Learning Chern Numbers of Multiband Topological Insulators with Gauge Equivariant Neural Networks

Longde Huang, Oleksandr Balabanov, Hampus Linander, Mats Granath
Daniel Persson, Jan E. Gerken

CHALMERS UNIVERSITY OF TECHNOLOGY

UNIVERSITY OF GOTHENBURG

WASP WALLENBERG AI, AUTONOMOUS SYSTEMS AND SOFTWARE PROGRAM

NeurIPS 2025

NEURAL INFORMATION PROCESSING SYSTEMS

Problem Setting

Topological Insulators

Topological Insulators: Materials that behave as insulators in their bulk but allow current to flow along their boundaries, due to underlying topological invariants (**Chern numbers**).

Tensor formulation

- The material: A $2m$ dimensional periodic grid Λ (Setting $m = 1$ for now)
  $\Rightarrow$ Spatial dimension $N_1 \times N_2 \times \cdots \times N_{2m} = N_{site}$;

Tensor formulation

- The material: A $2m$ dimensional periodic grid Λ (Setting $m = 1$ for now)
  $\Rightarrow$ Spatial dimension $N_1 \times N_2 \times \cdots \times N_{2m} = N_{site}$;
- Band: A $U(N)$ frame over M: $\{(U_k^x, U_k^y, W_k)\}_{k \in \Lambda}$
  $\Rightarrow$ Channel dimension $C(m) \times N \times N \times 2$;

Tensor formulation

- The material: A $2m$ dimensional periodic grid Λ (Setting $m = 1$ for now)
  $\Rightarrow$ Spatial dimension $N_1 \times N_2 \times \cdots \times N_{2m} = N_{site}$;
- Band: A $U(N)$ frame over M: $\{(U_k^x, U_k^y, W_k)\}_{k\in\Lambda}$
  $\Rightarrow$ Channel dimension $C(m) \times N \times N \times 2$;
- Chern Number: An integer $\tilde{C}$:

$$\tilde{C}\Big(\{W_k \,|\, k \in \Lambda\}\Big) = \frac{1}{2\pi} \sum_{k\in\Lambda} \operatorname{Im} \operatorname{Tr} \log W_k \,. \tag{1}$$

Tensor formulation

- The material: A $2m$ dimensional periodic grid Λ (Setting $m = 1$ for now)
  $\Rightarrow$ Spatial dimension $N_1 \times N_2 \times \cdots \times N_{2m} = N_{site}$;
- Band: A $U(N)$ frame over M: $\{(U_k^x, U_k^y, W_k)\}_{k\in\Lambda}$
  $\Rightarrow$ Channel dimension $C(m) \times N \times N \times 2$;
- Chern Number: An integer $\tilde{C}$:

$$\tilde{C}\Big(\{W_k \,|\, k \in \Lambda\} \Big) = \frac{1}{2\pi} \sum_{k\in\Lambda} \mathrm{Im}\,\mathrm{Tr}\,\log W_k \,. \tag{1}$$

- Gauge symmetry:

$$\tilde{C}\Big(\{W_k \,|\, k \in \Lambda\} \Big) = \tilde{C}\Big(\{\Omega_k^\dagger W_k \Omega_k \,|\, k \in \Lambda\} \Big) \quad \forall \{\Omega_k\}_{k\in\Lambda} \in \mathrm{U}(N)^{N_x \times N_y} \,. \tag{2}$$

Large Non-Abelian Gauge Symmetry

Gauge Symmetry

$$U_k^\mu \sim \Omega_k^\dagger U_k^\mu \Omega_{k+e_\mu}, \quad W_k \sim \Omega_k^\dagger W_k \Omega_k \tag{3}$$

- **Non-Commutative**: Unitary Matrix group.
- **Local**: One Unitary group at each grid point; traditional Lie group equivariance techniques (data augmentation, GCNN, etc.) fail.

GEBLNet: Local Gauge–Equivariant Network

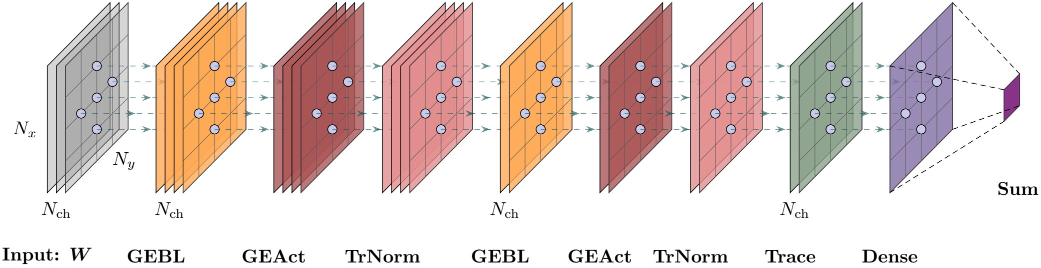


Figure: Architecture of GEBLNet [*Favoni et al. 2022*]. Each circle represents an independent grid point.

