52
ot6_06		18	136	$\mathbf{-32}$	$[0, 32, 0]$	14	6	55
ot6_07		19	140	-28	$[0, 30, 2]$	11	3	55
ot6_08		20	142	-18	$[0, 24, 6]$	11	2	51
ot6_15		18	136	$\mathbf{-33}$	$[0, 33, 0]$	14	8	56
ot6_10	3	19	147	-29	$[0, 31, 2]$	4	4	56
ot6_11		20	151	-22	$[0, 28, 6]$	3	4	56
ot6_12		20	151	-23	$[0, 29, 6]$	2	2	57
ot6_13		19	149	-31	$[0, 33, 2]$	6	5	59
ot6_14		21	153	-12	$[0, 24, 12]$	4	0	55
ot6_16		18	147	-36	$[0, 36, 0]$	12	9	61

Hasse cover counts: types 06 and 15

Type	$6 \to 4$	$4 \to 3$	$4 \to 2$	$3 \to 2$	$2 \to 1$
ot6_06	18	68	**54**	**142**	83
ot6_15	18	68	**56**	**140**	83

D.2 Label multiplicity across thresholds

To study multiplicity beyond the maximal-branch lattice, we retain all 32 activation labels. We compute the all-branch lattice over the same rational function field and match it at an exact rational realization of each type. In each of the 16 generic types, these labels give 32 distinct branch subspaces, so $w(V) = 1$ for every branch. Multiplicity therefore counts how many distinct branches contain an output. On the constant-output line, all 32 branches contribute. Retaining the nonmaximal branches also adds intersection flats in 11 types, as recorded by Δ_{all} in Table 1.

Across $1 \leq k \leq 32$, each type has 7–10 distinct generator antichains M_k, hence 7–10 distinct superlevel sets by Theorem 9. Several consecutive thresholds thus give the same set. These sets are nested as k increases, but the number of flats in their irredundant representations need not decrease. For all 16 generic types, retaining nonmaximal branches leaves the Möbius number, characteristic polynomial, and proper-part mod-2 Betti profile unchanged. The additional flats and branch membership carry information beyond these invariants.

Computing multiplicities on open strata. For each rational realization, we choose a rational point $y \in W^\circ$ by avoiding the finitely many proper lattice subflats of W. Such points exist because rational points are dense in W and W° is nonempty and relatively open. Testing y against all 32 branch

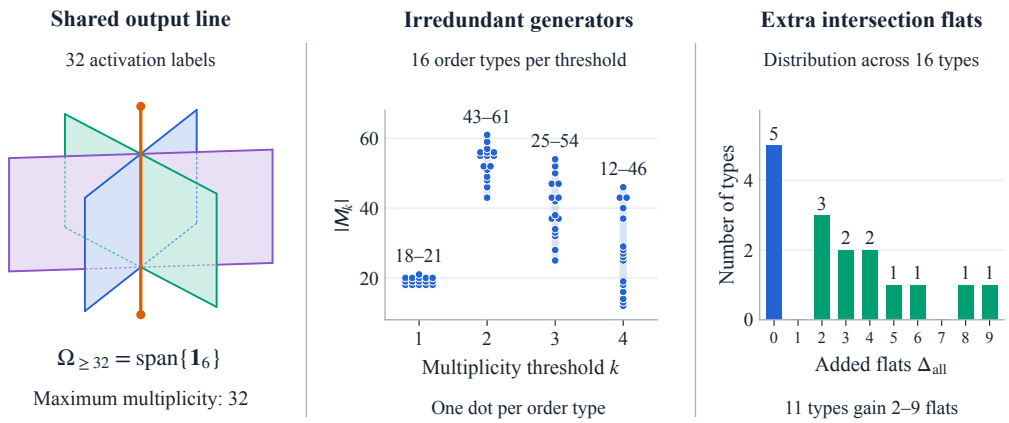

Figure 3: **Label multiplicity and additional intersection flats.** For six planar inputs and one hidden neuron, $(n, d, m) = (6, 2, 1)$. **Left:** schematic branch subspaces meet on the constant-output line, where all 32 labels contribute. **Middle:** for the first four thresholds $k = 1, \ldots, 4$, each dot gives the exact generator count $|\mathsf{M}_k|$ for one order type; horizontal offsets keep the dots distinct, and the annotations give each range. **Right:** the distribution of additional intersection flats when nonmaximal branches are retained. Five types gain none; the other 11 gain 2–9. The type-specific added-flat counts appear in Table 1.

annihilators recovers $\mathcal{B}_{\mathrm{lab}}(y)$; the resulting counts agree with the stratum formula of Theorem 9 on every computed stratum.

