# BOUNDARY ROBUSTNESS IS NONLOCAL: A KIRCHHOFF LAW FOR BALANCED ASSIGNMENT

**Anonymous authors**
Paper under double-blind review

## ABSTRACT

Local margins do not determine boundary robustness once a partition must give each region a fixed share of the data, as balanced Sinkhorn self-labelling, balanced token-to-expert matching and balanced transductive inference all do. We build two such partitions whose interface is identical in every local quantity, to machine precision in margin, density and conductance, yet which respond to the same shift by amounts differing by $10.6\times$; the ratio can be made arbitrarily large. The mechanism is exactly electrical: differentiating the mass constraints, the known implicit derivative of the semi-discrete dual, turns a shift into a current injection on the region-adjacency graph, so the potentials re-equilibrate through a Kirchhoff system and each interface moves by the voltage drop across its own edge over its local margin. Read one interface at a time this gives a per-unit-shift susceptibility $\sqrt{R_{ij}}/\gamma_{ij}$, an effective resistance over the local margin, whose network term is tight, and a sharp impossibility: local data does give a valid certificate for a current across the interface, the isolated-edge quantity $1/(h_{ij}\gamma_{ij})$ with $h_{ij}$ the interface conductance, but no local functional can tell whether it is informative and no local point predictor is uniformly accurate. The same sensitivity and the same impossibility hold for entropic Sinkhorn assignments, whose conductances are the closed-form co-assignment mass. Exact re-solution confirms the law through three perturbation channels and a constraint-off knockout removes the effect. On a self-labelling head, natural-image partitions and annotated tumour sections the law's trapezoidal form orders the interfaces that move at Spearman $\rho \geq 0.957$ per unit and beats the local current-conditioned competitor on every unit, while calibrated magnitudes degrade outside the linear regime, the shift-agnostic susceptibility ranks no better than a local score before the shift is seen, and warm-started Sinkhorn iteration wins on magnitude at higher cost.

## 1 INTRODUCTION

A mass constraint fixes how much of the data each region of a learned partition must receive. It prevents representation collapse and idle capacity, and it is everywhere in modern representation learning: balanced Sinkhorn self-labelling (Asano et al., 2020; Caron et al., 2020; Piran et al., 2024), Sinkhorn label allocation (Tai et al., 2021), balanced token-to-expert matching (Lewis et al., 2021), balanced transductive inference (Veilleux et al., 2021). We find that this one bookkeeping device changes what boundary robustness *is*. The standard way to judge whether a boundary will stay put is local: it is thought stable when the competing scores separate quickly as one leaves it, that is, when the local margin $\gamma$ is large. Margin bounds (Bartlett et al., 2017), margin-distribution predictors (Jiang et al., 2019) and randomised smoothing (Cohen et al., 2019) read the neighbourhood of the decision surface and nothing else. Under a mass constraint that neighbourhood is not enough, and not because the bounds are loose.

Figure 1 states the problem. Two mass-constrained partitions share every local quantity at $B_{12}$ to machine precision and differ only in the connectivity of regions *not* adjacent to it, yet under the same shift one moves $10.6$ times further: any functional of that local data returns one number and is wrong for at least one. We call this *frontier nonlocality*.

The mechanism is exactly electrical. Under mass constraints a partition is a semi-discrete optimal-transport tessellation whose dual potentials $w_k$ play the role of voltages (Aurenhammer et al., 1998;



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.

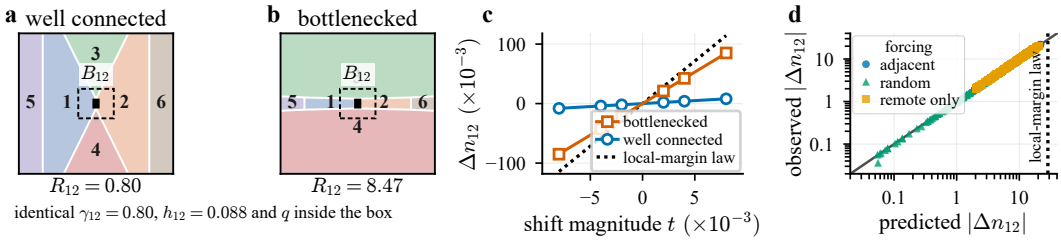


Figure 1: **Frontier nonlocality: the same local data, two different responses.** (a, b) Two mass-constrained partitions whose *local* data at $B_{12}$ is identical by construction, same $c_1, c_2$, segment and density, hence $\gamma_{12} = 0.80$ and $h_{12} = 0.088$ in both; only the remote programs differ, changing $R_{12}$ by $10.6\times$. (c) Under the identical shift they move by amounts differing by $10.6\times$, while a local-margin law must return one value (dotted). (d) Across 180 configurations, over which $R_{12}$ spans $10.9\times$, Theorem 2 predicts the re-solved displacement with $R^2 = 0.999998$, including for forcings injected only at cells that never touch $B_{12}$; axes in units of $10^{-3}$. *The local geometry is not wrong, it is insufficient: the missing variable is remote connectivity.*

Peyré & Cuturi, 2019); differentiating the constraints produces a weighted graph Laplacian $L_B$ on the region-adjacency graph, a classical identity used inside Newton solvers (Lévy, 2015; De Gournay et al., 2019; Kitagawa et al., 2019), and a shift enters as a *current injection* $b$ with $\mathbf{1}^\top b = 0$, which is conservation of mass. Read as a sensitivity statement this gives a first-order interface-motion law (Theorem 2) and, by Cauchy–Schwarz in the Laplacian energy, a bound whose network term is tight and in which the classical susceptibility $1/\gamma_{ij}$ becomes $\sqrt{R_{ij}}/\gamma_{ij}$ (Theorem 3). Effective resistance depends on every path between the two regions, so it is invisible to any neighbourhood of the interface. The impossibility this implies is sharper than "no local bound exists", which would be false: for a current injected across the interface, Rayleigh monotonicity turns the two-cell quantity $1/(h_{ij}\gamma_{ij})$ into a valid local certificate (Proposition 5). What local data cannot do is tell whether that certificate means anything, because two locally indistinguishable partitions respond by arbitrarily different amounts (Theorem 6).

**Assumptions and scope.** Neither the Laplacian identity nor its differentiation is new and neither is claimed here; what is new is reading that derivative one interface at a time, and what follows from doing so (Section 2 gives the credit in full). Everything below concerns partitions defined by Equation (1) under a mass constraint, exactly or entropically, satisfying Assumption 1, and the constraint must be active on the data whose boundaries matter: that holds for transductive balanced inference and for training-time target assignment but *not* at test time for systems that balance only while training, since BASE layers route greedily at inference (Lewis et al., 2021) and SwAV and SeLa use the balanced step to make targets. Equation (4) is first order, so it predicts magnitudes only inside the regime of Theorem 8; on real data, where corruption families are not $C^1$ in severity, we use its finite-step form and its reliable output is direction and ordering, while Equation (5) bounds that first-order term and not the realised displacement (Section 5.3).

**Contributions.**

- **Frontier nonlocality, and an impossibility.** Mass constraints make boundary sensitivity global: locally indistinguishable interfaces respond to the same shift by an arbitrary factor, so a local certificate stays valid but can be arbitrarily uninformative and no local point predictor is uniformly accurate (Proposition 5 and Theorems 6 and 7; realised for hard interfaces in Figure 1 and for entropic ones in Section 5.1).
- **The Boundary Circuit Law.** The known implicit derivative, read per interface, is a Kirchhoff system whose voltage drops give interface velocity, with a bound whose network term is tight, a shift-agnostic susceptibility $\Phi_{ij} = \sqrt{R_{ij}}/\gamma_{ij}$, and an entropic analogue whose conductances are the closed-form co-assignment mass of a Sinkhorn head (Theorems 2, 3 and 7).
- **Mechanism attribution, and an honest envelope.** A constraint-off knockout removes the effect ($R^2 = 1.0000$ against $-2.2250$), the circuit ranks ahead of eight alternative predictors and the two strongest of them lose paired within-partition tests; on three real systems and a *trained* head

with a drifting encoder ranks transfer far outside the linear regime while calibrated magnitudes do not, and the circuit matches $k = 50$ warm-started Sinkhorn updates on ranking at a fraction of their cost while losing to them on magnitude (Tables 1, 2, 4 and 5).

## 2 Related work

**Semi-discrete transport, and what we add.** Aurenhammer et al. (1998) give the concave dual and its gradient; that its Hessian is the Laplacian of the weighted dual graph, with the regularity and damped Newton solver this powers, is Kitagawa et al. (2019), the same sparse Hessian appearing in Lévy (2015); see (Mérigot, 2011; Bourne et al., 2020; Santambrogio, 2015; Peyré & Cuturi, 2019) for background. **We claim novelty for neither the identity nor its differentiation**: De Gournay et al. (2019) differentiate the optimal weights by the implicit function theorem, the computation in Theorem 2, with the entropic counterpart behind implicit Sinkhorn differentiation (Luise et al., 2018; Eisenberger et al., 2022). Prior stability results describe the perturbed solution as a whole, through a global $L^2$ error (Delalande & Mérigot, 2023), cell stability (Bansil & Kitagawa, 2022) or a breakdown point (Paindaveine & Passeggeri, 2024), and Robertson et al. (2024) obtain a Laplacian pseudo-inverse for linearised transport on a *given* graph. **New is the per-interface reading**: a resistance attached to one boundary, the proof that local data cannot say whether its own certificate is informative, and the experiments isolating it.

**Balanced assignment, and where the constraint is active.** Mass constraints entered representation learning as a remedy for collapse (Asano et al., 2020; Caron et al., 2020; Tai et al., 2021; Lewis et al., 2021; Piran et al., 2024). Two qualifications matter. Several such systems balance only during training, BASE routing greedily at inference and SwAV and SeLa using the balanced step for targets, so our question bites in transductive balanced inference, whose uniform-marginal assumption is itself fragile (Veilleux et al., 2021), and in training-time target or route stability (Dai et al., 2022). And neighbouring mechanisms balance only approximately, so we do not claim them: DINO (Caron et al., 2021) centres and sharpens, Switch (Fedus et al., 2022) uses a capacity factor, vector quantisation (van den Oord et al., 2017) prescribes no codeword mass, constrained $k$-means (Bradley et al., 2000) imposes minimum sizes. These works ask what the constraint buys; **we ask what it costs wherever it is active**.

**Margins, shift, and effective resistance.** Robustness analyses of decision boundaries are local by construction (Bartlett et al., 2017; Jiang et al., 2019; Cohen et al., 2019) and certify a fixed classifier against input perturbations, a different question; we use them only to name the habit of reading stability off the neighbourhood. Shift-performance predictors (Garg et al., 2022; Miller et al., 2021) work on aggregate accuracy, label-shift correction (Saerens et al., 2002; Alexandari et al., 2020) adjusts class priors, which in our language moves $m$ and injects a current but treats classes independently, transport also *models* shift (Courty et al., 2017; Naumann et al., 2026), and corruption benchmarks (Hendrycks & Dietterich, 2019; Koh et al., 2021) supply our perturbation families. Resistance distance (Klein & Randić, 1993; Doyle & Snell, 1984; Chandra et al., 1996) serves graph sparsification (Spielman & Srivastava, 2011) and explains over-squashing (Topping et al., 2022; Black et al., 2023; Di Giovanni et al., 2023; Attali et al., 2024), but there the graph is *given*; **ours is induced by the learned partition** and resistance measures the displacement of a boundary. Transport is likewise standard in spatial transcriptomics (Zeira et al., 2022; Cang et al., 2023) alongside domain segmentation (Singhal et al., 2024; Traag et al., 2019); we claim no priority there and use tumour interfaces (Andersson et al., 2021) to test a general theory.

## 3 Mass-constrained partitions and their circuit

**Setup and notation.** Let $\mathcal{X} \subset \mathbb{R}^d$ be compact with a probability density $q$, and let $c_1, \dots, c_K : \mathcal{X} \to \mathbb{R}$ be $C^1$ cost functions, one per *program* (prototype, class, code word, expert). A vector of *dual potentials* $w \in \mathbb{R}^K$ defines

$$\mathcal{C}_k(w) = \big\{x \in \mathcal{X} : \; c_k(x) - w_k \le c_j(x) - w_j \;\; \forall j\big\}, \qquad M_k(w) = \int_{\mathcal{C}_k(w)} q(x)\, dx. \tag{1}$$

Given target masses $m \in \Delta^{K-1}$, the *mass-constrained partition* solves $M(w) = m$, exactly the semi-discrete optimal-transport problem, in which $w$ maximises the concave dual $\sum_k m_k w_k +$

$\int_{\mathcal{X}} \min_k\{c_k(x) - w_k\}q(x)\,dx$ with gradient $m - M(w)$ (Aurenhammer et al., 1998; Peyré & Cuturi, 2019); it depends on $w$ only through differences, so we fix the gauge $\mathbf{1}^\top w = 0$. Writing $F_{ij}(x,t) = c_i(x,t) - c_j(x,t) - w_i(t) + w_j(t)$, the *interface* between $i$ and $j$ is its zero set

$$B_{ij} = \{x \in \mathcal{X} : F_{ij}(x) = 0\} \cap \partial\mathcal{C}_i \cap \partial\mathcal{C}_j, \qquad \gamma_{ij} = \inf_{x \in B_{ij}} \|\nabla_x(c_i - c_j)(x)\|, \tag{2}$$

the second quantity being the *local margin*. Write $V_{ij}(x)$ for the normal velocity of $B_{ij}$ at $x$ pointing from $\mathcal{C}_i$ into $\mathcal{C}_j$, so $V_{ij} > 0$ means $\mathcal{C}_i$ expands; the scalar robustness observable is the sup norm $\overline{V}_{ij} := \|V_{ij}\|_{L^\infty(B_{ij})}$, paired with the infimum margin in $\gamma_{ij}$. For affine cost differences, the case of every experiment here, $\|\nabla_x(c_i - c_j)\|$ is constant on $B_{ij}$ and $\overline{V}_{ij}$ is the common pointwise velocity; Appendix A collects all symbols. The circuit applies to a mechanism exactly when the Jacobian of its constraint residual with respect to the balancing duals is a connected symmetric weighted Laplacian, and Section 2 says which learning systems qualify.

**Assumption 1** (Regularity)**.** Over the interval considered: (A1) every cell has mass bounded below, $\min_k M_k \geq m_{\min} > 0$; (A2) each non-empty interface is $\mathcal{H}^{d-1}$-rectifiable with $\gamma_{ij} > 0$, higher-order junctions have zero $\mathcal{H}^{d-1}$ measure, $q$ is continuous and bounded near $\bigcup_{ij} B_{ij}$, and the region-adjacency graph $G_B$ is connected; (A3) $t \mapsto (c^t, q^t, m^t)$ is $C^1$ with $q^t$ a probability density and $\mathbf{1}^\top m^t = 1$; (A4) the conductances $h_{ij}$ of Equation (3) are *jointly continuous* in $(w,t)$ near $(w(0),0)$.

(A4) makes the mass Jacobian continuous and not merely defined, which is what the implicit function theorem needs in Theorem 2; it holds for the $C^{2,\alpha}$-regular problems of Kitagawa et al. (2019) and whenever the cost differences are affine and $q$ is continuous, as does connectivity in (A2) on a convex $\mathcal{X}$ (also checked numerically, Appendix C).

