

SYMMETRY AND GEOMETRY IN NEURAL REPRESENTATIONS

The Symmetry Budget

How much group structure should a network architecture assume?

Noor Islam S. Mohammad^{1,*} Mahmudul Hasan² Md. Basim Al Zahir Shammo³ Hasan Siddiki⁴ Jakaria Habib³

¹Istanbul Technical University ²Deakin University ³Pabna University of Science and Technology ⁴American International University–Bangladesh

*Casual Language and Multimodal Intelligence Lab · islam23@itu.edu.tr

An architecture is a symmetry assumption

Layer	Weights tied across
Convolution	translations
Group conv.	a finite group G
Deep sets	permutations
Dense	nothing

Tying buys **fewer parameters** at the price of an **assumption** that is almost never exactly true.

Images are not exactly shift-invariant;
dynamics are not exactly symmetric once
boundaries enter.



Everyone agrees the two ends exist.
Nobody can say where to stand.

Our question. Given n samples and a symmetry that fails by ε : *which* subgroup should the architecture tie over, and how sharply does that answer move with n ?

A model where the answer is computable

Linear maps on signals over the cyclic group C_d — architectures indexed by subgroups

ARCHITECTURE FAMILY

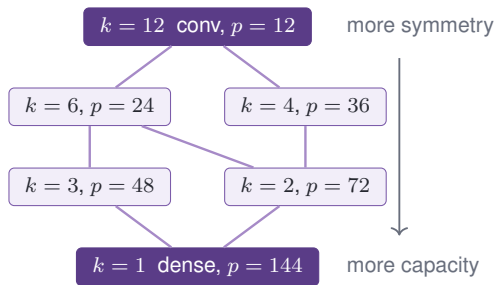
$$\mathcal{V}_k = \{ W \in \mathbb{R}^{d \times d} : WS^{d/k} = S^{d/k}W \}, \quad k \mid d$$

$$\dim \mathcal{V}_k = p_k = d^2/k$$

DATA

$$y = W^*x + \sigma\xi, \quad W^* = \underbrace{W_{\text{eq}}}_{\in \mathcal{V}_d} + \varepsilon \underbrace{R}_{\perp \mathcal{V}_d}$$

One integer k dials the architecture from an **unconstrained layer** (d^2 params) to a **circular convolution** (d params); ε is the task's **symmetry defect**.



Divisor lattice of $d = 12$: six architectures, nested as $\mathcal{V}_d \subseteq \mathcal{V}_k \subseteq \mathcal{V}_1$.

Every term is a closed form, not a bound

FINITE-SAMPLE EXCESS RISK (PROP. 3)

$$\mathbb{E} \left\| \hat{W}_k - W^\star \right\|_F^2 = \underbrace{\delta_k^2}_{\text{approximation}} + \underbrace{\frac{\sigma^2 p_k}{n}}_{\text{estimation}} + \underbrace{\frac{\delta_k^2 \text{tr}(\Pi_k)}{n}}_{\text{coupling (finite-}n\text{)}} + o(n^{-1})$$

GENERIC SYMMETRY BREAKING: R UNIFORM ON THE SPHERE OF $\mathcal{V}_d^\perp$ (PROP. 4)

$$\mathbb{E}_R \delta_k^2 = \varepsilon^2 \gamma_k, \quad \gamma_k = \frac{d}{d-1} \left(1 - \frac{1}{k} \right) \quad \text{— a pure dimension count}$$

k	1	2	3	4	6	12
$p_k = d^2/k$	144	72	48	36	24	12
γ_k closed form	0.0000	0.5455	0.7273	0.8182	0.9091	1.0000
γ_k measured (40 seeds)	0.0000	0.5414	0.7373	0.8170	0.9155	1.0000

Averaging over a subgroup destroys the symmetry-breaking signal at rate γ_k while returning only a $1/k$ fraction of the parameter savings. $d = 12, \varepsilon = 0.8$; agreement within 1.4%.

The lattice collapses to its endpoints

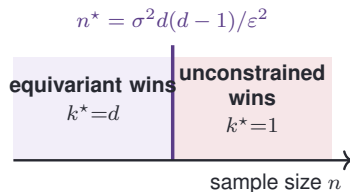
Theorem 5 and Corollary 6 — stated at finite n

$$R_{\text{fs}}(u) = \underbrace{\varepsilon^2 \frac{d}{d-1} (1-u) + \frac{\sigma^2 d^2}{n} u}_{\text{affine in } u=1/k} + \underbrace{\frac{\varepsilon^2 d^2}{n} \frac{d}{d-1} (1-u)u}_{\text{concave correction}} + o(n^{-1})$$

Theorem 5. Under generic symmetry breaking the minimizer over the lattice is an **endpoint** — $k=d$ or $k=1$ — for all $(n, \varepsilon, \sigma, d)$ outside a measure-zero boundary.