GEBLNet: Local Gauge–Equivariant Network

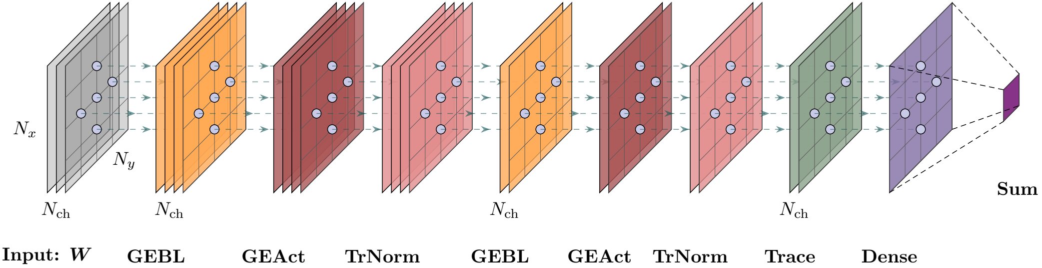


Figure: Architecture of GEBLNet [*Favoni et al. 2022*]. Each circle represents an independent grid point.

- **Purely local processing**: every point k is handled in isolation; no coupling between neighbours during feature extraction.
- **Gauge equivariance**: each layer manifestly preserves the gauge transformation and retains the matrix form outputs.

Implementation

- **Input:** $(W_k^\lambda) \in U(N)^{N_{\text{in}} \times N_{\text{site}}}$
- **Step 1:** Append identity matrix I to the input set
- **Step 2:**

$$W_k'^{\gamma} = \sum_{\mu,\nu} \alpha_{\gamma\mu\nu} \, W_k^\mu \, W_k^\nu$$

Intuition:

Constructs higher-order terms W^n from lower-order ones and form a polynomial.

Universal Approximation Theorem with GEBL

Implementation

- **Input:** $(W_k^\lambda) \in U(N)^{N_{\text{in}} \times N_{\text{site}}}$
- **Step 1:** Append identity matrix I to the input set
- **Step 2:**

$$W_k'^{\gamma} = \sum_{\mu,\nu} \alpha_{\gamma\mu\nu} \, W_k^{\mu} \, W_k^{\nu}$$

Intuition:

Constructs higher-order terms W^n from lower-order ones and form a polynomial.

Theorem (UAT for GEBLNet)

Let G be a compact Lie group. Assume the GEAct non-linearity is $\sigma = \tilde{\sigma} \circ \mathrm{Re}$ with $\tilde{\sigma}$ bounded and non-decreasing. Then GEBLNet can approximate any square-integrable conjugation invariant function $F \in L^2_{\text{class}}(G)$ to arbitrary accuracy.

Training Setup

Loss function

Given network output $f(W)$ and discrete Chern number $\tilde{C}$:

$$L_{\mathrm{g}} = \left\| f(W) - \tilde{C} \right\|_1, \qquad L_{\mathrm{std}} = \left\| \min\Big(\mathrm{std}\Big\{ g(W_k) \Big\}, \delta\Big) - \delta \right\|_1,$$

with $\delta = 0.5$.

$$\boxed{L_{\mathrm{total}} = L_{\mathrm{g}} + L_{\mathrm{std}}}$$

L_{std} prevents the network from collapsing to zero-outputs by enforcing point-to-point variation.

Model Capability Test: Band Number

Table: **Accuracy on a 5×5 grid.** GEBLNet is trained and evaluated for increasing band count N.

Bands N	4	5	6	7	8
Accuracy (%)	95.9	94.0	93.8	91.7	52.5

- High accuracy is retained up to $N = 7$ bands.

Training on Trivial Samples: Problem Statement

Setting

- Grid 5×5, $N = 4$ bands
- Only *topologically trivial* samples ($C = 0$)
- Need L_{std}