**Lemma 1** (The mass Jacobian is a graph Laplacian; known)**.** *Under Assumption 1, $M$ is differentiable at $w$ and*

$$\frac{\partial M_k}{\partial w_j} = -h_{kj}\ (k \neq j), \quad \frac{\partial M_k}{\partial w_k} = \sum_{j \neq k} h_{kj}, \quad h_{ij} = \int_{B_{ij}} \frac{q}{\|\nabla_x(c_i - c_j)\|}\, d\mathcal{H}^{d-1} \geq 0. \tag{3}$$

*Equivalently $\partial M/\partial w = L_B$, the weighted Laplacian of the* region-adjacency graph $G_B$ *on $\{1,\dots,K\}$ with edge conductances $h_{ij}$.*

Lemma 1 is not ours and neither is differentiating it (Section 2 gives the credit in full); what is new is what the derivative says about *one interface at a time*: Theorem 3, the impossibility of Theorem 6, and the experiments that isolate them.

## 4 The Boundary Circuit Law

Differentiating the constraint $M(w(t),t) = m(t)$ once gives the whole theory. This section reads that derivative one interface at a time: a Kirchhoff system, a bound whose network term is tight, an impossibility, and the entropic and finite-shift versions of both.

**Theorem 2** (Boundary Circuit Law; the implicit derivative, read per interface)**.** *Let $t \mapsto (c^t, q^t, m^t)$ satisfy Assumption 1, let $w(t)$ solve $M(w(t),t) = m(t)$ and define the* injected current $b_k := \dot{m}_k - \partial_t M_k|_w$*, which satisfies $\mathbf{1}^\top b = 0$. Then the potentials obey a Kirchhoff system on $G_B$, and every interface moves by the voltage drop across its own edge:*

$$\boxed{L_B\,\dot{w} = b, \qquad \dot{w} = L_B^\dagger b, \qquad V_{ij}(x) = \frac{(\dot{w}_i - \dot{w}_j) - \partial_t(c_i - c_j)(x)}{\|\nabla_x(c_i - c_j)(x)\|}} \tag{4}$$

*where $L_B^\dagger$ is the Moore–Penrose pseudo-inverse in the gauge $\mathbf{1}^\top\dot{w} = 0$. In particular $\overline{V}_{ij} = |\dot{w}_i - \dot{w}_j|/\gamma_{ij}$ when $\partial_t c \equiv 0$.*

Programs are nodes, interfaces are resistors of conductance $h_{ij}$, $\dot{w}$ is the voltage and the shift is a charge-conserving current injection. The two terms of Equation (4) differ in kind: $\partial_t(c_i - c_j)$ is local, what a drifting representation supplies, while $\dot{w}_i - \dot{w}_j$ is a functional of the *whole* graph. We call Equation (4) the *current-conditioned prediction*: it needs $b$, at deployment a measurable mass imbalance needing no labels.

**Theorem 3** (Effective-resistance bound with a tight network term)**.** *Let $R_{ij} = (e_i - e_j)^\top L_B^\dagger (e_i - e_j)$ be the effective resistance between $i$ and $j$ in $G_B$, let $\|b\|_{L_B^\dagger} = \sqrt{b^\top L_B^\dagger b}$ and $\delta_{ij} = \sup_{x \in B_{ij}} |\partial_t (c_i - c_j)(x)|$. Then*

$$|\dot{w}_i - \dot{w}_j| \;\le\; \sqrt{R_{ij}}\, \|b\|_{L_B^\dagger}, \qquad \boxed{\overline{V}_{ij} \;\le\; \frac{\sqrt{R_{ij}}\, \|b\|_{L_B^\dagger} + \delta_{ij}}{\gamma_{ij}}} \tag{5}$$

*with equality in the* first *inequality if and only if $b \propto e_i - e_j$; then $\dot{w}_i - \dot{w}_j = R_{ij}$ for $b = e_i - e_j$, so the network term cannot be improved. Remark 10 gives the Hausdorff form and the extra regularity it needs.*

Three distinctions matter throughout. Tightness is a property of the network term alone: the velocity bound is attained only when in addition $\delta_{ij} = 0$. Equation (5), the *first-order certificate*, bounds the first-order velocity and not the displacement realised over a finite shift, a gap Theorem 8 quantifies and Section 5.3 measures. And the factor it is usually quoted through, the *shift-agnostic susceptibility* $\Phi_{ij} := \sqrt{R_{ij}}/\gamma_{ij}$, is the response per unit of injected shift energy and not a bound on any displacement, since Equation (5) also contains $\|b\|_{L_B^\dagger}$ and $\delta_{ij}$: topology fixes the susceptibility, the realised shift supplies the current.

### 4.1 Local data certifies, but cannot tell whether the certificate is informative

**Definition 4** (Local data, local functional)**.** Fix $\varepsilon_0 > 0$, let $N$ be the $\varepsilon_0$-tube around $B_{ij}$, and call $\mathrm{loc}_{ij} = (B_{ij},\, c_i|_N,\, c_j|_N,\, q|_N,\, b_i,\, b_j)$ the *local data* of $(i, j)$. A *local functional* is any map on such tuples: it may compute $\gamma_{ij}$, $h_{ij}$, $|B_{ij}|$ and anything else supported on $N$, but not the other programs, the rest of $G_B$, or the current injected elsewhere. Entropically there is no hard interface and the counterpart is $\mathrm{loc}_{ij}^{\varepsilon} = (h_{ij}^{\varepsilon}, \gamma_{ij}, b_i, b_j)$, the arguments the local certificate consumes.

**Proposition 5** (A valid local certificate does exist)**.** *For $b = e_i - e_j$ and $\partial_t c \equiv 0$ we have $\overline{V}_{ij} = R_{ij}/\gamma_{ij}$, and Rayleigh monotonicity gives $R_{ij} \le 1/h_{ij}$, so*

$$\overline{V}_{ij} \;\le\; \frac{1}{h_{ij}\, \gamma_{ij}} \qquad \textit{(the} \text{ local certificate}\textit{)}, \tag{6}$$

*a local functional in the sense of Definition 4: the two-cell model, the response of the isolated edge $(i, j)$. The restriction to $b = e_i - e_j$ is essential, since for a current injected elsewhere, scaling it leaves $\mathrm{loc}_{ij}$ unchanged while the response grows without bound, so* no *local bound exists for general $b$.*

**Theorem 6** (No uniformly informative local certificate)**.** *Fix any $\rho > 1$. There exist two mass-constrained partitions* L *and* H*, each satisfying Assumption 1, with identical local data at the interface $(1, 2)$ in the sense of Definition 4, receiving the same injected current $b = e_1 - e_2$, with*

$$\overline{V}_{12}^{(\mathrm{H})} \;\ge\; \rho\, \overline{V}_{12}^{(\mathrm{L})}. \tag{7}$$

*Consequently, for every local functional $\mathcal{F}$:* (i) *used as a point predictor of $\overline{V}_{12}$ it returns one number for both worlds, so its multiplicative error is at least $\sqrt{\rho}$ on one of them;* (ii) *used as an upper bound valid on both, it over-estimates $\overline{V}_{12}^{(\mathrm{L})}$ by a factor at least $\rho$.*

The proof (Appendix B) is a Rayleigh-monotonicity argument: with the local data held fixed, $R_{12}$ can be placed anywhere in $(0, 1/h_{12}]$ by editing only the rest of the network, and the displacement follows it exactly when $b = e_1 - e_2$. Adding $n$ node-disjoint bridges to the low-response world gives ratio $1 + n\kappa/(2h_{12})$ with every cell mass bounded below, which is how an arbitrary $\rho$ is obtained; the planar two-cap family of Figure 1 is the one we can draw, and its ratio grows as $1/(2\beta)$ with a vanishing cell mass and algebraic connectivity, so there Assumption 1 holds for each $\beta > 0$ with $\beta$-dependent constants. Read with Proposition 5 this is the paper's impossibility statement: *local information can certify a mass-constrained interface but cannot tell whether the certificate is informative*. Corollary 11 completes the picture on average: a current isotropic in the energy inner product gives $\mathbb{E}[(\dot{w}_i - \dot{w}_j)^2] = R_{ij}$, so $\Phi_{ij}$ ranks by RMS response, while an unnormalised isotropic current makes $\widetilde{R}_{ij} = (e_i - e_j)^\top (L_B^\dagger)^2 (e_i - e_j)$ exact. Which ranks better is a property of the ensemble, not of the theory, and on real ensembles neither beats the local certificate.

### 4.2 ENTROPIC PARTITIONS: SINKHORN HEADS CARRY THE SAME CIRCUIT

Learning systems rarely solve Equation (1) exactly; they balance a temperature-$\varepsilon$ relaxation (Cuturi, 2013), which over a finite sample $\{x_v\}$ with weights $q_v$ reads $p_{vk} \propto \exp\big((w_k - c_k(x_v))/\varepsilon\big)$ subject to $\sum_v q_v p_{vk} = m_k$.

**Theorem 7** (Entropic circuit)**.** *The entropic mass map $M_k^\varepsilon(w) = \sum_v q_v p_{vk}(w)$ has Jacobian $\partial M^\varepsilon / \partial w = L_B^\varepsilon$, the Laplacian with conductances*

$$h_{ij}^\varepsilon = \frac{1}{\varepsilon} \sum_v q_v \, p_{vi} \, p_{vj} \ \geq 0. \tag{8}$$

*As $L_B^\varepsilon$ is a symmetric weighted Laplacian, connected whenever the co-assignment graph is, Theorems 2 and 3 and Proposition 5 hold verbatim with $L_B$ replaced by $L_B^\varepsilon$ and $h_{ij}^\varepsilon$ in place of $h_{ij}$.* Nonlocality transfers with them*: since $\dot{w}_i - \dot{w}_j = (e_i - e_j)^\top (L_B^\varepsilon)^\dagger \dot{b}$ depends on every entry of $L_B^\varepsilon$, two entropic partitions with the same $\mathrm{loc}_{ij}^\varepsilon$ respond differently and no functional of it is uniformly informative. Separately, for a partition of $\mathcal{X} \subset \mathbb{R}^d$ satisfying Assumption 1 and the uniform-gap condition (A5) of Appendix B, $h_{ij}^\varepsilon \to h_{ij}$ as $\varepsilon \to 0$.*

The hard-interface construction of Theorem 6 is not claimed verbatim in this setting; the algebraic statement above is, and Section 5.1 realises it on finite samples. Equation (8) is a closed form: an interface conductance is the *co-assignment mass* of its two programs over the temperature. It costs one $K \times K$ Gram matrix, needs no geometry, and is what every real system below uses, so those experiments rest on the Jacobian identity and never on the $\varepsilon \to 0$ limit or on (A5). Theorem 7 concerns the balanced entropic *optimum*, not an arbitrary iterate: every partition here is solved to $\|M^\varepsilon(w) - m\|_\infty < 10^{-13}$, and Section 5.3 measures how far a fixed budget is from it.

### 4.3 WHEN THE FIRST-ORDER LAW STOPS BEING VALID

**Theorem 8** (Linear-response regime; proof in Appendix B)**.** *Suppose on $[0, T]$ that Assumption 1 holds, the diagram keeps its combinatorial type, and (A6) $t \mapsto L_B(t)$ and $t \mapsto b(t)$ are $C^1$, which holds when each $h_{ij}$ is a boundary integral over a smoothly varying interface and $(c^t, q^t, m^t)$ is $C^2$; note that (A3) alone gives only $b$ continuous. Then $w \in C^2$ and*

$$\|w(t) - w(0) - t\dot{w}(0)\| \ \leq \ \tfrac{1}{2} t^2 \sup_{s \in [0,t]} \frac{\|\dot{b}(s)\| + \|\dot{L_B}(s)\| \, \|\dot{w}(s)\|}{\lambda_2(L_B(s))}. \tag{9}$$

*The* absolute *error is therefore $O(t^2)$, with a constant that degrades as the circuit approaches disconnection; if the first-order displacement is non-degenerate, $\|\dot{w}(0)\| \geq c > 0$, the* relative *error is $O(t)$. Equation* (9) *controls the potentials, and the interface displacement inherits it divided by $\gamma_{ij}$ once $\partial_t(c_i - c_j)$ is $C^1$ and $\gamma_{ij}$ is bounded below.*

Theorem 8 excludes a change of combinatorial type, and Proposition 12 in Appendix B shows the excluded event is benign: a newly adjacent pair enters with $h_{ij} \leq \|q\|_\infty \gamma_{ij}^{-1} \mathcal{H}^{d-1}(B_{ij}) \to 0$, so $L_B$ is continuous through the birth, though continuity is not differentiability and the constant can still fail there.

## 5 EXPERIMENTS

**Protocol.** Every prediction uses the *unperturbed* partition only: conductances from Equation (3) (synthetic, exact boundary integrals) or Equation (8) (real data, closed form), never from a finite difference of the solver, and $b$ is the mass imbalance the shift creates at frozen potentials. Ground truth is the displacement realised after re-solving exactly at the shifted parameters; in two dimensions we clip polygons exactly and integrate by Gauss quadrature, so masses are met to $1.2 \times 10^{-15}$ and the observables carry no discretisation error. Real corruption families are not $C^1$ in severity, so there Equation (4) is applied in finite-step form, which needs only differentiability of $M$ at $w$. The statistical unit is the partition, image, patient or shift; correlations on real data are computed *within* a unit and averaged, with a 95% bootstrap interval over units (Appendix C).

### 5.1 CAN LOCAL INFORMATION DETERMINE BOUNDARY ROBUSTNESS?

Figure 1 realises Theorem 6. Across 180 configurations, $B_{12}$, $\gamma_{12} = 0.80$ and $h_{12} = 0.088$ are identical ($\gamma$ to machine zero, $h$ to $3.1 \times 10^{-14}$ relative) while $R_{12}$ spans $10.9\times$, because only the wiring of the *other* regions changes; for the two worlds in panels (a, b) the resistance and displacement ratios are $10.6\times$ and $10.6\times$. Over the sweep the displacement spans $10.9\times$ and Theorem 2 predicts it with $R^2 = 0.999998$, against $R^2 = -9.44$ for a local-margin law, and this is no artefact of injecting current at the interface: a random zero-sum forcing gives $R^2 = 0.999958$ and one applied *only* to cells that never touch $B_{12}$ gives $R^2 = 0.999997$. Shrinking the pinned interface drives the ratio to $35.5\times$ at the $1/(2\beta)$ rate of Appendix B but also drives the smallest cell mass to $0.0046$, which is why an arbitrary $\rho$ is proved by the graph-level construction instead (Appendix C.7).

**The same thing without a hard interface.** Theorem 7 predicts that entropic assignments inherit the phenomenon. We build 6 finite entropic partitions sharing $\gamma_{12}$ exactly and $h^{\varepsilon}_{12}$ to a relative spread of $8.1 \times 10^{-12}$, so Equation (6) returns the same number for all of them, differing only in where the remaining programs sit. Their responses span $2.11\times$, tracking $R_{12}$, and the circuit predicts the re-solved entropic response to $5.9 \times 10^{-4}$ relative. Unlike the planar family nothing degenerates: the smallest cell mass is $0.0134$ and $\lambda_2 \geq 0.0043$.