D.3 Intrinsic exceptional loci

We return to the maximal-branch incidence structure of Appendix D.1 and determine where its generic type changes within a fixed realization cell. The polynomial in Theorem 8 gives a sufficient exceptional set, but may include extra points: a selected minor can vanish while the full matrix retains its generic rank.

Fix the affine gauge of Appendix D.1 and write $\mathcal{R}^{\mathrm{g}}_\chi \subset \mathbb{R}^6$ for its realization set in the six free coordinates. Affine invariance recovers the full exceptional locus from this slice. Let $\mathcal{T}_\chi(X)$ record all branch dimensions, their labeled equality and containment relations, and the dimension-labeled maximal-branch intersection poset. Two types agree when these branch data agree and their posets admit a dimension-preserving isomorphism matching maximal branches with the same activation labels. Thus we compare incidence, rather than the positions of the subspaces in $\mathbb{R}^6$. Define the exceptional set by

$$E_\chi = \{X \in \mathcal{R}^{\mathrm{g}}_\chi : \mathcal{T}_\chi(X) \neq \mathcal{T}^{\mathrm{gen}}_\chi\},$$

where $\mathcal{T}^{\mathrm{gen}}_\chi$ is the generic type. Its intrinsic exceptional ideal is $\mathfrak{D}_\chi = I(\overline{E_\chi}^{\mathrm{Zar}})$, taken in $\mathbb{Q}[u, v, r, s, w, z]$ localized at the orientation determinants of the affine gauge. The ideal consists of the elements of this localized ring that vanish on E_χ; all inverted determinants are nonzero there.

Theorem 14 (Intrinsic exceptional loci and principal ideals)**.** *For each uniform planar six-point order type at width one, there are $2 \leq q_\chi \leq 20$ nonassociated irreducible factors $F_{\chi,1}, \ldots, F_{\chi,q_\chi}$ with*

$$E_\chi = \mathcal{R}^{\mathrm{g}}_\chi \cap \bigcup_{j=1}^{q_\chi} Z(F_{\chi,j}).$$

Here q_χ counts irreducible algebraic hypersurfaces. In the orientation-localized coordinate ring, the intrinsic radical is principal:

$$\mathfrak{D}_\chi = \bigcap_{j=1}^{q_\chi} (F_{\chi,j}) = \left(\prod_{j=1}^{q_\chi} F_{\chi,j} \right). \tag{20}$$

Proof by exact computation. We first specify a finite family of matrices that controls the type. Singleton and pairwise ranks for all 32 labels fix branch dimensions, equality, containment, and maximality by Eq. (18). When these ranks are generic, the maximal labels are fixed. Appendix C.2, with $p = 4$, then controls their intersection poset using at most four branches. We therefore use all singletons and pairs, together with triples and quadruples of the generic maximal branches. Constancy of these ranks implies constancy of the type.

For each cell, factor selected generic-rank minors after removing orientation factors. For each irreducible candidate f, recompute the whole-matrix rank over $\mathrm{Frac}(\mathbb{Q}[u, v, r, s, w, z]/(f))$, and retain the rank-drop factors whose zero sets meet the strict cell inequalities. Each retained factor has a smooth rational witness; excluded rank-drop factors have no real point in the cell, including singular points. Additional minors cover the feasible parts of the discarded selected-minor loci. Outside the retained hypersurfaces, these covers supply a nonzero generic-rank minor for every controlling matrix. Hence every exceptional point lies on a retained hypersurface.

For the converse, each retained factor f has an associated matching matrix C_J for a subset J of generic maximal labels whose rank drops modulo f. Thus every generic-rank minor of C_J is divisible by f, and the rank drop persists at every point of $Z(f)$ in the cell. If branch dimensions or labeled equality or containment relations change, the type is already exceptional. Otherwise the labels in J remain maximal and their branch ranks are unchanged. Eq. (17) then forces $\dim \bigcap_{a\in J} V_a$ to exceed its generic value, ruling out a dimension-preserving poset isomorphism matching the same labels. This proves the reverse inclusion, also at intersections of hypersurfaces. There is therefore no additional higher-codimension exceptional component outside their union. These are finite exact checks for each of the 16 order types: quotient-field ranks detect the drops, strict orientation inequalities determine real feasibility, and additional minors verify rank preservation on the remaining candidate loci.

Each retained factor is irreducible over $\mathbb{Q}$ and has a smooth rational point. It is therefore geometrically irreducible: otherwise its Galois-conjugate geometric components would all pass through that point, making it singular. Its smooth real patch in the open cell is Zariski dense in the hypersurface. In the orientation-localized polynomial ring, the factors are pairwise nonassociated prime elements. Their ideal intersection is generated by the squarefree product, giving Eq. (20). □

Thus a fixed order type determines the generic incidence structure, while additional polynomial relations describe its exceptional configurations. The next section makes these relations explicit in two cells.