Corollary 6. The endpoints exchange rank at

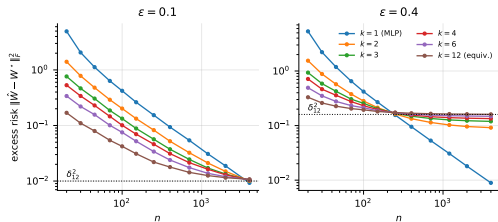
$$n^*(\varepsilon) = \frac{\sigma^2 d(d-1)}{\varepsilon^2} \iff \log n^* = -2 \log \varepsilon + c$$



No interior subgroup, at any sample size.

A phase transition — not a gradual interpolation.

The prediction is parameter-free, and it lands



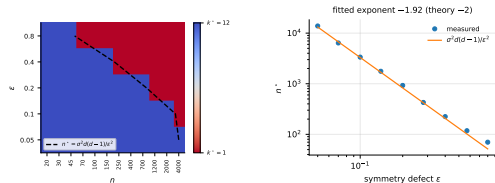
Convolution plateaus at its floor δ_{12}^2 ; dense descends without one. A $4\times$ defect, a $16\times$ earlier crossover.

60/60 & 48/48

cells predicted exactly ($d = 12, 16$)

-1.92 ± 0.07

fitted exponent vs. predicted -2



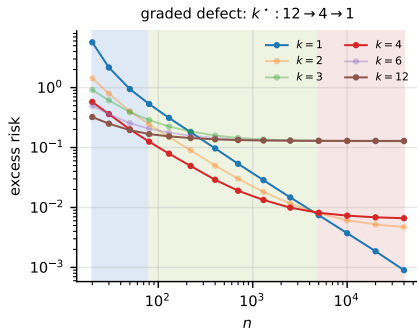
Measured k^* over the (ϵ, n) grid, with the parameter-free boundary overlaid. Only the two endpoints ever appear.

never interior

realized $k^* \in \{1, 12\}$ in every cell

Theorem 5 predicts its own failure mode

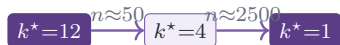
Give the defect structure and the lattice reopens



GRADED TARGET (PROP. 7)

$W^* = u_{12} + b u_4 + c u_1$, $(b, c) = (0.35, 0.08)$ — far from convolutional, close to C_4 -equivariant.

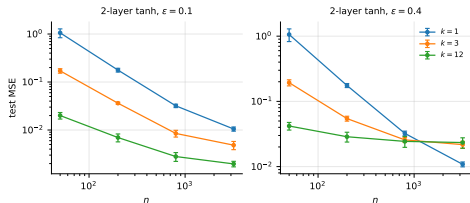
The optimum **descends the lattice in discrete steps**:



Predicted and measured k^* agree on **14/14** sample sizes, and the measured window matches the closed form: **two decades of n** , not a knife edge.

The practical reading. If the target has a genuine, smaller symmetry group, architecture selection by risk *finds that group* before falling through to no symmetry at all.

What we checked, and what we did not



Two-layer tanh networks, gradients projected onto $\mathcal{V}_k$ at every step (10 seeds, 8 sample sizes). The crossover survives nonlinearity and moves with ϵ — a qualitative control, not a scaling law.

HONEST LIMITS

- Linear model, **cyclic** group, Gaussian inputs, homoscedastic noise — what makes Props. 1–7 exact.
- Collapse rests on γ_k , a count valid for **one-dimensional irreps**. Prop. 8 says exactly when it survives for non-abelian groups — and when it fails.
- ϵ^{-2} holds at fixed σ ; structured noise replaces σ^2 by a k -dependent level. **Not** a claim about real vision architectures.
- OPEN: estimating δ_k from a real dataset — CIFAR-10 under translation — rather than constructing it.

Three things to take away

- 1 Architecture choice is a choice of **subgroup**, and in the cyclic linear model the trade-off is **computed, not bounded**: $n^* = \sigma^2 d(d-1)/\varepsilon^2$, verified at $d = 8, 12, 16$.
- 2 Under **generic** symmetry breaking, partial weight tying is **not a hedge**: the lattice collapses to its endpoints at every sample size. Sharp transition, not smooth interpolation.
- 3 Under **structured** breaking the interior returns. The useful question is not *how much* symmetry to assume, but **which symmetry the defect leaves intact**.

Code, data, and a per-claim validation harness (47 unit tests, full protocol in < 4 CPU-minutes):

github.com/CLMI-Lab/symmetry-budget

islam23@itu.edu.tr