### 5.2 DOES THE BALANCE CONSTRAINT CAUSE IT?

Table 4 drives the law through three physically distinct channels: prior shift, cost drift and covariate shift. In every case the predicted velocity matches the exactly re-solved one with $R^2 \geq 0.99999$, unit slope and relative error decaying with log-log slope $\approx 1$ over 2.5 decades, which is a correctness check on the linearisation, not a test of the hypothesis. The knockout is: the *same* cost perturbation on the *same* prototypes with the constraint switched off makes the purely local law exact ($R^2 = 1.0000$) and switching it back on makes that law worse than predicting no motion ($R^2 = -2.2250$). The first half is definitional, since without the constraint $\dot{w} \equiv 0$; the informative half is that reinstating it costs the local law all of its accuracy. The same test runs on *real* features, with the same log-log slope $0.998$ and an entropic Laplacian matching the true Jacobian to $2.2 \times 10^{-8}$ (Appendix C.7).

**Beyond the guaranteed regime.** Growing the shift on 90 random partitions keeps the relative error linear in $t$ over three and a half decades, and two findings go beyond Theorem 8 (Figure 2c, d; detail in Appendix C.7). *Creating or destroying an adjacency was benign in every instance we generated* though the theorem excludes it, which Proposition 12 explains without becoming a finite-shift guarantee; what hurts is the *cell collapse* the theorem does predict, the $9\%$ of instances that lose most of a cell's mass having median error $0.164$ against $0.088$. And a vanishing margin does *not* make an interface explosively fragile, because $\gamma_{ij}$ also appears in $h_{ij} = \int_{B_{ij}} q/\gamma_{ij}$ and the two cancel: over a $37.5\times$ margin collapse the naive $1/\gamma$ law is wrong by $36.4\times$ while the circuit tracks the observed plateau to $2.7 \times 10^{-4}$.

**Is it really the effective resistance?** A predictor can look good for the wrong reason, so we compare the circuit against eight alternatives on the same interfaces (Table 5): three local, two global but not electrical, and three corruptions that keep its ingredients, namely conductances permuted across edges, conductances rewired at fixed weighted degree, and the resistance of the *wrong* graph. Under the ensemble the theory is stated for, the circuit ranks at $\rho = 0.875$, ahead of the degree-preserving shuffle ($0.814$), the local margin ($0.785$), the permuted circuit ($0.701$) and the wrong graph ($0.579$). That ensemble is built from the true $L_B$, so Corollary 11 makes $\Phi_{ij}$ its exact RMS predictor by construction and the informative quantity is the *gap* to the controls; under the plain isotropic ensemble $\Phi_{ij}$ falls to $0.429$ and the $(L_B^{\dagger})^2$ form takes over ($0.739$).

The shuffle is a strong control and we do not overstate the gap: per partition $\rho_{\text{circuit}} - \rho_{\text{shuffle}} = 0.061$ $[0.036, 0.093]$ on $82\%$ of partitions (Wilcoxon $p < 0.001$), and residualised on the local variables the true resistance retains $0.535$ against $0.339$ for the shuffle (Appendix C.7). The defensible statement is that *the exact conductance network carries information beyond degree-preserving structure*, not that the latter carries none.

Table 1: **The circuit on real systems, against the honest competitor.** Eight image corruptions at six severities and five raw-count corruptions at four, over 48 shifts, 200 images and 7 patients; Spearman $\rho$ is computed within a unit and averaged, with a bootstrap interval over units, and sign accuracies are in the text. *uses re-solved conductances, so it is a diagnostic, not a predictor. The two-cell row is the current-conditioned isolated-edge predictor $(b_i - b_j)/(2h_{ij}\gamma_{ij})$, the local competitor that does use the measured shift, its slope a scale artefact and omitted; the last block is $k$ Sinkhorn column updates warm-started at the clean potentials (Section 5.3).

| | CIFAR-10 head | | Pets | | HER2$^+$ tumours | |
|---|---|---|---|---|---|---|
| predictor | $\rho$ [95% CI] | slope | $\rho$ [95% CI] | slope | $\rho$ [95% CI] | slope |
| first-order law Equation (4) | $+0.95$ $[+0.92, +0.97]$ | 0.36 | $+0.93$ $[+0.93, +0.94]$ | 0.57 | $+0.95$ $[+0.94, +0.97]$ | 0.88 |
| predictive trapezoid | $+0.99$ $[+0.98, +0.99]$ | 0.35 | $+0.97$ $[+0.97, +0.97]$ | 0.69 | $+0.96$ $[+0.93, +0.98]$ | 1.12 |
| exact trapezoid* | $+0.97$ $[+0.95, +0.98]$ | 1.01 | $+0.97$ $[+0.96, +0.97]$ | 0.86 | $+0.95$ $[+0.94, +0.96]$ | 0.90 |
| two-cell Equation (6) | $+0.80$ $[+0.78, +0.81]$ | – | $+0.86$ $[+0.85, +0.86]$ | – | $+0.64$ $[+0.48, +0.78]$ | – |
| $k = 1$ warm Sinkhorn | $+0.88$ $[+0.87, +0.89]$ | 2.55 | – | – | – | – |
| $k = 5$ warm Sinkhorn | $+0.91$ $[+0.89, +0.92]$ | 1.92 | – | – | – | – |
| $k = 20$ warm Sinkhorn | $+0.96$ $[+0.95, +0.96]$ | 1.23 | – | – | – | – |
| $k = 50$ warm Sinkhorn | $+0.98$ $[+0.98, +0.99]$ | 1.06 | – | – | – | – |

### 5.3 THREE REAL SYSTEMS: RANKS TRANSFER, MAGNITUDES DO NOT

The three systems are a balanced self-labelling head (ten class prototypes on frozen ResNet-18 features of CIFAR-10, a temperature-2.42 Sinkhorn head assigning 5,000 images at 0.736 accuracy), 200 Oxford-IIIT Pet images cut into $K = 6$ mass-balanced regions of a frozen patch-feature space, and 8 HER2$^+$ tumour sections with pathologist annotations (Andersson et al., 2021), 7 of which enter the interface analysis leave-one-patient-out with target masses from the *training* patients, so no test label enters. *Ranks transfer far outside the linear regime*: the current-conditioned law orders interfaces at $\rho = 0.950$, 0.934 and 0.954 and its trapezoidal form at 0.987, 0.968 and 0.957, with sign accuracy 97.9%, 94.4% and 97.7%, and both beat the local current-conditioned competitor on every system, paired per unit: the first-order law by 0.151 $[0.117, 0.184]$, 0.077 $[0.070, 0.084]$ and 0.318 $[0.177, 0.470]$ on 88%, 97% and 100% of units, the trapezoid by 0.189, 0.111 and 0.321 on 100%, 100% and 100%. *Magnitudes do not*: the first-order slope on CIFAR-10 falls from 0.67 to 0.32 as corruption grows and the head loses 28 accuracy points, the first-order law drops below $R^2 = 0$ from severity 0.35 and the deployable trapezoid from 0.60, and only the exact trapezoid, which needs re-solved conductances, keeps slope 1.01 and $R^2 = 0.744$ (Appendix C.7). The usable claim on real data is direction and order, not calibrated displacement.

**What the circuit is worth against re-solving.** The law needs the shifted data, and with those in hand one can simply iterate: $k$ Sinkhorn column updates warm-started at the clean potentials cost one pass each and converge to the exact re-solve (Table 1). A single update is diagonal, using only each cell's own mass residual, and already reaches $\rho = 0.883$; the budget must grow to $k = 50$ to match the trapezoidal circuit (0.984 against 0.987), by which point it costs 3.12 s against 0.08 s. That gap is the part of the response a node-local update cannot see, the nonlocality of Theorem 6 measured without synthetic geometry. The conclusion cuts both ways: *for ranking the circuit is a* 38*-fold saving, for calibrated magnitudes iterate instead*, since even $k = 1$ attains $R^2 = 0.21$ where the first-order circuit is at $-1.31$. The same runs audit the equilibrium Theorem 7 assumes: the median marginal violation is $2.8 \times 10^{-2}$ at $k = 1$ and $4.1 \times 10^{-3}$ at $k = 50$, against $10^{-13}$ throughout.

**Where the shift-agnostic score stands.** Equation (5) is first order and the finite-step data say so: realised displacement exceeds it in 20.3% of tumour rows against 0.0% and 2.3% on the vision systems, falling to 9.3% on the trapezoidal circuit. More importantly, $\Phi_{ij}$ ranks significantly on *none* of the three systems once the tumour masses come from the training patients (0.210 $[-0.270, 0.614]$ there), and on CIFAR-10 and Pets the shift-agnostic *local* score $1/(h_{ij}\gamma_{ij})$ beats it (0.195 and 0.064 against 0.027 and $-0.001$, non-overlapping intervals). Real corruption ensembles are not the one Corollary 11 assumes, and this is the sharpest limitation we report: *before a shift arrives the network score is no better than a local one; the nonlocal advantage appears once the shift is measured.*

**A head that is trained, and an encoder that drifts.** All three systems above freeze the costs, so the direct term of Equation (4) never acts. We therefore train the head the paper is about: from the

same partition, Table 2 first trains the $10$ prototypes with the SwAV objective, then fine-tunes the last residual block of the encoder jointly with them. Prototype drift reaches $27.0\%$ and the combined drift $65.7\%$, far outside the linear regime, and the circuit survives both. Under head training the network term is predicted essentially exactly ($\rho = 0.995$, $R^2 = 0.971$, scored against the realised potential difference rather than the full displacement) and the error of the full law grows as $O(\text{drift})$ with log-log slope $1.12$. Under joint drift the trapezoidal form is accurate in both senses at once ($\rho = 0.993$, slope $0.99$, $R^2 = 0.977$), and the knockout is decisive: keeping the cost-drift term and dropping the re-equilibration gives $\rho = -0.190$ and sign accuracy $44.9\%$, no better than a coin. The direct cost term is not small in norm but does not explain the *signed* response; the network carries $89\%$ of the unsigned motion and all of its direction.

**What the circuit buys, and what it costs.** Where the shift is given, the trapezoidal circuit finds the most displaced interface $89.6\%$ / $74.1\%$ / $46.3\%$ of the time on the three systems against $0.0\%$ / $12.7\%$ / $22.3\%$ for a margin ranking, and where it is not, using Corollary 11 to choose which $B = \lceil K/4 \rceil$ programs to *anchor* removes $34.1\%$ of the displacement energy under re-solution against $31.4\%$ for the best non-electrical rule, a gap of $0.028$ $[0.014, 0.042]$ that is significant but small (Appendix C.7). The audit costs one $K \times K$ Gram matrix and $K$ sparse solves: at $N = 200{,}000$, $K = 16$ the balanced assignment takes $6.39$ s and the circuit plus *all* pairwise resistances $0.155$ s, never above $5.3\%$ overhead on the tested grid (Appendix C.6).

## 6 Discussion, limitations and conclusion

**Why nonlocality appears.** Nothing in Theorem 2 is specific to transport: it needs the Jacobian of the constraint residual with respect to the balancing duals to be a connected symmetric weighted Laplacian, which a conservation law across regions produces and Lemma 1 and Theorem 7 supply. Nonlocality is the price of balance, and Table 4 shows the switch directly.

**What the right robustness object is, and when.** Before the shift is seen the answer is not the network: on our three real systems the shift-agnostic $\Phi_{ij}$ ranks no better than the local $1/(h\gamma)$, because real corruption ensembles are not the energy-isotropic one Corollary 11 assumes. Once it is measured the answer is the network: the current-conditioned law beats the local current-conditioned predictor on every system, its trapezoidal form on every unit, and that form matches a $k = 50$ Sinkhorn budget on ranking at $1/38$ of its cost. For calibrated displacement, iterate instead. That split, not a single score, is what we would tell a practitioner.

**When it stops being quantitatively accurate.** Inside the regime of Theorem 8 the law predicts magnitudes; outside it, ranks and signs survive much further. On the corruption grid the deployable forms drop below $R^2 = 0$ beyond moderate severity while rank correlations stay above $0.957$, and Equation (5), bounding only the first-order term, is exceeded on $20.3\%$ of tumour rows. Cell collapse, driving $\lambda_2(L_B) \to 0$, is the failure mode the theory names and the experiments confirm; adjacency birth and death, which it excludes, was benign everywhere we looked, which Proposition 12 explains without promising it in general.

**What is and is not covered, and the conclusion.** Exact mass-constrained partitions and entropic balanced assignments are covered, one system is trained rather than frozen, and the impossibility is realised both with hard interfaces and on finite entropic samples. Not covered: production-scale $K$ (we test $K \leq 24$), a self-supervised head with thousands of prototypes, a learned codebook or a capacity-factor router, whose Jacobians we do not verify to be Laplacian, and encoders larger than a fine-tuned ResNet-18 block; the three frozen systems also hold the costs fixed, so the direct term of Equation (4) is exercised only synthetically and by Table 2. A mass constraint turns boundary robustness from a local geometric property into a global network one: a shift injects a current, potentials re-equilibrate through a Kirchhoff system, and each boundary moves by the voltage drop across its own edge over its margin, so two locally indistinguishable boundaries can respond by arbitrarily different amounts and a valid local certificate can say nothing. The object to compute is the circuit, and the current that drives it.

## AI USE STATEMENT

In this work we used generative AI tools in the following ways, all of which we have reviewed and for which we take full responsibility.

Generative AI was used throughout this work. It helped draft, polish and review the text, search for and format references, and propose the title; coding agents processed the datasets and implemented the methods; AI took part in the discussion of the results, generated the plots, and helped to create slides for internal meetings. Generative AI was not used to generate or annotate data. Every closed form was verified computationally, the code was tested before any result was reported, and every reference was checked against its publisher or archive page. The authors reviewed all AI-assisted content and take full responsibility for the paper, its claims and its code.

## REPRODUCIBILITY STATEMENT

All theoretical claims are stated with their assumptions in Assumption 1 and proved in full in Appendix B; the construction behind Theorem 6 is given explicitly, including the closed-form cap weights Equation (10) that pin the interface, so the figure can be rebuilt from the text alone. Appendix C specifies every component needed to reproduce the experiments: the exact polygonal transport solver and its tolerances, the conductance estimators, the six synthetic layouts, the two shift ensembles, the corruption grids for images and for raw counts, the frozen encoder and the feature pipeline, the leave-one-patient-out protocol, the optimiser, learning rates and checkpoint schedule of the trained head and the drifting encoder, and the point at which each hyper-parameter was selected. The three real datasets are public (Krizhevsky, 2009; Parkhi et al., 2012; Andersson et al., 2021) and we list the exact subsets used. Every number in the paper is generated programmatically from the stored result files rather than transcribed, and all randomness is seeded. Correctness of the implementation is checked against analytic values by unit tests that verify the mass Jacobian against the boundary-integral Laplacian, the tightness case of Theorem 3, and the entropic limit of Theorem 7; the numbers those tests produce are reported in Appendix C, so a reimplementation can be checked against them directly. The bridge-conductance limit and the verification of Assumption 1 along the path of Theorem 6 are themselves a script whose output is quoted in Appendix B, and the unit-level intervals and paired tests of Table 1 are recomputed from the stored per-interface rows rather than from summary statistics.

## ETHICS STATEMENT

This work is a sensitivity analysis of a class of learning algorithms and involves no human subjects, no new data collection and no deployment. Two of the three real datasets are standard public benchmarks (Krizhevsky, 2009; Parkhi et al., 2012). The third is a published, de-identified spatial-transcriptomics resource of HER2-positive breast tumours released under CC-BY-4.0 by its authors (Andersson et al., 2021); we use it only as a testbed for a general theory, we do not attempt re-identification, and we make no diagnostic or prognostic claim. The partitions we produce for that dataset are not clinical outputs and should not be used as such.