D.4 Two explicit realization cells

We now make the exceptional loci explicit in two cells to show how the orientation inequalities restrict algebraic rank drops. All 20 rank-drop factors for Cell A meet its open realization cell. For Cell B, only 15 of 17 do: two are excluded by positive-bracket identities. This distinction determines which factors contribute to the intrinsic exceptional locus. For the explicit coordinates and bracket catalogs below, we index the six samples by $0, 1, 2, 3, 4, 5$.

Coordinates and branch structure. Cell A represents type ot6_09 in Table 1, has hull size five, and uses the affine gauge

$$x_0^A = (0, 0), \quad x_2^A = (1, 0), \quad x_3^A = (1, 1), \quad x_1^A = (u, v),\ x_4^A = (r, s),\ x_5^A = (w, z). \tag{21}$$

Its chirotope χ_A, in lexicographic triple order, is $---+++-+++++---$.

Cell B represents type ot6_01, with all six points on the convex hull. Its affine gauge is

$$x_0^B = (0, 0), \quad x_1^B = (1, 0), \quad x_2^B = (0, 1), \quad x_3^B = (u, v),\ x_4^B = (r, s),\ x_5^B = (w, z), \tag{22}$$

with chirotope $\chi_B = +++--+--+--+-++$ in the same triple order. Table 2 records the generic maximal-branch lattice sizes and the factor counts for both cells. In Cell $\ell \in \{A, B\}$, write $[ijk] = \det(\widetilde{x}_i^\ell, \widetilde{x}_j^\ell, \widetilde{x}_k^\ell)$. We use $E_A, \mathfrak{D}_A$ and $E_B, \mathfrak{D}_B$ for the exceptional sets and ideals defined in Appendix D.3 for these chirotopes.

Table 2: **Lattice invariants and exceptional factors for Cells A and B.** Both cells have 32 activation patterns. Cell A has hull size five and 20 intrinsic factors F_j^A; Cell B has hull size six and 15 factors F_j^B. **Upper block:** maximal-branch lattice sizes and successive factor counts. **Lower block:** both complete factor catalogs. Cell A factors 16–20 have bracket degree two; all other factors have degree one, assigning degree one to each bracket before fixing the gauge. Product degrees refer to the displayed representatives, with gauge degree computed in (u, v, r, s, w, z). A dash indicates that Cell B has no factor with that index.

Quantity	Cell A ($h = 5$)	Cell B ($h = 6$)
Generic maximal-branch lattice elements	134	117
Generic maximal branches	20	18
Controlling matrices	6,513	4,404
Selected-minor candidates	37	26
Factors causing a rank drop	20	17
Rank-drop factors meeting the cell	20	15
Product degree (bracket / gauge)	25/37	15/22

j	F_j^A	F_j^B	j	F_j^A	F_j^B
1	$[013]+[034]$	$[012]+[024]$	11	$[014]-[015]$	$[134]+[145]$
2	$[012]+[024]$	$[024]-[025]$	12	$[134]+[145]$	$[234]+[245]$
3	$[235]-[234]$	$[012]-[014]$	13	$[124]+[145]$	$[134]-[135]$
4	$[024]-[025]$	$[014]-[015]$	14	$[345]-[234]$	$[014]-[145]$
5	$[013]-[012]$	$[012]+[023]$	15	$[015]-[012]$	$[023]-[235]$
6	$[012]+[023]$	$[023]-[025]$	16	$[034][125]-[013][245]$	—
7	$[023]+[035]$	$[012]-[013]$	17	$[012][345]-[024][135]$	—
8	$[025]-[023]$	$[013]-[015]$	18	$[012][345]+[023][145]$	—
9	$[135]-[134]$	$[123]-[234]$	19	$[013][245]-[012][345]$	—
10	$[014]-[013]$	$[234]-[235]$	20	$[012][345]-[015][234]$	—

Proposition 15 (Explicit intrinsic factors for two cells)**.** *For the factors listed in Table 2, E_A is exactly the union of 20 irreducible hypersurfaces $Z(F_j^A)$ restricted to the cell. Likewise, E_B is exactly the union of 15 irreducible hypersurfaces $Z(F_j^B)$ restricted to the cell. Within each catalog the factors are pairwise nonassociated. Consequently, in the respective orientation-localized coordinate rings,*

$$\mathfrak{D}_A = \bigcap_{j=1}^{20} (F_j^A) = \left(\prod_{j=1}^{20} F_j^A \right), \qquad \mathfrak{D}_B = \bigcap_{j=1}^{15} (F_j^B) = \left(\prod_{j=1}^{15} F_j^B \right). \tag{23}$$

Proof by exact computation. Apply the controlling-matrix construction of Appendix D.3 in each affine gauge. Factoring selected generic-rank minors and testing the whole-matrix rank over the corresponding quotient fields leaves 20 rank-drop factors for Cell A and 17 for Cell B. Exact factorization and pairwise greatest common divisors verify the irreducibility and nonassociation of the catalog factors.