A method that predicts which decision boundaries will move under distribution shift is dual-use in a mild sense: the same susceptibility score that tells an operator where to invest in robustness also tells an adversary which interface is cheapest to disturb. We judge the risk to be low (the score is computed from a model's own partition, so an adversary who can compute it already has white-box access), and on balance favourable, because the main practical effect is to make a class of failures that is currently invisible to local analysis measurable in advance. The theory also suggests a fairness-relevant hypothesis that we did not test and therefore do not claim: since a program's response scales with its effective resistance, poorly connected programs should be the more shift-sensitive ones, and whether that correlates with group size is an empirical question our experiments do not answer. Practitioners who rely on balance constraints for equitable capacity allocation should be aware that balance at training time does not imply equal stability at deployment time.

## REFERENCES


Amr Alexandari, Anshul Kundaje, and Avanti Shrikumar. Maximum likelihood with bias-corrected calibration is hard-to-beat at label shift adaptation. In *International Conference on Machine Learning*, pp. 222–232, 2020.

Alma Andersson, Ludvig Larsson, Linnea Stenbeck, Fredrik Salmén, Anna Ehinger, Sunny Z Wu, Ghamdan Al-Eryani, Daniel Roden, Alexander Swarbrick, Åke Borg, et al. Spatial deconvolution of HER2-positive breast cancer delineates tumor-associated cell type interactions. *Nature Communications*, 12(1):6012, 2021.

Yuki M Asano, Christian Rupprecht, and Andrea Vedaldi. Self-labelling via simultaneous clustering and representation learning. In *International Conference on Learning Representations*, 2020.

Hugo Attali, Davide Buscaldi, Nathalie Pernelle, and Fragkiskos D Malliaros. Graph rewiring in GNNs to mitigate over-squashing and over-smoothing: a survey. *arXiv preprint arXiv:2411.17429*, 2024.

Franz Aurenhammer, Friedrich Hoffmann, and Boris Aronov. Minkowski-type theorems and least-squares clustering. *Algorithmica*, 20(1):61–76, 1998.

Mohit Bansil and Jun Kitagawa. Quantitative stability in the geometry of semi-discrete optimal transport. *International Mathematics Research Notices*, 2022(10):7354–7389, 2022.

Peter L Bartlett, Dylan J Foster, and Matus J Telgarsky. Spectrally-normalized margin bounds for neural networks. In *Advances in Neural Information Processing Systems*, volume 30, 2017.

Mitchell Black, Zhengchao Wan, Amir Nayyeri, and Yusu Wang. Understanding oversquashing in GNNs through the lens of effective resistance. In *International Conference on Machine Learning*, pp. 2528–2547, 2023.

David P Bourne, Pieter JJ Kok, Steven M Roper, and Wilbert DT Spanjer. Laguerre tessellations and polycrystalline microstructures: a fast algorithm for generating grains of given volumes. *Philosophical Magazine*, 100(21):2677–2707, 2020.

Paul S Bradley, Kristin P Bennett, and Ayhan Demiriz. Constrained k-means clustering. Technical Report MSR-TR-2000-65, Microsoft Research, 2000.

Zixuan Cang, Yanxiang Zhao, Axel A Almet, Adam Stabell, Raul Ramos, Maksim V Plikus, Scott X Atwood, and Qing Nie. Screening cell–cell communication in spatial transcriptomics via collective optimal transport. *Nature Methods*, 20(2):218–228, 2023.

Mathilde Caron, Ishan Misra, Julien Mairal, Priya Goyal, Piotr Bojanowski, and Armand Joulin. Unsupervised learning of visual features by contrasting cluster assignments. In *Advances in Neural Information Processing Systems*, volume 33, 2020.

Mathilde Caron, Hugo Touvron, Ishan Misra, Hervé Jégou, Julien Mairal, Piotr Bojanowski, and Armand Joulin. Emerging properties in self-supervised vision transformers. In *Proceedings of the IEEE/CVF International Conference on Computer Vision*, pp. 9630–9640, 2021.

Ashok K Chandra, Prabhakar Raghavan, Walter L Ruzzo, Roman Smolensky, and Prasoon Tiwari. The electrical resistance of a graph captures its commute and cover times. *Computational Complexity*, 6(4):312–340, 1996.

Jeremy Cohen, Elan Rosenfeld, and Zico Kolter. Certified adversarial robustness via randomized smoothing. In *International Conference on Machine Learning*, pp. 1310–1320, 2019.

Nicolas Courty, Rémi Flamary, Amaury Habrard, and Alain Rakotomamonjy. Joint distribution optimal transportation for domain adaptation. In *Advances in Neural Information Processing Systems*, volume 30, 2017.

Marco Cuturi. Sinkhorn distances: lightspeed computation of optimal transport. In *Advances in Neural Information Processing Systems*, volume 26, 2013.

Damai Dai, Li Dong, Shuming Ma, Bo Zheng, Zhifang Sui, Baobao Chang, and Furu Wei. StableMoE: Stable routing strategy for mixture of experts. In *Proceedings of the 60th Annual Meeting of the Association for Computational Linguistics*, pp. 7085–7095, 2022.

Frédéric De Gournay, Jonas Kahn, and Léo Lebrat. Differentiation and regularity of semi-discrete optimal transport with respect to the parameters of the discrete measure. *Numerische Mathematik*, 141(2):429–453, 2019.

Alex Delalande and Quentin Mérigot. Quantitative stability of optimal transport maps under variations of the target measure. *Duke Mathematical Journal*, 172(17):3321–3357, 2023.

Francesco Di Giovanni, Lorenzo Giusti, Federico Barbero, Giulia Luise, Pietro Liò, and Michael M Bronstein. On over-squashing in message passing neural networks: the impact of width, depth, and topology. In *International Conference on Machine Learning*, pp. 7865–7885, 2023.

Peter G Doyle and J Laurie Snell. *Random Walks and Electric Networks*. Carus Mathematical Monographs. Mathematical Association of America, 1984.

Marvin Eisenberger, Aysim Toker, Laura Leal-Taixé, Florian Bernard, and Daniel Cremers. A unified framework for implicit sinkhorn differentiation. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, pp. 499–508, 2022.

William Fedus, Barret Zoph, and Noam Shazeer. Switch transformers: scaling to trillion parameter models with simple and efficient sparsity. *Journal of Machine Learning Research*, 23(120):1–39, 2022.

Saurabh Garg, Sivaraman Balakrishnan, Zachary C Lipton, Behnam Neyshabur, and Hanie Sedghi. Leveraging unlabeled data to predict out-of-distribution performance. In *International Conference on Learning Representations*, 2022.

Kaiming He, Xiangyu Zhang, Shaoqing Ren, and Jian Sun. Deep residual learning for image recognition. In *IEEE Conference on Computer Vision and Pattern Recognition*, pp. 770–778, 2016.

Dan Hendrycks and Thomas Dietterich. Benchmarking neural network robustness to common corruptions and perturbations. In *International Conference on Learning Representations*, 2019.

Yiding Jiang, Dilip Krishnan, Hossein Mobahi, and Samy Bengio. Predicting the generalization gap in deep networks with margin distributions. In *International Conference on Learning Representations*, 2019.

Jun Kitagawa, Quentin Mérigot, and Boris Thibert. Convergence of a newton algorithm for semi-discrete optimal transport. *Journal of the European Mathematical Society*, 21(9):2603–2651, 2019.

Douglas J Klein and Milan Randić. Resistance distance. *Journal of Mathematical Chemistry*, 12(1): 81–95, 1993.

Pang Wei Koh, Shiori Sagawa, Henrik Marklund, Sang Michael Xie, Marvin Zhang, Akshay Balsubramani, Weihua Hu, Michihiro Yasunaga, Richard Lanas Phillips, Irena Gao, et al. WILDS: A benchmark of in-the-wild distribution shifts. In *International Conference on Machine Learning*, pp. 5637–5664, 2021.

Alex Krizhevsky. Learning multiple layers of features from tiny images. Technical report, University of Toronto, 2009.

Bruno Lévy. A numerical algorithm for $L_2$ semi-discrete optimal transport in 3D. *ESAIM: Mathematical Modelling and Numerical Analysis*, 49(6):1693–1715, 2015.

Mike Lewis, Shruti Bhosale, Tim Dettmers, Naman Goyal, and Luke Zettlemoyer. BASE layers: simplifying training of large, sparse models. In *International Conference on Machine Learning*, pp. 6265–6274, 2021.

Giulia Luise, Alessandro Rudi, Massimiliano Pontil, and Carlo Ciliberto. Differential properties of sinkhorn approximation for learning with wasserstein distance. In *Advances in Neural Information Processing Systems*, volume 31, 2018.

Quentin Mérigot. A multiscale approach to optimal transport. *Computer Graphics Forum*, 30(5): 1583–1592, 2011.

John P Miller, Rohan Taori, Aditi Raghunathan, Shiori Sagawa, Pang Wei Koh, Vaishaal Shankar, Percy Liang, Yair Carmon, and Ludwig Schmidt. Accuracy on the line: on the strong correlation between out-of-distribution and in-distribution generalization. In *International Conference on Machine Learning*, pp. 7721–7735, 2021.

Philip Naumann, Jacob Kauffmann, Klaus-Robert Müller, and Grégoire Montavon. Reliable modeling of distribution shifts via displacement-reshaped optimal transport. *arXiv preprint arXiv:2605.04965*, 2026.

Maxime Oquab, Timothée Darcet, Théo Moutakanni, Huy Vo, Marc Szafraniec, Vasil Khalidov, Pierre Fernandez, Daniel Haziza, Francisco Massa, Alaaeldin El-Nouby, et al. DINOv2: Learning robust visual features without supervision. *Transactions on Machine Learning Research*, 2024.

Davy Paindaveine and Riccardo Passeggeri. On the robustness of semi-discrete optimal transport. *arXiv preprint arXiv:2410.19596*, 2024.

Omkar M Parkhi, Andrea Vedaldi, Andrew Zisserman, and CV Jawahar. Cats and dogs. In *IEEE Conference on Computer Vision and Pattern Recognition*, pp. 3498–3505, 2012.

Gabriel Peyré and Marco Cuturi. Computational optimal transport: with applications to data science. *Foundations and Trends in Machine Learning*, 11(5-6):355–607, 2019.

Zoe Piran, Michal Klein, James Thornton, and Marco Cuturi. Contrasting multiple representations with the multi-marginal matching gap. In *International Conference on Machine Learning*, pp. 40827–40842, 2024.

Sawyer Robertson, Zhengchao Wan, and Alexander Cloninger. Resistance distance and linearized optimal transport on graphs. *arXiv preprint arXiv:2404.15261*, 2024.

Marco Saerens, Patrice Latinne, and Christine Decaestecker. Adjusting the outputs of a classifier to new a priori probabilities: a simple procedure. *Neural Computation*, 14(1):21–41, 2002.

Filippo Santambrogio. *Optimal Transport for Applied Mathematicians*. Birkhäuser, Cham, 2015.

Vipul Singhal, Nigel Chou, Joseph Lee, Yifei Yue, Jinyue Liu, Wan Kee Chock, Li Lin, Yun-Ching Chang, Erica Mei Ling Teo, Jonathan Aow, et al. BANKSY unifies cell typing and tissue domain segmentation for scalable spatial omics data analysis. *Nature Genetics*, 56(3):431–441, 2024.

Daniel A Spielman and Nikhil Srivastava. Graph sparsification by effective resistances. *SIAM Journal on Computing*, 40(6):1913–1926, 2011.

Kai Sheng Tai, Peter Bailis, and Gregory Valiant. Sinkhorn label allocation: semi-supervised classification via annealed self-training. In *International Conference on Machine Learning*, pp. 10065–10075, 2021.

Jake Topping, Francesco Di Giovanni, Benjamin Paul Chamberlain, Xiaowen Dong, and Michael M Bronstein. Understanding over-squashing and bottlenecks on graphs via curvature. In *International Conference on Learning Representations*, 2022.

Vincent A Traag, Ludo Waltman, and Nees Jan Van Eck. From Louvain to Leiden: guaranteeing well-connected communities. *Scientific Reports*, 9(1):5233, 2019.

Aaron van den Oord, Oriol Vinyals, and Koray Kavukcuoglu. Neural discrete representation learning. In *Advances in Neural Information Processing Systems*, volume 30, 2017.

Olivier Veilleux, Malik Boudiaf, Pablo Piantanida, and Ismail Ben Ayed. Realistic evaluation of transductive few-shot learning. In *Advances in Neural Information Processing Systems*, volume 34, 2021.

Ron Zeira, Max Land, Alexander Strzalkowski, and Benjamin J Raphael. Alignment and integration of spatial transcriptomics data. *Nature Methods*, 19(5):567–575, 2022.

# A NOTATION

| Symbol | Meaning |
|---|---|
| $\mathcal{X} \subset \mathbb{R}^d$, $q$ | feature domain and its probability density |
| $K$, $c_k$, $\mu_k$ | number of programs, cost of program $k$, its prototype |
| $w \in \mathbb{R}^K$ | dual potentials ("voltages"), gauge $\mathbf{1}^\top w = 0$ |
| $\mathcal{C}_k(w)$, $M_k(w)$ | generalised Laguerre cell of program $k$ and its mass |
| $m \in \Delta^{K-1}$ | target masses (the balance / capacity constraint) |
| $B_{ij}$, $F_{ij}$ | interface between $i$ and $j$, and its defining function Equation (2) |
| $\gamma_{ij}$ | local margin $\inf_{B_{ij}} \|\nabla_x(c_i - c_j)\|$ |
| $h_{ij}$ | interface conductance, Equation (3) |
| $G_B$, $L_B$, $L_B^\dagger$ | region-adjacency graph, its Laplacian, its pseudo-inverse |
| $\lambda_2(L_B)$ | algebraic connectivity of $G_B$ |
| $b$ | injected current $\dot{m} - \partial_t M\|_w$, satisfies $\mathbf{1}^\top b = 0$ |
| $\dot{w}$, $V_{ij}(x)$ | potential velocity, normal interface velocity at $x$ (from $i$ into $j$) |
| $R_{ij}$ | effective resistance $(e_i - e_j)^\top L_B^\dagger (e_i - e_j)$ |
| $\Phi_{ij} = \sqrt{R_{ij}}/\gamma_{ij}$ | shift-agnostic susceptibility, Section 4 |
| $\varepsilon$, $p_{vk}$, $h_{ij}^\varepsilon$ | Sinkhorn temperature, soft assignment, entropic conductance |
| $t$, $\delta_{ij}$ | shift parameter, $\sup_{B_{ij}} \|\partial_t(c_i - c_j)\|$ |
| $\Delta n_{ij}$ | realised normal displacement of $B_{ij}$ |

# B PROOFS

## B.1 PROPOSITION 9

**Proposition 9** (Degenerate cases)**.** *If $w$ is constant, $\mathcal{C}_k(w)$ is the nearest-cost partition: for $c_k = \|x - \mu_k\|^2$ the Voronoi or nearest-centroid partition, and for $c_k = \|x - \mu_k\|^2 - r_k^2$ the power diagram whose interfaces are radical axes. Conversely a constrained partition reduces to the unconstrained one exactly when $m$ happens to equal the nearest-cost masses. Away from that coincidence the constraint is what makes the potentials unequal, and Theorem 2 shows it is also what couples them.*

If $w_k \equiv \omega$ for all $k$ then $c_k(x) - w_k \le c_j(x) - w_j$ reduces to $c_k(x) \le c_j(x)$, which is the nearest-cost rule. For $c_k(x) = \|x - \mu_k\|^2$ this is the Voronoi partition of $\{\mu_k\}$, and the assignment coincides with nearest-centroid. For $c_k(x) = \|x - \mu_k\|^2 - r_k^2$ the condition $\|x - \mu_i\|^2 - r_i^2 = \|x - \mu_j\|^2 - r_j^2$ is the definition of the radical axis of the two spheres, so the partition is the power diagram. Conversely, replacing $w_k$ by $w_k + \omega$ leaves Equation (1) unchanged, so the partition depends on $w$ only through its differences: equal potentials are exactly the unconstrained case. □

## B.2 LEMMA 1

This is classical (Aurenhammer et al., 1998; Kitagawa et al., 2019; De Gournay et al., 2019); we recall the one-line derivation. Increasing $w_j$ by $\eta$ lowers $c_j - w_j$, so the level set $\{F_{ij} = 0\}$ moves into $\mathcal{C}_i$. Since $F_{ij}$ is $C^1$ with $\|\nabla_x F_{ij}\| \ge \gamma_{ij} > 0$ on $B_{ij}$, the normal displacement at $x \in B_{ij}$ is $\eta / \|\nabla_x(c_i - c_j)(x)\| + o(\eta)$, so the mass transferred from $\mathcal{C}_i$ to $\mathcal{C}_j$ is $\eta \int_{B_{ij}} q / \|\nabla_x(c_i - c_j)\| \, d\mathcal{H}^{d-1} + o(\eta) = \eta h_{ij} + o(\eta)$. Junctions contribute at order $o(\eta)$ because they have zero $\mathcal{H}^{d-1}$ measure. Hence $\partial M_i / \partial w_j = -h_{ij}$ for $i \ne j$, and $\partial M_i / \partial w_i = \sum_{j \ne i} h_{ij}$ because $\sum_k M_k \equiv 1$ forces the rows of the Jacobian to sum to zero. That matrix is by definition the weighted Laplacian of $G_B$. □

## B.3 THEOREM 2

*Kirchhoff system.* By Lemma 1 and (A3), $(w, t) \mapsto M(w, t)$ is $C^1$ on a neighbourhood of $(w(0), 0)$ with $\partial_w M = L_B$. Restrict to the gauge hyperplane $\mathcal{G} = \{w : \mathbf{1}^\top w = 0\}$; on $\mathcal{G}$ the map $L_B|_{\mathcal{G}}$ is invertible because $G_B$ is connected (A2), so $\ker L_B = \operatorname{span}\{\mathbf{1}\}$. The constraint $G(w, t) := M(w, t) - m(t) = 0$ has $\partial_w G|_{\mathcal{G}}$ invertible, so the implicit function theorem gives a unique $C^1$ branch $t \mapsto w(t) \in \mathcal{G}$ with

$$L_B \dot{w} + \partial_t M\big|_w = \dot{m} \qquad \Longleftrightarrow \qquad L_B \dot{w} = \dot{m} - \partial_t M\big|_w = b.$$

Solvability requires $b \perp \ker L_B$, i.e. $\mathbf{1}^\top b = 0$; this holds because $\sum_k m_k(t) \equiv 1$ gives $\mathbf{1}^\top \dot{m} = 0$ and $\sum_k M_k(w,t) = \int_{\mathcal{X}} q^t \equiv 1$ gives $\mathbf{1}^\top \partial_t M|_w = 0$. In the gauge $\mathbf{1}^\top \dot{w} = 0$ the solution is $\dot{w} = L_B^\dagger b$.

*Interface velocity.* Fix $x_0 \in B_{ij}(0)$ with $\nabla_x F_{ij}(x_0, 0) \neq 0$ and let $n = \nabla_x F_{ij}/\|\nabla_x F_{ij}\|$. By the implicit function theorem applied to $s \mapsto F_{ij}(x_0 + sn, t)$ there is a $C^1$ curve $t \mapsto x(t) = x_0 + s(t)n + o(t)$ with $F_{ij}(x(t), t) = 0$ and $s(0) = 0$. Differentiating at $t = 0$, $\nabla_x F_{ij} \cdot \dot{x} + \partial_t F_{ij} = 0$, i.e. $V_{ij}\|\nabla_x F_{ij}\| = -\partial_t F_{ij}$ with $V_{ij} = \dot{s}$. Since $\partial_t F_{ij} = \partial_t(c_i - c_j) - (\dot{w}_i - \dot{w}_j)$, this is exactly the third identity in Equation (4). □

### B.4 Theorem 3

$L_B \succeq 0$ implies $L_B^\dagger \succeq 0$, so $\langle u, v\rangle_* = u^\top L_B^\dagger v$ is an inner product on $\mathbf{1}^\perp = \mathrm{range}(L_B)$. Both $e_i - e_j$ and $b$ lie in $\mathbf{1}^\perp$. Cauchy–Schwarz gives

$$|\dot{w}_i - \dot{w}_j| = |(e_i - e_j)^\top L_B^\dagger b| = |\langle e_i - e_j, b\rangle_*| \leq \|e_i - e_j\|_* \|b\|_* = \sqrt{R_{ij}}\, \|b\|_{L_B^\dagger},$$

with equality iff $b$ and $e_i - e_j$ are parallel in $\mathbf{1}^\perp$, i.e. $b = \lambda(e_i - e_j)$. Taking $b = e_i - e_j$ gives $\dot{w}_i - \dot{w}_j = R_{ij}$ and $\|b\|_{L_B^\dagger} = \sqrt{R_{ij}}$, so both sides equal $R_{ij}$ and the constant cannot be reduced. Substituting into Equation (4) and bounding $|\partial_t(c_i - c_j)| \leq \delta_{ij}$ and $\|\nabla_x(c_i - c_j)\| \geq \gamma_{ij}$ on $B_{ij}$ gives Equation (5). □

**Remark 10** (Hausdorff form)**.** If in addition $B_{ij}(s)$ is a compact embedded $C^1$ hypersurface for every $s \in [0, t]$, admits a tubular neighbourhood of uniform radius, and does not change topology on $[0, t]$, then the normal flow is a bijection between $B_{ij}(0)$ and $B_{ij}(t)$ and $d_{\mathrm{H}}(B_{ij}(t), B_{ij}(0)) \leq \int_0^t \sup_x |V_{ij}(x, s)|\, ds$. Mere rectifiability with $\gamma_{ij} > 0$ is not enough for this step, which is why Theorem 3 does not state it.

### B.5 Corollary 11

**Corollary 11** (Average case)**.** *If the injected current is isotropic in the energy inner product, $b = L_B^{1/2} u$ with $u \sim \mathcal{N}(0, I - \frac{1}{K}\mathbf{1}\mathbf{1}^\top)$, then $\mathbb{E}[(\dot{w}_i - \dot{w}_j)^2] = R_{ij}$, so for a cost-independent shift $\sqrt{\mathbb{E}V_{ij}^2} = \Phi_{ij}$: the susceptibility ranks interfaces by RMS response. For an unnormalised isotropic $b$ the exact predictor is instead $\widetilde{R}_{ij} = (e_i - e_j)^\top (L_B^\dagger)^2 (e_i - e_j)$, again purely global.*

Let $L_B = U\Lambda U^\top$ with $\Lambda_{11} = 0$, and $L_B^{1/2} = U\Lambda^{1/2}U^\top$. For $u$ isotropic on $\mathbf{1}^\perp$, $\mathbb{E}[uu^\top] = P := I - \frac{1}{K}\mathbf{1}\mathbf{1}^\top$ and $L_B^{1/2} P L_B^{1/2} = L_B$. With $b = L_B^{1/2} u$,

$$\mathbb{E}\big[(\dot{w}_i - \dot{w}_j)^2\big] = (e_i - e_j)^\top L_B^\dagger L_B^{1/2} P L_B^{1/2} L_B^\dagger (e_i - e_j) = (e_i - e_j)^\top L_B^\dagger L_B L_B^\dagger (e_i - e_j) = R_{ij},$$

using $L_B^\dagger L_B L_B^\dagger = L_B^\dagger$. Dividing by $\gamma_{ij}$ gives $\sqrt{\mathbb{E}V_{ij}^2} = \Phi_{ij}$ when $\partial_t c \equiv 0$. If instead $u$ is normalised to $\|u\| = 1$ the same computation gives $R_{ij}/(K-1)$, a per-partition constant times $R_{ij}$, so the ranking is unchanged. For an *unnormalised* isotropic $b$ the same argument gives $\mathbb{E}[(\dot{w}_i - \dot{w}_j)^2] = (e_i - e_j)^\top (L_B^\dagger)^2 (e_i - e_j) =: \widetilde{R}_{ij}$, which is again a purely global circuit quantity; we report both ensembles in Table 5. □

### B.6 Proposition 5

Take $b = e_i - e_j$ and $\partial_t c \equiv 0$. By Theorem 2, $\dot{w}_i - \dot{w}_j = (e_i - e_j)^\top L_B^\dagger (e_i - e_j) = R_{ij}$ and $V_{ij}(x) = (\dot{w}_i - \dot{w}_j)/\|\nabla_x(c_i - c_j)(x)\|$, so taking the supremum over $x \in B_{ij}$ and using $\inf_{B_{ij}} \|\nabla_x(c_i - c_j)\| = \gamma_{ij}$ gives $\overline{V}_{ij} = R_{ij}/\gamma_{ij}$ exactly; this is why the observable of $\overline{V}_{ij}$ is the supremum and not a reference-point velocity. Rayleigh monotonicity says that deleting edges cannot decrease an effective resistance; deleting every edge except $(i, j)$ leaves a single resistor of conductance $h_{ij}$, whose effective resistance is $1/h_{ij}$. Hence $R_{ij} \leq 1/h_{ij}$ and $|V_{ij}| \leq 1/(h_{ij}\gamma_{ij})$. Both $h_{ij}$ and $\gamma_{ij}$ are determined by $\mathrm{loc}_{ij}$ of Definition 4, so the bound is a local functional. For a general current the same is false: take $b$ supported on two cells neither of which is $i$ or $j$, so $b_i = b_j = 0$ and $\mathrm{loc}_{ij}$ does not move when $b$ is scaled, while $\dot{w}_i - \dot{w}_j$ scales with it. □

### B.7 THEOREM 6

*Graph-level statement, which is the route to an arbitrary $\rho$.* Fix the local data, so $h_{12} > 0$ and $\gamma_{12} > 0$ are fixed, and take $b = e_1 - e_2$ with $\partial_t c \equiv 0$, so $\overline{V}_{12} = R_{12}/\gamma_{12}$ exactly. Cutting every path between 1 and 2 other than the direct edge gives $R_{12} = 1/h_{12}$; adding $n$ node-disjoint two-edge paths of conductance $\kappa$ gives $R_{12} = (h_{12} + n\kappa/2)^{-1}$. The two worlds therefore satisfy $\overline{V}_{12}^{(\mathrm{H})}/\overline{V}_{12}^{(\mathrm{L})} = 1 + n\kappa/(2h_{12})$, which exceeds $\rho$ once $n \geq 2h_{12}(\rho - 1)/\kappa$, while $\mathrm{loc}_{12}$ is the same in both. Each of the two worlds is a fixed finite partition, so Assumption 1 holds for it with its own $m_{\min} > 0$ and $\lambda_2 > 0$; only $K$ grows with $\rho$. This is the construction the theorem quantifies over.

*Planar realisation, which is what Figure 1 runs.* Work on the unit square with uniform density $q \equiv 1$ and quadratic costs. Put $\mu_1 = (\frac{1}{2} - a, \frac{1}{2})$, $\mu_2 = (\frac{1}{2} + a, \frac{1}{2})$ and $w_1 = w_2 = 0$, so that $B_{12} \subset \{x_1 = \frac{1}{2}\}$ and $\gamma_{12} = 4a$ regardless of everything else. Add two "cap" programs $\mu_3 = (\frac{1}{2}, \frac{1}{2} + H)$, $\mu_4 = (\frac{1}{2}, \frac{1}{2} - H)$ with

$$w_3 = w_4 = H^2 - a^2 - 2H\beta. \tag{10}$$

On the line $x_1 = \frac{1}{2}$, writing $u = x_2 - \frac{1}{2}$, we have $c_1 - w_1 = a^2 + u^2$ and $c_3 - w_3 = (u - H)^2 - w_3$; equating gives $u = (H^2 - w_3 - a^2)/(2H) = \beta$ for *every* $H > 0$. The triple point 1–2–3 therefore sits at $x_2 = \frac{1}{2} + \beta$ and, by symmetry, 1–2–4 at $x_2 = \frac{1}{2} - \beta$, so

$$B_{12} = \{\tfrac{1}{2}\} \times [\tfrac{1}{2} - \beta, \tfrac{1}{2} + \beta], \qquad h_{12} = \frac{2\beta}{4a} = \frac{\beta}{2a},$$

both independent of $H$. Setting the target masses to $m := M(w)$ makes $w$ the exact mass-constrained solution by construction.

*What the near caps actually give.* It is tempting to write $R_{12}^{(\mathrm{L})} \leq (h_{12} + h_b)^{-1}$ with $h_b = \lim_{\beta \to 0} h_{13}$, but that is false, and the construction is more rigid than it looks. For quadratic costs every cost difference is affine, so each interface is a segment cut out of the square by a half-plane arrangement whose defining functions depend affinely on $\beta$ through Equation (10). When $B_{13}$ leaves through the top edge, as it does for the near caps, its vertical extent is $\frac{1}{2} - \beta$ and its normal is parallel to $\mu_3 - \mu_1 = (a, H)$, so $|B_{13}|/\gamma_{13} = (\frac{1}{2} - \beta)/(2a)$ and hence $h_{13}(\beta) = (\frac{1}{2} - \beta)/(2a)$ exactly, so

$$h_{12} + h_{13} \;=\; \frac{\beta}{2a} + \frac{\frac{1}{2} - \beta}{2a} \;=\; \frac{1}{4a} \qquad \text{independently of } \beta, \qquad \text{hence} \qquad R_{12}^{(\mathrm{L})} = 4a$$

for every $\beta$: the near-cap world has a *constant* effective resistance, which is why Figure 1a reads $R_{12} = \gamma_{12} = 0.80$. For the far caps $\gamma_{13} = 2\sqrt{a^2 + H^2} \to \infty$ with $|B_{13}| \leq \sqrt{2}$, so the four bridge conductances vanish and $R_{12}^{(\mathrm{H})} \to 1/h_{12} = 2a/\beta$. The ratio is therefore

$$\frac{R_{12}^{(\mathrm{H})}}{R_{12}^{(\mathrm{L})}} \;\longrightarrow\; \frac{2a/\beta}{4a} \;=\; \frac{1}{2\beta} \qquad (H \to \infty),$$

which exceeds any $\rho$ once $\beta < 1/(2\rho)$, while both worlds share $\mu_1, \mu_2, w_1, w_2$, the density, $B_{12}$, $\gamma_{12}$ and $h_{12}$ exactly.