The zero set of each Cell A factor, and of 15 Cell B factors, has a smooth rational witness in the corresponding open cell. The remaining two Cell B factors are excluded by its orientation inequalities. Put $\langle ijk \rangle = \chi_B(ijk)[ijk]$, which is strictly positive throughout Cell B, and write these two factors as

$$C_1 = -rvz + rz + su + svw - sw - uz,$$
$$C_2 = ruz - rvw + su - sw - uz + vw.$$

Exact polynomial identities give

$$\begin{aligned}(-C_1)\langle 024\rangle &= \langle 012\rangle\langle 045\rangle\langle 234\rangle + \langle 014\rangle\langle 023\rangle\langle 245\rangle,\\ (-C_2)\langle 013\rangle &= \langle 012\rangle\langle 035\rangle\langle 134\rangle + \langle 014\rangle\langle 023\rangle\langle 135\rangle.\end{aligned} \tag{24}$$

Thus $C_1, C_2 < 0$ throughout Cell B, so their zero sets cannot contribute to its intrinsic exceptional locus. This leaves precisely the 20 and 15 factors listed in Table 2.

The completeness and ideal arguments of Theorem 14, applied to these factors and smooth rational witnesses, give the stated exceptional sets and Eq. (23). □

The catalogs in Table 2 are given up to nonzero rational scalars, orientation units, and bracket identities.

At generic points of the 15 exceptional hypersurfaces in Cell B, the maximal-branch lattice has 115 elements on nine components and 111 on the other six, compared with 117 away from the exceptional locus. Intersections of components can produce further changes, illustrated next in Cell A.

D.5 Further stratification of exceptional configurations

The hypersurface union determines where the generic incidence type fails. Its intersections distinguish further lattice types within the exceptional locus. In Cell A, we examine how these special input configurations change higher-order branch intersections in output space, then follow the changes along a path within the exceptional locus.

Intersecting exceptional hypersurfaces. Take the three Cell A factors $J = F_5^A$, $G = F_{13}^A$, and $H = F_{18}^A$ from Table 2. Bracket identities give the equivalent forms

$$J = [013] - [012], \qquad G = [125] + [245], \qquad H = [024][135] - [025][134].$$

For each of the eight choices of which factors vanish, we solve the vanishing equations on a rational coordinate chart, with the remaining factors among J, G, H nonzero. Exact closure over each chart's function field gives 134 maximal-branch lattice elements when none vanish, 132 when one vanishes, 130 when two vanish, and 128 when all three vanish. Each chart has a rational witness in Cell A attaining its generic lattice size.

The same 20 labels represent distinct four-dimensional maximal branches in all eight computations. At the rational witnesses, an exact comparison with labels fixed gives identical intersection dimensions for every pair of these branches, despite the different lattice sizes. Thus even the labeled pairwise dimensions do not determine the maximal-branch lattice; the distinction requires intersections of three or more branches.

Exceptional parameters along a path. To follow a further specialization within $Z(J) \cap Z(G)$, take the path $X(t)$, $0 \le t \le 1$, with

$$x_1^A(t) = (0, 1 + t), \qquad x_5^A(t) = (3, 2 - t),$$

and x_0^A, x_2^A, x_3^A as in Eq. (21), $x_4^A = (2, 3)$. All orientation determinants retain the signs of Cell A. Along this path, $J \equiv G \equiv 0$ and $H = t^2 + 3t - 1$. The ambient exceptional ideal therefore pulls back to zero and cannot distinguish changes along the path. We instead compare each parameter with the lattice generic along $X(t)$.

Over $\mathbb{Q}(t)$, the lattice has 20 maximal branches and 130 elements. We factor the generic-rank minors and test whether each factor forces a rank drop in the corresponding matrix. This yields the squarefree exceptional polynomial

$$\Delta_{X,\mathrm{rel}}(t) = t(t-1)(t^2 + 3t - 1). \tag{25}$$

Its zeros in $[0, 1]$ are exactly the parameters where the lattice differs from this generic type. Their lattice sizes are

$$t = 0 : 122, \qquad t_* = \frac{\sqrt{13} - 3}{2} : 128, \qquad t = 1 : 128.$$

Every other parameter in $[0, 1]$, including $t = 1/2$, has the path-generic 130-element lattice. The endpoints lie on additional exceptional hypersurfaces from Table 2, while the interior irrational root

t_* marks the crossing of $Z(H)$ within $Z(J) \cap Z(G)$. Sampling only $\{0, \frac{1}{2}, 1\}$ would detect the exceptional endpoints but miss this additional change.