*Assumption 1 along the planar family, stated honestly.* For each fixed $\beta > 0$ both worlds satisfy Assumption 1: every interface is a segment, $\gamma_{12} = 4a$ is constant, $G_B$ is connected because the caps touch both centre cells, and the cell masses are positive. The constants are *not* uniform in $\beta$. With far caps the centre cells are a band of height about $2\beta$, so $m_{\min} \to 0$, and $R_{12}^{(\mathrm{H})} \leq 2/\lambda_2$ forces $\lambda_2 \to 0$ in the high-response world: in this family large ratios and a degenerating spectrum arrive together, and the first-order velocities they compare are exactly the regime in which Theorem 8's constant grows. The graph-level construction above is the one that gives an arbitrary $\rho$ without degeneration, and Theorem 7 gives a third realisation on finite entropic samples with all masses bounded below. Section 5.1 measures the planar family: over $\beta \in [0.0025, 0.080]$, which moves $h_{12}$ by a factor $0.080/0.0025$, the near-cap $R_{12}$ stays at $4a$ to machine precision, the ratio grows from $5.4\times$ to $35.5\times$ (against $1/(2\beta)$ in the limit $H \to \infty$), the smallest cell mass falls to $0.0046$ and $\lambda_2(L_B)$ to $0.035$.

*From the ratio to the two impossibility statements.* Let $\mathcal{F}$ be a local functional and write $V_{\mathrm{L}} = |V_{12}^{(\mathrm{L})}| \leq V_{\mathrm{H}}/\rho$ with $V_{\mathrm{H}} = |V_{12}^{(\mathrm{H})}|$. Because the two worlds have the same $\mathrm{loc}_{12}$, $\mathcal{F}$ returns one value $F$ on both. (i) If $F$ is a point predictor and its multiplicative error $\max\{F/V, V/F\}$ were below $\sqrt{\rho}$ on both worlds, then $V_{\mathrm{H}} < \sqrt{\rho}\, F < \rho V_{\mathrm{L}} \leq V_{\mathrm{H}}$, a contradiction. (ii) If $F \geq V$ on both worlds then $F \geq V_{\mathrm{H}} \geq \rho V_{\mathrm{L}}$, so the bound over-estimates the low-response world by a factor at least $\rho$. □

### B.8 PROPOSITION 12

**Proposition 12** (The Laplacian survives an adjacency change)**.** *Let $t_0$ be a time at which the pair $(i, j)$ becomes adjacent, so $\mathcal{H}^{d-1}(B_{ij}(t)) \to 0$ as $t \downarrow t_0$. If $q$ is bounded above and $\gamma_{ij}$ below on a neighbourhood of the emerging interface, then $h_{ij}(t) \le \|q\|_\infty \gamma_{ij}^{-1} \mathcal{H}^{d-1}(B_{ij}(t)) \to 0$, so $L_B$ is continuous at $t_0$ even though $G_B$ gains an edge; the same holds for a vanishing interface. Continuity is not differentiability, so (A4) and the constant of Theorem 8 can still fail at $t_0$.*

By Equation (3) and the stated bounds, $h_{ij}(t) = \int_{B_{ij}(t)} q/\|\nabla_x(c_i - c_j)\|\, d\mathcal{H}^{d-1} \le \|q\|_\infty \gamma_{ij}^{-1} \mathcal{H}^{d-1}(B_{ij}(t))$, and the right-hand side tends to 0 as $\mathcal{H}^{d-1}(B_{ij}(t)) \to 0$. Every other entry of $L_B$ is a boundary integral over an interface that varies continuously, so $L_B$ is continuous at $t_0$; the new edge enters with weight 0. The same estimate applies to a vanishing interface. Differentiability can fail because $t \mapsto \mathcal{H}^{d-1}(B_{ij}(t))$ need not be $C^1$ at the birth time, which is why (A4) is assumed separately in Theorem 8. □

### B.9 THEOREM 7

*Jacobian.* With $p_{vk} = e^{(w_k - c_{vk})/\varepsilon} / \sum_l e^{(w_l - c_{vl})/\varepsilon}$ we have $\partial p_{vk}/\partial w_j = \varepsilon^{-1}(p_{vk}\delta_{kj} - p_{vk}p_{vj})$, so

$$\frac{\partial M_k^\varepsilon}{\partial w_j} = \frac{1}{\varepsilon} \sum_v q_v \big(p_{vk}\delta_{kj} - p_{vk}p_{vj}\big) = \begin{cases} -h_{kj}^\varepsilon \le 0, & k \ne j, \\ \frac{1}{\varepsilon} \sum_v q_v p_{vk}(1 - p_{vk}) = \sum_{j \ne k} h_{kj}^\varepsilon, & k = j, \end{cases}$$

using $\sum_j p_{vj} = 1$. This matrix is symmetric, has non-positive off-diagonals and zero row sums, hence is a weighted graph Laplacian. The *algebraic* arguments behind Theorems 2 and 3 and Proposition 5 use only those three properties and therefore transfer verbatim. The geometric construction of Theorem 6 does not: it builds hard interfaces, and in the entropic case every pair can carry positive co-assignment. Theorem 7 states what does transfer, and Section 5.1 realises it on finite samples.

*Limit.* **(A5)**: there are $\delta, \kappa > 0$ such that on the tube $\mathcal{T}_\delta = \{|F_{ij}| < \delta\}$ around $B_{ij}$ one has $\min_{l \ne i,j}(c_l - w_l) - \min(c_i - w_i, c_j - w_j) \ge \kappa$ and $\|\nabla_x F_{ij}\| \ge \gamma_{ij}/2$. Consider the continuum version $h_{ij}^\varepsilon = \varepsilon^{-1} \int_{\mathcal{X}} q\, p_i p_j$ and write $F = F_{ij}$.

*(i) Outside the tube.* On $\mathcal{X} \setminus \mathcal{T}_\delta$ we have $|F| \ge \delta$, so one of $p_i, p_j$ is at most $e^{-\delta/\varepsilon}$ and $p_i p_j \le e^{-\delta/\varepsilon}$. Since $q$ is a probability density, that part of the integral is at most $\varepsilon^{-1} e^{-\delta/\varepsilon} \to 0$.

*(ii) Inside the tube.* By (A5), for $l \notin \{i, j\}$, $e^{-(c_l - w_l)/\varepsilon} \le e^{-\kappa/\varepsilon} \min(e^{-(c_i - w_i)/\varepsilon}, e^{-(c_j - w_j)/\varepsilon})$, so the softmax denominator is $(1 + O(e^{-\kappa/\varepsilon}))$ times its two-term truncation and

$$p_i p_j = \tfrac{1}{4} \operatorname{sech}^2\Big(\frac{F}{2\varepsilon}\Big)\big(1 + O(e^{-\kappa/\varepsilon})\big) \qquad \text{uniformly on } \mathcal{T}_\delta.$$

By (A5) $\|\nabla_x F\| \ge \gamma_{ij}/2 > 0$ on $\mathcal{T}_\delta$, so the co-area formula applies with a locally bounded Jacobian factor and

$$\frac{1}{\varepsilon} \int_{\mathcal{T}_\delta} q\, \tfrac{1}{4} \operatorname{sech}^2\Big(\frac{F}{2\varepsilon}\Big) = \int_{-\delta}^{\delta} \phi_\varepsilon(s)\, G(s)\, ds, \qquad \phi_\varepsilon(s) = \frac{\operatorname{sech}^2(s/2\varepsilon)}{4\varepsilon},$$

where $G(s) = \int_{\{F = s\}} q \|\nabla_x F\|^{-1} d\mathcal{H}^{d-1}$. Here $\phi_\varepsilon \ge 0$ with $\int_{\mathbb{R}} \phi_\varepsilon = 1$ and $\int_{|s| > \eta} \phi_\varepsilon \to 0$ for every $\eta > 0$, so $(\phi_\varepsilon)$ is an approximate identity. $G$ is bounded on $[-\delta, \delta]$ (because $q$ is bounded, $\|\nabla_x F\|$ is bounded below, and the level sets have uniformly bounded $\mathcal{H}^{d-1}$ measure inside the tube) and continuous at $s = 0$ by (A2), so dominated convergence gives $\int \phi_\varepsilon G \to G(0) = h_{ij}$. Adding (i) and (ii), $h_{ij}^\varepsilon \to h_{ij}$. □

*Remark on (A5) and the junction set.* (A5) is an assumption, not a consequence of (A2): a uniform gap on a whole tube does not follow from the junction set being $\mathcal{H}^{d-1}$-null, because the closure of $B_{ij}$ reaches those junctions. The clean way to use (A2) is to excise them. Fix $\eta > 0$, let $B_{ij}^\eta$ be $B_{ij}$ minus the $\eta$-neighbourhood of the junction set, and apply the argument above on a tube around $B_{ij}^\eta$, where a uniform gap does hold by compactness; the discarded part contributes at most $\|q\|_\infty \gamma_{ij}^{-1} \mathcal{H}^{d-1}(B_{ij} \setminus B_{ij}^\eta)$ to both $h_{ij}$ and $h_{ij}^\varepsilon$, which tends to 0 with $\eta$ by (A2). Letting $\varepsilon \to 0$ and then $\eta \to 0$ gives the same conclusion without assuming a uniform gap up to the junctions. The boundedness and continuity of $G$ at $s = 0$ are likewise properties of the excised tube. Appendix C.7 measures the convergence numerically; the real systems never use this limit.

### B.10 THEOREM 8

(A6) is what licenses the second derivative: $\dot{w} = L_B^\dagger b$ with $L_B$ and $b$ of class $C^1$ and $L_B^\dagger$ smooth in $L_B$ on the set of connected Laplacians, so $w \in C^2$ on $[0, T]$. Note that (A3) alone does not give this, since it makes $(c, q, m)$ only $C^1$ and hence $b$ merely continuous; (A6) is implied by the same data being $C^2$. Differentiating $L_B\dot{w} = b$ gives $\dot{L_B}\dot{w} + L_B\ddot{w} = \dot{b}$; projecting on $\mathbf{1}^\perp$, $\ddot{w} = L_B^\dagger(\dot{b} - \dot{L_B}\dot{w})$ and $\|L_B^\dagger\|_2 = \lambda_2(L_B)^{-1}$, so

$$\|\ddot{w}(s)\| \;\le\; \frac{\|\dot{b}(s)\| + \|\dot{L_B}(s)\|\|\dot{w}(s)\|}{\lambda_2(L_B(s))} \qquad \text{for every } s \in [0, t].$$

Writing the remainder in integral form, $w(t) - w(0) - t\dot{w}(0) = \int_0^t (t - s)\ddot{w}(s)\, ds$, which is a vector identity and does not need a per-coordinate Lagrange point, and bounding the integrand by the supremum gives Equation (9). If $\|\dot{w}(0)\| \ge c > 0$ then the first-order displacement is $\Theta(t)$ and dividing gives the relative $O(t)$ statement; without that non-degeneracy the denominator can itself be $O(t^2)$ and no relative rate follows. Finally, if $\partial_t(c_i - c_j)$ is $C^1$ in $t$ and $\gamma_{ij} \ge \gamma > 0$, then $\overline{V}_{ij}$ is a $C^1$ function of $(\dot{w}, \partial_t c)$ with Lipschitz constant $\gamma^{-1}$ and inherits the same rate; the interface displacement statement used in Section 5.2 is this one. When an adjacency is created or destroyed, $h_{ij}$ leaves (or reaches) 0 with a one-sided derivative, $\dot{L_B}$ jumps, (A6) fails and the $O(t^2)$ remainder is no longer uniform; when a cell vanishes, $\lambda_2(L_B) \to 0$ and the bound blows up. □

## C IMPLEMENTATION AND EXPERIMENTAL DETAIL

### C.1 EXACT POLYGONAL TRANSPORT IN TWO DIMENSIONS

For costs of the form $c_k(x) = \|x - \mu_k\|^2 + a_k^\top x + s_k$ every $F_{ij}$ is affine in $x$, so Laguerre cells are convex polygons. We therefore avoid grids entirely: each cell is obtained by Sutherland–Hodgman clipping of the domain against the $K - 1$ half-planes $\{F_{ik} \le 0\}$, masses are fan-triangulated and integrated with a tensor Gauss–Legendre rule under the Duffy transform (order 10; order 16 for the normalisation constant), and each interface $B_{ij}$ is obtained *exactly* as the sub-interval of the line $\{F_{ij} = 0\}$ that survives intersection with the remaining half-planes, which reduces to a one-dimensional interval intersection. The conductance $h_{ij}$ is then a one-dimensional Gauss–Legendre integral (order 20) along that segment. Nothing in the observables is discretised.

The potentials are found by damped Newton on the concave semi-discrete dual, with the Newton matrix taken to be the geometric Laplacian of Lemma 1, never a finite difference, and a backtracking rule that halves the step until no cell has lost more than half the smallest mass and the residual has decreased. Tolerance is $\|M(w) - m\|_\infty < 10^{-11}$; the realised value across all synthetic experiments is at most $1.2 \times 10^{-15}$. Connectivity of $G_B$ (part of Assumption 1) is checked as $\lambda_2(L_B) > 10^{-5}$ and instances failing it are discarded; on real data the grounded solve falls back to a least-squares solution if the entropic Laplacian is numerically singular, which did not occur for any configuration reported here.

**Unit tests.** On random 8-cell partitions with a smooth non-uniform density: the analytic Gauss rule reproduces $\int_{[0,1]^2} e^x \sin 3y$ to $6.7 \times 10^{-16}$; masses sum to 1 to $10^{-12}$; the geometric Laplacian matches a central finite difference of the mass map to $2.1 \times 10^{-9}$ relative; the solver reaches $1.4 \times 10^{-15}$ mass error; the first-order law reproduces $\Delta w/t$ with relative error $1.17 \times 10^{-3}$ at $t = 10^{-3}$ falling to $3.50 \times 10^{-5}$ at $t = 3 \times 10^{-5}$, i.e. exactly $O(t)$; $R_{ij} \le 1/h_{ij}$ (Rayleigh) holds; and the bound of Theorem 3 is attained to $10^{-10}$ at $b = e_i - e_j$.

### C.2 THE TWIN-WORLD CONSTRUCTION (FIGURE 1)

Unit square, uniform density, $a = 0.20$, $\beta = 0.035$, $\mu_{1,2} = (\frac{1}{2} \mp a, \frac{1}{2})$ with $w_1 = w_2 = 0$; caps at $\mu_{3,4} = (\frac{1}{2}, \frac{1}{2} \pm H)$ with weights given by Equation (10); two spectator programs at $(0.12, \frac{1}{2})$ and $(0.88, \frac{1}{2})$ with weights $-0.02$, jittered by $\pm 0.03$ across configurations. Target masses are set to $m := M(w)$, which makes the configuration an exact mass-constrained solution by construction. $H$ is swept logarithmically over $[0.09, 8]$, giving 180 configurations in which $\gamma_{12}$ is identical to machine

zero and $h_{12}$ to $3.1 \times 10^{-14}$ relative, while $R_{12}$ spans $10.9\times$. Three forcings are applied: $b = e_1 - e_2$, a random zero-sum $b$, and $b = e_5 - e_6$ which touches only the two spectators.

### C.3 Synthetic library (Tables 4 and 5, Figure 2)

Table 4 draws 70 partitions per channel with $K \in \{4, 8, 16\}$ and three densities (uniform; a three-component Gaussian mixture; a linear gradient), rejecting configurations with a cell mass below $0.35/K$ or a disconnected adjacency graph. The perturbation magnitudes are $t \in \{10^{-4}, 3 \times 10^{-4}, 10^{-3}, 3 \times 10^{-3}, 10^{-2}, 3 \times 10^{-2}\}$; scatter plots are at $t = 10^{-3}$. Table 5 draws 96 partitions over six layouts (Poisson, chain, ladder, jittered grid, dumbbell, and two clusters joined by a two-cell bridge) with $K \in \{6, 10, 16, 24\}$, and applies 24 shifts per partition from each of the two ensembles of Corollary 11. The degree-preserving shuffle moves conductance around alternating 4-cycles, which leaves every weighted degree exactly unchanged; the permutation control permutes the conductances across edges; the wrong-graph control uses a 3-nearest-neighbour graph on the prototypes with conductance $1/\|\mu_i - \mu_j\|$.

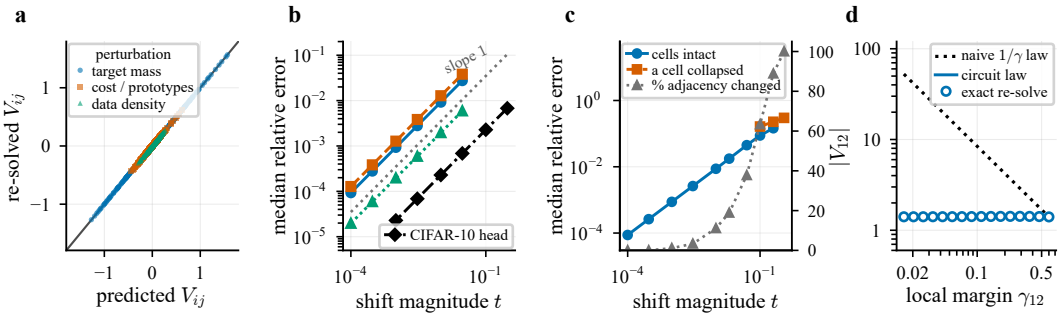


Figure 2: **The law, its error, and its operating regime.** (a) Predicted against exactly re-solved interface velocity, three perturbation channels. (b) Median relative error against shift magnitude: slope 1 on log-log axes for all three channels and for a controlled reweighting shift of a real Sinkhorn head. (c) The error is $\approx 0.870\,t$ whether or not adjacencies are created or destroyed, but roughly doubles once a cell collapses. (d) Driving $\gamma_{12} \to 0$ changes the naive $1/\gamma$ prediction by $37.5\times$ while the realised response changes by $1.014\times$; the circuit tracks the measured plateau to $2.7 \times 10^{-4}$.

### C.4 The two trapezoidal predictors

Three estimators of the same displacement appear throughout, and they differ only in which Laplacian they use. Write $L_B^0 = L_B(c^0, w^0)$ for the clean circuit, $\widehat{L_B}^1 = L_B(c^1, w^0)$ for the circuit re-evaluated on the *shifted* costs at the *frozen* clean potentials, and $L_B^1 = L_B(c^1, w^1)$ for the circuit of the re-solved partition. With $b$ the measured mass imbalance at frozen potentials,

$$\text{first-order law:} \quad \Delta w = (L_B^0)^\dagger b, \tag{11}$$

$$\text{predictive trapezoid:} \quad \Delta w = \big(\tfrac{1}{2}(L_B^0 + \widehat{L_B}^1)\big)^\dagger b, \tag{12}$$

$$\text{exact trapezoid (diagnostic):} \quad \Delta w = \big(\tfrac{1}{2}(L_B^0 + L_B^1)\big)^\dagger b. \tag{13}$$

Only Equations (11) and (12) are predictors: $\widehat{L_B}^1$ needs one pass over the shifted sample to recompute the co-assignment Gram matrix Equation (8) at the *clean* potentials and no solve, which is the same information $b$ already requires, whereas $L_B^1$ needs the answer. Equation (13) is reported only to separate the second-order term from prediction error. The interface displacement in each case is $(\Delta w_i - \Delta w_j - \Delta(c_i - c_j))/\gamma_{ij}$, with the cost term absent whenever the costs are frozen.

### C.5 Real systems

**CIFAR-10 balanced head.** Frozen ImageNet ResNet-18 (He et al., 2016) penultimate features of images upsampled to $112 \times 112$ (we use a supervised encoder for reproducibility on a laptop; a self-supervised one such as DINOv2 (Oquab et al., 2024) would only change the feature map,

not the circuit); feature standardisation and a 48-dimensional PCA basis fitted on 10,000 training images and frozen; class prototypes are the training-split class means in that basis. The deployment set is the first 5,000 test images with uniform target masses $m_k = 1/10$ and Sinkhorn temperature $\varepsilon = 2.42$ (two per cent of the median cost spread). Corruptions are the eight families of Appendix C.5 at six severities. The controlled reweighting shift of Section 5.2 multiplies each item's weight by $1 + t\xi_v$ with $\xi$ zero-mean and $\|\xi\|_\infty = 1$; it changes the deployment distribution without touching the encoder, which is why it can be made arbitrarily small.

**Oxford-IIIT Pet.** 200 test images at $256 \times 256$; patch features from the second residual stage of the same frozen encoder ($32 \times 32$ patches, $\ell_2$-normalised, projected onto a 24-dimensional PCA basis fitted on 60 held-out images), augmented with the patch coordinates scaled by 0.6. Each image is partitioned into $K = 6$ regions with uniform target masses, the prototypes coming from $k$-means on the clean image. The pixel trimaps that ship with the dataset are used only to restrict the pool to images that have one; no trimap-derived quantity is reported, because the object of the experiment is interface motion under corruption, which needs no ground-truth segmentation.

**HER2$^+$ breast cancer.** The resource of Andersson et al. (2021) contains 36 sections; 8 of them carry per-spot pathologist annotations, one per patient, and we restrict to the 12,689 genes common to all of them. For each held-out patient: counts are CPM-normalised and $\log 1p$-transformed, the top 2,000 variable genes are selected and a PCA basis fitted *on the other seven patients*; the first 20 components are used, and features are mixed with their lattice neighbours at weight 0.5. Class prototypes are the training-patient class means. 7 patients retain at least three annotated classes with at least five spots and enter the interface analysis. The target masses are the class proportions of the *training* patients, never of the held-out slide, so no test label enters the partition; this is also the only choice available at deployment. It is a harder target than the slide's own proportions, and reaching it needs the temperature continuation described in Appendix C. We do not report a segmentation comparison: with honest targets the balanced partition recovers pathologist labels no better than nearest centroid, which is unsurprising since it optimises mass balance rather than accuracy, and the object of this paper is interface *motion*. The two hyper-parameters (number of components, neighbourhood weight) were chosen by an inner leave-one-out over the first seven patients before the eighth was examined, and then frozen. Raw-count corruptions are Poisson resampling under log-normal library-size or gene-efficiency multipliers, Bernoulli dropout, a smooth Gaussian spatial artefact, and neighbourhood aggregation (a resolution change), each at four severities, with the whole pipeline re-run from counts.

Table 2: **Trained self-labelling head, and joint encoder drift** (Section 5.3). The same CIFAR-10 partition as Table 1, moved by training rather than by corruption, over 8 and 5 checkpoints. Prototype and encoder drift make $\partial_t(c_i - c_j)$ non-zero, the term the corruption experiments never exercise. Predictions use the clean circuit $L_B(0)$ only. "Cost drift only" keeps that term and drops the re-equilibration: it is the control that decides whether the circuit is needed.

| predictor | $\rho$ | sign | slope | $R^2$ |
|---|---|---|---|---|
| **Head training**: prototypes drift 0.8–27.0%, encoder frozen | | | | |
| full law Equation (4) | $+0.759$ | 85.0% | 0.64 | $+0.46$ |
| predictive trapezoid | $+0.803$ | 88.6% | 0.92 | $+0.81$ |
| cost drift only (no re-equilibration) | $-0.136$ | 48.1% | $-0.05$ | $-17.42$ |
| two-cell certificate Equation (6) | $+0.390$ | 55.6% | – | – |
| network term $\dot{w}_i - \dot{w}_j$ alone | $+0.995$ | 99.2% | 0.91 | $+0.97$ |
| **Joint drift**: encoder and prototypes trained together, combined drift 46.7–65.7% | | | | |
| full law Equation (4) | $+0.961$ | 92.9% | 0.63 | $+0.57$ |
| predictive trapezoid | $+0.993$ | 96.9% | 0.99 | $+0.98$ |
| cost drift only (no re-equilibration) | $-0.190$ | 44.9% | $-1.85$ | $-0.04$ |
| two-cell certificate Equation (6) | $+0.663$ | 78.2% | – | – |
| network term $\dot{w}_i - \dot{w}_j$ alone | $+0.966$ | 92.4% | 0.63 | $+0.59$ |

**Trained head and drifting encoder (Table 2).** The clean partition is the CIFAR-10 one above, reproduced bit for bit. Stage A trains the 10 prototypes only: balanced Sinkhorn targets $Q$ are computed once on the training features at the initial prototypes, and the prototypes then descend the

SwAV cross-entropy $-\sum_v \sum_k Q_{vk} \log p_{vk}(\mu)$ with Adam at learning rate $0.01$, checkpointed at $8$ step counts spaced geometrically so that the relative prototype drift covers $0.8\%$ to $27.0\%$. Stage B additionally unfreezes the last residual block of the encoder and descends the same objective on mini-batches of $64$ images with Adam at learning rate $10^{-4}$, BatchNorm kept in evaluation mode so that no running statistic drifts; the $5$ checkpoints re-extract the deployment features from scratch. At each checkpoint the injected current, the prediction and the prediction uses the clean circuit $L_B^0$ and the trapezoidal correction uses $\frac{1}{2}(L_B^0 + \widehat{L_B}^1)$ of Equation (12), never a re-solved circuit; ground truth is the re-solved balanced assignment at the drifted costs. The cost-drift term $\partial_t(c_i - c_j)$ is evaluated at one reference point per interface, the co-assignment barycentre $z^\star_{ij} = \sum_v q_v p_{vi} p_{vj} z_v / \sum_v q_v p_{vi} p_{vj}$ of the clean partition, which is the same weighting that defines $h^\varepsilon_{ij}$ in Equation (8). Reported statistics pool the checkpoints; the per-checkpoint mean is $0.857$ $[0.749, 0.947]$ for stage A and $0.966$ $[0.948, 0.983]$ for stage B.

**Entropic twin worlds (Section 5.1).** $6$ samples of $6{,}000$ points in the plane, anisotropic Gaussian with standard deviations $(0.6, 1.7)$, $K = 6$ programs at fixed $\mu_{1,2} = (\mp 0.45, 0)$, two fixed spectators at $(\pm 1.9, 0)$ and two that move to $(0, \pm D)$ for $D$ between $0.9$ and $6.0$; those two exchange data with programs 1 and 2 only at large $|y|$, away from the 1–2 midplane. Temperature $\varepsilon = 0.06$, target masses set self-consistently to $M^\varepsilon(0)$ so that $w = 0$ solves each instance. The remaining freedom is a single multiplier on the weight of the points in the slab $|x| < 0.22$, which we bisect to $10^{-13}$ so that $h^\varepsilon_{12}$ matches the reference world; $\gamma_{12}$ is exactly equal by construction. The comparison is therefore at fixed $\mathrm{loc}^\varepsilon_{12}$ in the sense of Definition 4.

**Warm-started Sinkhorn baseline (Section 5.3).** The clean state of the CIFAR-10 head is loaded from the run of Section 5.3, so the circuit columns of Table 1 are the same numbers, checked to $3 \times 10^{-15}$ relative. One update is $w \leftarrow w + \varepsilon \log(m / M^\varepsilon(w))$ followed by re-centring, the standard dual Sinkhorn step; it is diagonal in $k$ and uses only that cell's own mass residual. Timings are wall clock on the same laptop CPU for the whole grid of $48$ shifts, so they compare like with like; the circuit timing includes forming $\widehat{L_B}^1$.

**Corruption grid.** All image families share one normalised severity axis $a \in [0, 1]$ mapped to a per-family maximum: brightness $+0.45$, contrast reduction $0.65$, desaturation $1.0$, gamma $1{+}1.6$, Gaussian noise $\sigma = 0.18$, Gaussian blur $\sigma = 3$, JPEG quality down to $5$, and downsampling by $6\times$. Raw-count families use log-normal $\sigma$ up to $0.4$ (library size, gene efficiency), dropout up to $0.3$, spatial-artefact amplitude up to $0.6$ and neighbour mixing up to $0.6$.

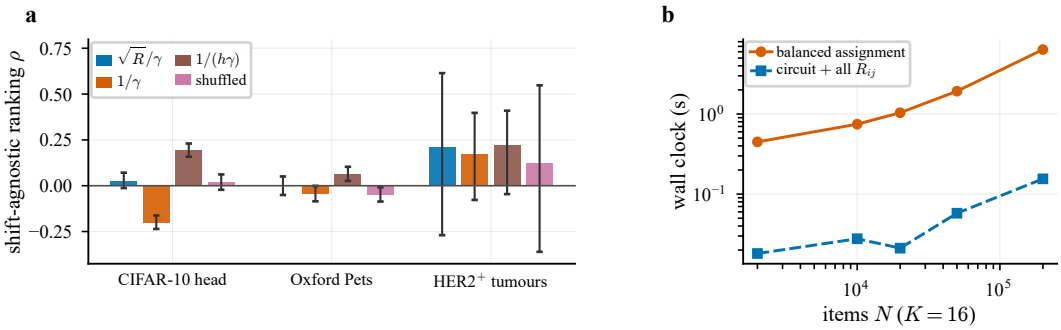


Figure 3: **The shift-agnostic negative, and the cost of the audit.** (a) Ranking *before* the shift is seen on the three real systems: the susceptibility $\Phi_{ij}$ against the local margin, the two-cell model and a degree-preserving shuffle, with $95\%$ bootstrap intervals over units. $\Phi_{ij}$ is not significant on any of the three, and the *local* two-cell score is ahead of it and significant on both vision systems, because real corruption ensembles are structured rather than energy-isotropic; what works across all three is the current-conditioned law of Table 1, which uses the measured shift. (b) Cost of the audit against the cost of the balanced assignment it audits.

### C.6 COST MEASUREMENTS

Table 3 and Figure 3b give the measured wall-clock cost of the audit; Figure 3a gives the shift-agnostic ranking discussed in Section 6.

Table 3: Wall-clock cost of the audit against the cost of the balanced assignment it audits, on a laptop CPU (median of three runs, $d = 32$ features).

| $N$ | $K$ | assignment (s) | circuit (s) | all $R_{ij}$ (s) | overhead |
|---|---|---|---|---|---|
| 2,000 | 16 | 0.449 | 0.0176 | 0.0005 | 4.0% |
| 10,000 | 16 | 0.747 | 0.0273 | 0.0004 | 3.7% |
| 50,000 | 16 | 1.928 | 0.0573 | 0.0004 | 3.0% |
| 200,000 | 16 | 6.387 | 0.1547 | 0.0004 | 2.4% |
| 20,000 | 4 | 0.652 | 0.0213 | 0.0003 | 3.3% |
| 20,000 | 8 | 0.745 | 0.0249 | 0.0003 | 3.4% |
| 20,000 | 16 | 1.039 | 0.0207 | 0.0004 | 2.0% |
| 20,000 | 32 | 1.585 | 0.0389 | 0.0007 | 2.5% |
| 20,000 | 64 | 2.959 | 0.0943 | 0.0635 | 5.3% |
| 20,000 | 128 | 4.427 | 0.1221 | 0.0636 | 4.2% |

### C.7 ADDITIONAL RESULTS

**The pinned-interface sweep (Section 5.1).** Shrinking the pinned segment holds the near-cap resistance at $0.800$ to $6.9 \times 10^{-16}$ relative while the response ratio grows to $35.5\times$, matching the $1/(2\beta)$ rate derived in Appendix B. The smallest cell mass falls to $0.0046$ over the same sweep: in this planar family large ratios and a degenerating spectrum arrive together, which is why Theorem 6 obtains an arbitrary $\rho$ from the graph-level construction instead.