The endpoint $t = 1$ also provides the comparison used in Section 4.3. Relative to $t = 1/2$, all 32 branch dimensions and all labeled pairwise intersection dimensions are unchanged. For the three labels 010111, 101001, and 101101, however, the matching rank drops from 11 to 10. Their branch ranks remain four, so Eq. (17) gives a triple-intersection dimension increase from one to two, despite unchanged pairwise data.

E One-wall NTK discontinuity and spectral response

At a ghost wall, the adjacent branches have the same output closure, while the wall-image closure has one fewer dimension. We ask whether equal branch spaces also imply the same local output dynamics. The empirical neural tangent kernel (NTK) determines how parameter gradient flow acts on the output residual (Jacot et al., 2018). We compute its exact change across a single wall and relate its range to the branch spaces. This connects the wall geometry of Section 3 to local optimization dynamics; a short spectral example then illustrates how the response varies with the data and parameter normalization.

E.1 NTK jumps and local output dynamics

Fix a true single-wall parameter θ_0 between labels a and b, where only neuron j on sample i has zero preactivation. Let a label the negative side and b the positive side. On side $S \in \{-, +\}$, write $s^{(k),S}$ for neuron k's activation pattern, so $s_i^{(j),-} = 0$ and $s_i^{(j),+} = 1$. Evaluating all parameters at θ_0, set $A_k^S = D_{s^{(k),S}} \widetilde{X}$ and $r_k^S = A_k^S u_k$. Ordering parameters by $(\beta_0, \alpha_1, u_1, \ldots, \alpha_m, u_m)$ gives the one-sided Jacobian

$$J_S = \left[\mathbf{1}_n, r_1^S, \alpha_1 A_1^S, \ldots, r_m^S, \alpha_m A_m^S\right]. \tag{26}$$

Write $K_S^0 = J_S J_S^\top$ for the one-sided empirical NTK with respect to the Euclidean metric on these parameters. The crossing neuron has the same activation values $r_j^- = r_j^+$ at the wall, so only its affine-parameter block changes. Let $Q = \widetilde{X}\widetilde{X}^\top$, $v = D_{s^{(j),-}} Q e_i$, and $q_{ii} = \|\widetilde{x}_i\|^2$.

Proposition 16 (Exact one-wall NTK update)**.** *With the above side convention,*

$$K_+^0 - K_-^0 = \alpha_j^2 \left(e_i v^\top + v e_i^\top + q_{ii} e_i e_i^\top\right), \qquad \operatorname{rank}(K_+^0 - K_-^0) \le 2. \tag{27}$$

The update is nonzero if $\alpha_j \neq 0$; reversing the crossing direction negates it.

Proof. From Eq. (26), the one-sided kernel is

$$K_S^0 = \mathbf{1}_n \mathbf{1}_n^\top + \sum_{k=1}^m \left[r_k^S (r_k^S)^\top + \alpha_k^2 A_k^S (A_k^S)^\top\right].$$

At the wall, $\widetilde{x}_i^\top u_j = 0$, so $r_j^- = r_j^+$ and all output-weight columns agree. Since $A_j^+ = A_j^- + e_i \widetilde{x}_i^\top$, expanding the remaining Gram difference gives

$$\begin{aligned} A_j^+ (A_j^+)^\top - A_j^- (A_j^-)^\top &= e_i (\widetilde{x}_i^\top (A_j^-)^\top) + (A_j^- \widetilde{x}_i) e_i^\top + \|\widetilde{x}_i\|^2 e_i e_i^\top \\ &= e_i v^\top + v e_i^\top + q_{ii} e_i e_i^\top. \end{aligned}$$

Its range lies in $\operatorname{span}\{e_i, v\}$, proving rank at most two. Since $s_i^{(j),-} = 0$, we have $v_i = 0$, while the constant input coordinate gives $q_{ii} \ge 1$. Thus $\alpha_j \neq 0$ implies $(K_+^0 - K_-^0)_{ii} = \alpha_j^2 q_{ii} > 0$. □

Branch spaces and local dynamics. If every output weight α_k is nonzero, the columns $\alpha_k A_k^S$ span $\operatorname{im} A_k^S$, while $r_k^S = A_k^S u_k$ adds no direction. Hence Eq. (26) gives

$$\operatorname{im} K_-^0 = \operatorname{im} J_- = V_a, \qquad \operatorname{im} K_+^0 = \operatorname{im} J_+ = V_b.$$

At a ghost wall, Theorem 3 gives $V_a = V_b \supsetneq G_{a,b}$. Under the same nonzero-weight assumption, the two kernels therefore have the same range and rank, but act differently on this common subspace by Proposition 16.