**The law on real features (Section 5.2).** Reweighting the deployment sample of the CIFAR-10 head by $q_v \mapsto q_v(1 + t\xi_v)$ with $\xi$ a fixed random field, the relative error of Equation (4) falls from $6.8 \times 10^{-3}$ at $t = 0.3$ to $2.3 \times 10^{-5}$ at $t = 10^{-3}$, a log-log slope of $0.998$, and the entropic Laplacian Equation (8) matches a finite difference of the true entropic mass map to $2.2 \times 10^{-8}$ relative.

**Drift, per checkpoint (Section 5.3).** Under head training the per-checkpoint rank correlation of the full law falls from $0.998$ at $0.8\%$ drift to $0.604$ at $27.0\%$, and its slope from $1.00$ to $0.34$; the aggregate log-log error slope is $1.12$.

**The operating regime in detail (Section 5.2).** Over $90$ random partitions the median relative error of Equation (4) is $\approx 0.874\,t$ at $t = 10^{-4}$ and $\approx 0.895\,t$ at $t = 0.1$. At $t = 0.1$, $63\%$ of the instances have changed combinatorial type, which Theorem 8 excludes, yet their error is indistinguishable from that of the rest; the $9\%$ that lose most of a cell's mass have median error $0.164$ against $0.088$ for the others. In the margin sweep, driving $\gamma_{12} \to 0$ over a $37.5\times$ range moves the realised response by only $1.014\times$: for a current across the edge $\gamma_{ij}$ appears in $h_{ij}$ as well as in the velocity and the two cancel. That cancellation needs an affine cost difference and a current across the edge, and it sits in tension with $\Phi_{ij} \propto 1/\gamma_{ij}$, which answers the different question of an ensemble average.

**Residualised attribution (Section 5.2).** Against the local margin the per-partition gap is $0.090$ $[0.064, 0.119]$, and the bootstrap interval on the residualised resistance is $[0.488, 0.581]$. After residualising on $\log \gamma_{ij}$, $\log h_{ij}$ and $\log|B_{ij}|$, the permuted and wrong-graph resistances retain partial $\rho = 0.038$ and $0.038$. We report no partial statistic for $1/\gamma$: for an affine cost difference $\log h = \log|B| + \log \bar{q} - \log \gamma$, so $\log \gamma$ is nearly a combination of two of the controls (within-partition $R^2 = 0.94$) and its residual is noise.

**Choosing which programs to anchor (Section 5.3).** A practitioner who can *anchor* $B = \lceil K/4 \rceil$ of the $K$ programs, attenuating their share of the shift by $\kappa = 0.10$, must choose which. Corollary 11 makes the expected squared displacement an explicit function of $L_B^\dagger$, which gives a greedy rule; we compare it against anchoring the smallest margins, the smallest weighted degrees or the smallest masses and against a random choice, and score all five by *re-solving* the partition under sampled

isotropic shifts rather than by the energy-isotropic surrogate the circuit rule optimises. Over 60 partitions the circuit rule removes 34.1% of the displacement energy and the margin rule 19.0%, but the margin rule is the wrong yardstick, being worse than a random choice (22.9%). The honest comparison is with the min-degree rule at 31.4%, which the circuit beats by $0.028\ [0.014, 0.042]$ on 65% of instances (Wilcoxon $p < 0.001$): significant, and small. Figure 4d shows the residual energy against $B$.

**The entropic circuit converges to the geometric one.** For a fixed 8-cell partition of the unit square we compare the closed-form entropic conductances Equation (8), evaluated on a $700 \times 700$ sample, against the exact boundary integrals Equation (3) of the same partition. The median relative discrepancy falls monotonically from 0.263 at $\varepsilon = 3 \times 10^{-2}$ to 0.0002 at $\varepsilon = 3 \times 10^{-4}$, confirming Theorem 7; below that the sampling error of the finite grid dominates. The analytic entropic Laplacian matches a finite difference of the entropic mass map to $5.1 \times 10^{-9}$ relative at every temperature, and the first-order law holds on the entropic solver with log-log error slope 0.998.

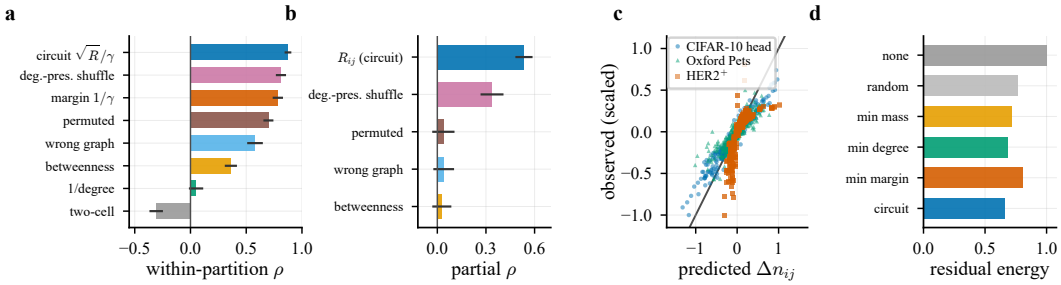


Figure 4: **Attribution, real systems, and use.** (a) Within-partition ranking correlation against eight controls, mean over 96 partitions, 95% bootstrap CIs. (b) Partial Spearman after removing $\log\gamma$, $\log h$, $\log|B_{ij}|$: the permuted and wrong-graph resistances fall to zero, the degree-preserving shuffle does not. (c) Predicted against observed displacement in the three real systems. (d) Residual displacement energy after anchoring $B$ programs.

Table 4: **The Boundary Circuit Law against exact re-solution, and the knockout.** Random mass-constrained partitions ($K \in \{4, 8, 16\}$, three densities), 4899 interfaces; sign accuracy is 99.8% or above throughout. "No re-equilibration" drops the $\dot{w}_i - \dot{w}_j$ term, "two-cell" is Equation (6) used as a predictor. The last row is not a fourth channel but the knockout: the same cost perturbation with the constraint switched *off*, where $\dot{w} \equiv 0$ and the purely local law is exact by Theorem 2.

| | | Circuit (Equation (4)) | | | $R^2$ of controls | |
|---|---|---|---|---|---|---|
| Perturbation channel | #int. | $R^2$ | slope | $\mathrm{d}\log\mathrm{err}/\mathrm{d}\log t$ | no re-equil. | two-cell |
| target mass $m(t) = m + tb$ | 1225 | 0.999998 | 0.9996 | 0.998 | −0.0028 | −2080.0 |
| cost $\mu_k(t) = \mu_k + td_k$ | 1218 | 0.999995 | 1.0001 | 1.000 | −2.2250 | −7773.9 |
| density $q_t = q + t\eta$ | 1238 | 1.000000 | 1.0000 | 1.000 | −0.0040 | −375.7 |
| cost shift, *constraint off* | 1218 | 0.999991 | 1.0003 | 0.997 | 1.0000 | 1.0 |

**All predictors.** Table 5 adds to the predictors discussed in Section 5.2 the inverse algebraic connectivity $1/\lambda_2(L_B)$, which is global but carries no edge information, and a constant. Both are uninformative within a partition, as they must be: $\lambda_2$ is the same number for every interface of a given partition, so its within-partition rank correlation is undefined. Pooled over all interfaces, where between-partition variation does let them move, they reach Spearman $+0.087$ and $-0.043$ against 0.875 for the circuit, so neither is a competitor even with the confound left in.

**Scope of the real-data shifts.** On the three corruption systems the cost functions are held fixed (frozen encoder, frozen prototypes) and the shift acts on the data distribution, so the $\partial_t(c_i - c_j)$ term of Equation (4) vanishes there. That term is exercised by the cost channel of Table 4 and, on real learned features, by Table 2, where the prototypes and then the encoder drift under training.

Table 5: **Mechanism attribution** (Section 5.2): every predictor, on 2745 interfaces over 96 partitions with $K \in \{6, 10, 16, 24\}$, scored by realised RMS displacement over 24 shifts and compared *within* each partition. The two column groups are the shift ensembles of Corollary 11; the last is the partial Spearman after regressing out $\log \gamma_{ij}$, $\log h_{ij}$ and $\log|B_{ij}|$. A dagger marks a predictor that is constant inside a partition, so that a within-partition rank correlation is undefined and the pooled value over all interfaces is given instead.

| | energy-isotropic | | isotropic | | partial $\rho$ |
|---|---|---|---|---|---|
| predictor | $\rho$ | $R^2$ | $\rho$ | $R^2$ | (local removed) |
| circuit $\sqrt{R_{ij}}/\gamma_{ij}$ | $+0.875$ | $+0.887$ | $+0.429$ | $+0.292$ | $+0.535$ [$+0.488, +0.581$] |
| circuit, isotropic form | $+0.493$ | $+0.302$ | $+0.739$ | $+0.700$ | $+0.418$ [$+0.369, +0.466$] |
| local margin $1/\gamma_{ij}$ | $+0.785$ | $+0.807$ | $+0.251$ | $+0.135$ | – |
| two-cell model $1/(h\gamma_{ij})$ | $-0.311$ | $+0.143$ | $+0.075$ | $+0.024$ | – |
| inverse interface size | $-0.354$ | $+0.163$ | $-0.033$ | $+0.050$ | – |
| inverse weighted degree | $+0.049$ | $+0.055$ | $+0.205$ | $+0.140$ | $-0.380$ [$-0.424, -0.334$] |
| shortest-path edge betweenness | $+0.364$ | $+0.193$ | $+0.078$ | $+0.035$ | $+0.028$ [$-0.024, +0.080$] |
| inverse algebraic connectivity $1/\lambda_2$ | $+0.087^\dagger$ | $+0.005^\dagger$ | $+0.121^\dagger$ | $+0.012^\dagger$ | – |
| degree-preserving conductance shuffle | $+0.814$ | $+0.844$ | $+0.346$ | $+0.241$ | $+0.339$ [$+0.273, +0.401$] |
| permuted conductances | $+0.701$ | $+0.683$ | $+0.227$ | $+0.152$ | $+0.038$ [$-0.025, +0.099$] |
| resistance of the wrong graph ($k$-NN) | $+0.579$ | $+0.542$ | $+0.215$ | $+0.113$ | $+0.038$ [$-0.018, +0.097$] |
| constant (no predictor) | $-0.043^\dagger$ | $-0.001^\dagger$ | $-0.069^\dagger$ | $-0.003^\dagger$ | – |

**Replication.** Each synthetic configuration is generated once from a fixed master seed; replication comes from the number of independent instances, not from repeated seeds on one instance: 180 configurations in Figure 1, 70 partitions per channel in Table 4, 96 partitions $\times$ 24 shifts in Table 5, 90 in Section 5.2 and 60 in Section 5.3. All reported intervals are bootstrap intervals over those instances (or over images / patients / corruption settings on real data). Image corruptions are deterministic given a severity except for Gaussian noise, which uses a fixed seed; raw-count corruptions use two seeds per family and severity.

**Interface triage at deployment.** Table 6 gives the full ranking task of Section 5.3: for every corruption setting of every real system, each candidate score ranks that system's interfaces and is compared against the ranking by realised displacement. The circuit wins on every system and every metric, and the two controls that destroy the circuit while keeping its ingredients (a shuffled circuit, a margin-only score) fall back to roughly chance at the top of the list.

Table 6: Interface triage: each score ranks a system's interfaces within one corruption setting, and is compared with the ranking by realised displacement. Top-$k$ is the fraction of the truly top-$k$ interfaces recovered; nDCG uses the realised displacement as gain.

| real system | ranking | top-1 | top-3 | nDCG |
|---|---|---|---|---|
| CIFAR-10 head | circuit | 68.8% | 73.6% | 0.970 |
| | circuit trapezoid | 89.6% | 84.0% | 0.992 |
| | two-cell | 2.1% | 6.2% | 0.826 |
| | margin | 0.0% | 0.0% | 0.736 |
| | shuffled | 0.0% | 9.0% | 0.779 |
| Oxford Pets | circuit | 63.4% | 75.0% | 0.976 |
| | circuit trapezoid | 74.1% | 82.3% | 0.989 |
| | two-cell | 8.1% | 33.9% | 0.853 |
| | margin | 12.7% | 24.5% | 0.795 |
| | shuffled | 11.3% | 23.3% | 0.792 |
| HER2+ tumours | circuit | 56.0% | 78.5% | 0.990 |
| | circuit trapezoid | 46.3% | 72.6% | 0.986 |
| | two-cell | 9.1% | 41.1% | 0.907 |
| | margin | 22.3% | 43.8% | 0.874 |
| | shuffled | 21.1% | 41.5% | 0.857 |

**Per-severity behaviour on real systems.** Table 7 reports, for the CIFAR-10 head and the tumour sections, the slope and $R^2$ of predicted against observed displacement at each severity, for the first-order law and for its trapezoidal version that uses the midpoint Laplacian $\frac{1}{2}(L_B^0 + L_B^1)$. The midpoint version recovers part of the deficit, which identifies the residual as the second-order term of Theorem 8, the shift changes the circuit as well as the potentials, rather than as a failure of the mechanism.

Table 7: Per-severity behaviour of the first-order law and of its midpoint (trapezoidal) form on the two systems for which a severity axis is defined.

| | | | first order | | predictive trapezoid | | exact trapezoid | | |
|---|---|---|---|---|---|---|---|---|---|
| system | sev. | $\|b\|_{L^\dagger}$ | slope | $R^2$ | slope | $R^2$ | slope | $R^2$ | viol. |
| CIFAR-10 head | 0.05 | 0.125 | 0.67 | $+0.745$ | 0.71 | $+0.807$ | 0.89 | $+0.946$ | 0.0% |
| | 0.10 | 0.245 | 0.68 | $+0.735$ | 0.67 | $+0.723$ | 0.90 | $+0.915$ | 0.0% |
| | 0.20 | 0.862 | 0.67 | $+0.692$ | 0.73 | $+0.845$ | 0.85 | $+0.891$ | 0.0% |
| | 0.35 | 1.433 | 0.45 | $-0.377$ | 0.56 | $+0.328$ | 0.90 | $+0.824$ | 0.0% |
| | 0.60 | 2.627 | 0.33 | $-1.816$ | 0.39 | $-1.264$ | 1.04 | $+0.757$ | 0.0% |
| | 1.00 | 7.149 | 0.32 | $-2.048$ | 0.28 | $-4.039$ | 1.15 | $+0.645$ | 0.0% |
| | all | – | 0.36 | $-1.313$ | 0.35 | $-1.852$ | 1.01 | $+0.744$ | 0.0% |
| HER2$^+$ tumours | 0.1 | 0.381 | 0.86 | $+0.507$ | 1.03 | $+0.847$ | 0.87 | $+0.457$ | 17.6% |
| | 0.25 | 0.380 | 0.96 | $+0.437$ | 1.19 | $+0.765$ | 0.98 | $+0.400$ | 18.7% |
| | 0.5 | 0.390 | 0.96 | $+0.398$ | 1.21 | $+0.723$ | 1.00 | $+0.387$ | 21.3% |
| | 0.75 | 0.389 | 0.90 | $+0.327$ | 1.17 | $+0.631$ | 0.90 | $+0.276$ | 21.7% |
| | 1.0 | 0.390 | 0.78 | $+0.295$ | 1.03 | $+0.638$ | 0.83 | $+0.343$ | 22.0% |
| | all | – | 0.88 | $+0.372$ | 1.12 | $+0.697$ | 0.90 | $+0.357$ | 20.3% |