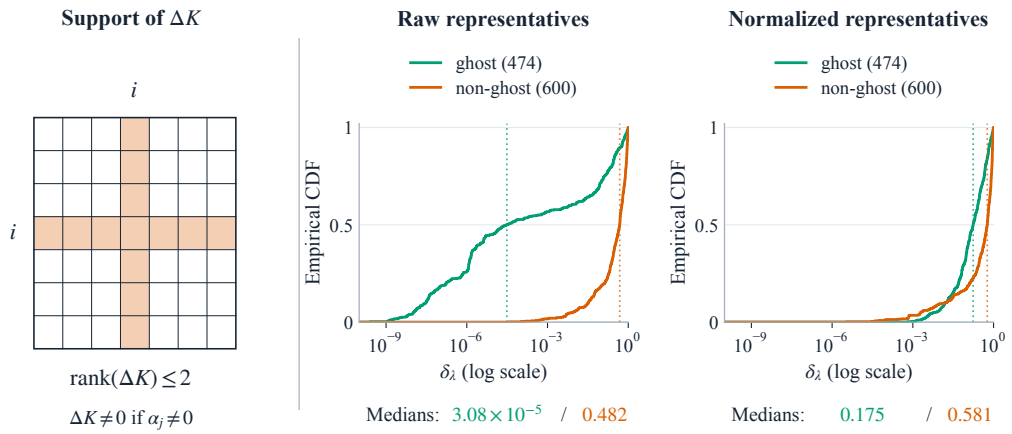

Figure 4: **NTK jumps and normalization-dependent spectral response. Left:** schematic support of the one-wall update $\Delta K = K_+^0 - K_-^0$; entries outside row and column i vanish. **Middle and right:** empirical cumulative distribution functions (CDFs) of δ_λ under the raw and normalized conventions, with identical logarithmic axes. Green denotes ghost observations and orange denotes the non-ghost reference set. Each convention has 474 ghost and 600 reference observations from three output-weight choices per wall. Dotted lines mark the medians reported below each plot.

For unnormalized squared loss $\frac{1}{2}\|E_X(\theta) - y^\star\|^2$, the output-velocity fields induced by parameter gradient flow have one-sided limits $\dot{y}_\pm = -K_\pm^0(E_X(\theta_0) - y^\star)$ at θ_0. Hence

$$\dot{y}_+ - \dot{y}_- = -(K_+^0 - K_-^0)\big(E_X(\theta_0) - y^\star\big). \tag{28}$$

The residual therefore determines whether the kernel jump changes the output velocity; at exact fit both velocities vanish. Kernel jumps also occur without ghosts: at width one, Theorem 3 excludes ghosts, yet the update is nonzero whenever $\alpha_j \neq 0$.

The empirical NTK uses the Euclidean metric of parameter space and is not invariant under the ReLU gauge $(\alpha_j, u_j) \mapsto (\alpha_j/c, cu_j)$, $c > 0$, which preserves the represented function. Thus equal branch spaces do not fix the kernel spectrum. Parameterization also affects activation-boundary dynamics (Chen & Ge, 2024). This dependence motivates the following numerical comparison, which also varies the input normalization and wall representatives.

E.2 Spectral response under input and parameter normalization

We illustrate this parameter dependence on 158 ghost walls, with 200 non-ghost walls as a reference. In the deterministic census order, we take the first 50 non-ghost walls for each of $(d, m) = (1, 1)$ and $(1, 2)$, and the first 100 for $(d, m) = (2, 2)$. Each wall is evaluated at crossing-neuron output weights $\alpha_j \in \{0.5, 1, 2\}$; all remaining output weights are fixed and nonzero. The reference set spans different input dimensions and widths and includes walls where the kernel rank changes. The raw convention uses the original linear-program wall representatives. The normalized convention centers each input coordinate and rescales it to unit root-mean-square value, recomputes the representatives, and normalizes each augmented hidden-unit vector to unit Euclidean norm. The invertible affine input normalization preserves the branch spaces associated with each activation label. The wall labels and output-weight choices are shared; the transformed inputs and recomputed representatives can change the represented function.

For the smallest positive one-sided eigenvalues $\lambda_\pm = \lambda_{\min}^+(K_\pm^0)$, define the relative jump $\delta_\lambda = |\lambda_+ - \lambda_-|/\max\{\lambda_+, \lambda_-\}$. The numerical Jacobian rank determines how many kernel eigenvalues are retained as positive, with the numerical convention specified in Appendix G.1. In Figure 4, normalization markedly increases the median jump for the same ghost-wall labels. Thus equality of the adjacent branch spaces coexists with spectral responses that depend on the data and parameter normalization.

F Discussion: implications for learning

The geometric results distinguish two aspects of fitting a finite batch: moving between parameter settings with the same predictions, and recovering activation labels from those predictions. We discuss how these questions change when small prediction errors are allowed. Wall-image distances give lower bounds on the training loss along parameter paths, while distances to alternative branches control the stability of activation explanations.

F.1 Ghost directions and loss barriers

Mode connectivity asks whether fitted networks can be joined through low-loss parameters (Garipov et al., 2018; Nguyen, 2019). Exact-fit connectivity is its zero-loss case. For squared loss, write

$$L_y(\theta) = \frac{1}{2n}\|E_X(\theta) - y\|_2^2.$$

At a point in the closure of a shared facet, the output belongs to $G_{a,b}$. Crossing that wall therefore costs at least $\operatorname{dist}(y, G_{a,b})^2/(2n)$. To account for paths that avoid a particular wall, consider the adjacency graph of all strict ordered activation cells, assigning each edge the weight

$$c_{a,b}(y) = \frac{\operatorname{dist}(y, G_{a,b})^2}{2n}.$$

Unlike $\Gamma_G(y)$, this graph also includes cells whose branch closures do not contain y.

Corollary 17 (A global loss-barrier bound)**.** *For distinct inputs, let $\theta_0, \theta_1 \in \mathcal{F}_y$ lie in distinct strict ordered cells C_{a_0}, C_{a_1}. Then*

$$\inf_{\theta:\theta_0 \rightsquigarrow \theta_1} \max_{t\in[0,1]} L_y(\theta(t)) \;\geq\; \min_{P:a_0 \rightsquigarrow a_1} \max_{(a,b)\in P} c_{a,b}(y),$$

where the infimum is over continuous parameter paths on $[0,1]$ and P ranges over paths in the full activation adjacency graph. If a_0, a_1 are disconnected in $\Gamma_G(y)$, the right-hand side is strictly positive.

Proof. Let a continuous parameter path have maximum loss h. Its endpoints belong to the same path component of the semialgebraic sublevel set $\{L_y \leq h\}$, so they admit a semialgebraic path within that set. Apply the incident-chamber construction in the proof of Theorem 4. Each edge a, b in the resulting walk has a point θ_* on this path in the closure of its shared facet. Hence $E_X(\theta_*) \in G_{a,b}$ and $c_{a,b}(y) \leq L_y(\theta_*) \leq h$. The minimum over graph paths is therefore at most h; taking the infimum over parameter paths proves the bound. A zero-weight edge has $y \in G_{a,b} \subseteq V_a \cap V_b$, so a zero-weight path would also be a path in $\Gamma_G(y)$. Finiteness of the graph gives strict positivity when such a path is absent. □

An exact barrier in the five-point example. For the width-two witnesses θ_a, θ_b in Appendix B.1, the minimum possible maximum loss is exactly $1/80$. The cell $a = (11100, 00111)$ has four adjacent walls. Direct calculation gives squared distances $1/6, 1/6, 1/8, 1/8$ from $y = (5, 3, 2, 3, 5)^\top$ to their image closures. Every graph path leaving a therefore has an edge of weight at least $1/80$. Corollary 17 gives the lower bound, including for parameter paths that cross several walls simultaneously.

For attainment, normalize both witnesses to output weights $(1, 1)$ by positive homogeneity, preserving their predictions. In the parameter ordering of Appendix B.1, the common wall contains

$$\theta_* = \big(-3/8;\, 1, (-7/4, 21/4);\, 1, (19/8, -4)\big).$$

With $q = (-1, 1, 1, -2, 1)^\top$, direct evaluation gives $E_X(\theta_*) = y + q/8$, the orthogonal projection of y onto $G_{a,b}$, and $L_y(\theta_*) = 1/80$. Join each normalized witness to θ_* by a straight segment in its closed activation cell, keeping the output weights fixed. Evaluation is affine on each segment, so its output traces the segment between y and $y + q/8$ and its loss never exceeds $1/80$. Concatenating with the output-preserving normalization paths attains the bound. Thus no path connects these witnesses while keeping the loss below $1/80$, even when arbitrary routes through parameter space are allowed.

F.2 Activation explanations and stability

Small perturbations of the target outputs and fitting errors may allow activation labels that are incompatible with the original target. We seek a margin within which these labels remain excluded. The weighted lattice separates two sources of algebraic label ambiguity. For a maximal branch $V \in \mathcal{M}_X$, almost every $y \in V$, with respect to Lebesgue measure on V, satisfies $b_{\mathrm{lab}}(y) = w(V)$. To see this, exclude the finitely many intersections $V \cap V'$ with distinct branches V'. Maximality makes each a proper subspace of V; outside their union, only labels with branch equal to V contain y. Thus branch weights describe ambiguity present at generic outputs, while the intersection lattice locates the additional explanations at special outputs.

For a fixed target y, exclusion is controlled by the distances to branches that do not contain it. Define

$$\delta(y) = \min_{a \in \mathcal{A}_X^{(m)} : y \notin V_a} \operatorname{dist}(y, V_a),$$

with $\delta(y) = +\infty$ when the index set is empty. Otherwise $\delta(y) > 0$, since there are finitely many closed branch subspaces.

Corollary 18 (Stability of activation explanations)**.** *For a finite batch X, let $\widetilde{y} = y + e$ with $\|e\|_2 \le \eta$. If a network in a strict activation cell with multiset label a satisfies $\|E_X(\theta) - \widetilde{y}\|_2 \le r$ and $\eta + r < \delta(y)$, then $a \in \mathcal{B}_{\mathrm{lab}}(y)$. In particular, if $\mathcal{B}_{\mathrm{lab}}(y)$ is a singleton, every such network has the same activation multiset.*

Proof. If $y \notin V_a$, then $E_X(\theta) \in V_a$ implies

$$\delta(y) \le \operatorname{dist}(y, V_a) \le \|y - E_X(\theta)\|_2 \le \eta + r,$$

contradicting the assumed margin. □

For a maximal branch of weight one, the generic-output observation gives the singleton case. The margin then quantifies how much combined target perturbation and fitting error can be tolerated while excluding other activation multisets. This identifies an activation explanation on the finite batch; parameters within that explanation can still vary. The lattice determines which branches compete, while a numerical margin also depends on their positions in output space.

G Computational details and future directions

G.1 Computational methods and examples

Computations use integer or rational input configurations. Activation patterns are enumerated as strict affine-halfspace sign patterns; the signed-fiber computations test every binary mask by the normalized linear inequalities of Appendix B.3, including for repeated inputs. Branch spaces are represented by row-reduced bases of $B_a^\top$. Ranks and intersections are computed in exact arithmetic using SymPy; intersections use dual normal spaces or the matching matrix in Eq. (16). Linear programs check activation and signed-fiber feasibility; rational witnesses and Farkas certificates verify the decisions as in Appendix B.3. The code provides solver-free replay of fiber certificates, exact constructions of both types of intersection lattice, and SMT-based replay of nonlinear real-feasibility queries.

NTK spectral calculations use double precision. For an $n \times q$ Jacobian J, its numerical rank counts singular values above $\sigma_{\max}(J) \max\{n, q\}\epsilon_{\mathrm{mach}}$, where $\epsilon_{\mathrm{mach}} = 2^{-52}$, following the default NumPy threshold. If this rank is r_J, the r_J largest eigenvalues of $JJ^\top$ are retained; the smallest of these defines $\lambda_{\min}^{+}$.

Table 3 collects examples of wall geometry, fiber connectivity, and lattice structure. The complete generic six-point classification and exceptional loci are given in Appendix D.

Computational resources. The recorded environment is Linux x86-64, Python 3.9.25, NumPy 1.26.4, SciPy 1.13.1, SymPy 1.13.3, and JAX/JAXlib 0.4.30. Nonlinear feasibility queries were replayed with Z3 5.0.0. Runs used a shared dual-socket AMD EPYC 9554 CPU host with four parallel workers.

Table 3: **Summary of exact geometric computations.** Multiplicity uses six planar inputs and width one. The lattice examples use 18 labeled six-point cells and six labeled seven-point cells, selected to exhibit within-cell variation. Fiber components are counted in the ordered parameter space.

Geometric object	Instances	Result
Single walls	6,008 wall cases	158 walls with ghost directions
Two-sample fibers	24 width-one examples	8 constant targets: contractible fibers; 16 nonconstant targets: two components
Signed fiber graph	11 rational examples	Empty fibers; nonempty fibers have 1, 2, 6, or 8 path components
Lattice variation	24 within-cell segments; endpoints and midpoints	Distinct lattice types at the two endpoints of every segment
Label multiplicity	16 generic six-point types	Retaining nonmaximal branches adds strata in 11 types

G.2 Future directions and broader implications

Future directions. An immediate direction is to develop more efficient lattice and fiber-graph algorithms for larger batches and widths. Extending the geometric analysis to deeper networks leads to polynomially parameterized image closures, while multiple outputs introduce rank constraints through shared hidden features. Connecting these geometric descriptions to training trajectories is another direction for future work.

Broader impact. Understanding how different network representations produce the same predictions helps clarify what finite training data can reveal about activation structure and parameter connectivity. The geometric distinctions developed here can inform the interpretation of such ambiguity in learned models. The explicit constructions and exact certificates also provide reproducible reference cases for connecting mathematical analysis with empirical studies of neural-network behavior.