Result: GEBLNet without TrNorm fails

- Most seeds $\to$ outputs ≈ 0

Trace Explosion

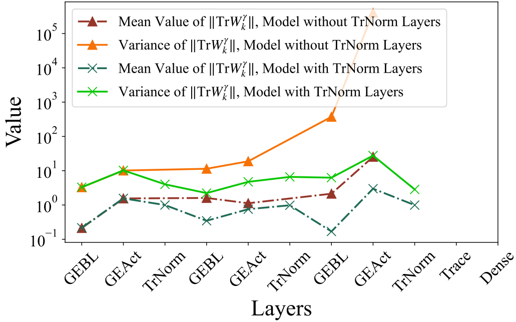


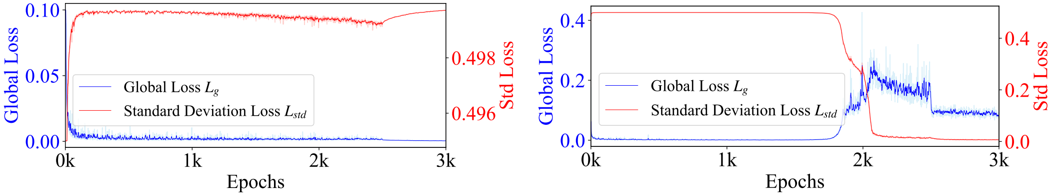
Figure: Statistics of trace per layer, with or without TrNorm

Trace Normalization (TrNorm)

$$W_k'^\gamma = \frac{W_k^\gamma}{\max(\varepsilon, \operatorname{mean}_\gamma\{|\operatorname{tr} W_k^\gamma|\})}$$

- Normalization over each channel so that $\operatorname{mean}_\gamma\{|\operatorname{tr} W_k^\gamma|\}$.
- Cancels trace explosion after every GEBL layer.

TrNorm Fixing the collision



Loss curves, seed 83. Left: no TrNorm (collapse). Right: with TrNorm (converges).

- **Turning point** (std. loss < 0.49) occurs between 1.3 k–1.9 k epochs for 4/5 seeds.
- Post-rescale accuracy on non-trivial data $92\% - 95\%$ (table below).

Generalization: Higher Dimensional Models

Higher Order Chern Number

For general $2M$ dimensional multiband system, the discretized Mth Chern number could be defined as

$$\tilde{C}_M = \frac{M!}{(2M)!(\pi i)^M} \sum_k \sum_{\mu_1,\ldots,\mu_{2M}} \mathrm{Tr}\epsilon_{\mu_1,\ldots,\mu_{2M}} \prod_{t=1}^{M} \log W_k^{\mu_{2t-1},\mu_{2t}}. \tag{4}$$

Generalization: Higher Dimensional Models

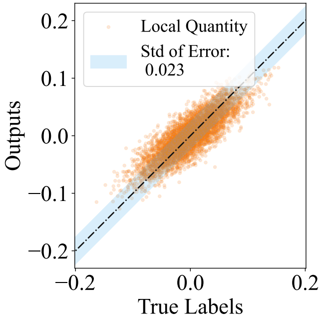
Higher Order Chern Number

For general $2M$ dimensional multiband system, the discretized Mth Chern number could be defined as

$$\tilde{C}_M = \frac{M!}{(2M)!(\pi i)^M} \sum_k \sum_{\mu_1,\ldots,\mu_{2M}} \mathrm{Tr} \epsilon_{\mu_1,\ldots,\mu_{2M}} \prod_{t=1}^{M} \log W_k^{\mu_{2t-1},\mu_{2t}}. \tag{4}$$

- For M=2 (four spatial dimensions) each point carries $\binom{4}{2} = 6$ loops (channels): $W_k^{\mu\nu}$.
- The approximation is not necessarily integer; evaluation metric is then mean absolute error (MAE), not exact accuracy.

Generalization: Higher Dimensional Models



Local quantity-truth comparison

- Points lie close to the line $y = x$: model reproduces the *local* topological quantity with high precision.
- Global MAE≈ 0.25: well within rounding error of the true second Chern number.

Generalization: Higher Dimensional Models

Theorem (Higher-Dimensional UAT for GEBLNet)

Let G be the unitary group $U(M)$. Under the same assumptions on the nonlinearity, GEBLNet can approximate arbitrarily well any square-integrable function $f \in L^2(G^K)$ such that

$$f(g_1, \ldots, g_K) = f(hg_1h^{-1}, \ldots, hg_Kh^{-1}), \quad \forall h \in G.$$

Conclusion

Our Model

GEBLNet: a gauge-equivariant neural network for predicting Chern numbers of multiband topological insulators.

Key contributions

- *Theory:* universal-approximation theorem: a sufficiently wide GEBLNet can represent *any* gauge-invariant function (incl. Chern number).
- *Practice:* high accuracy up to $N = 7$ bands.
- *Architecture:* introduced **TrNorm** layer; removes trace explosion and enables learning from topologically trivial data alone.
- *Beyond 2D:* extended to 4D grids; MAE ≈ 0.25 on the second Chern number.

Limitations and Future Work

Current gaps

- 4D discretization is approximate, not strictly integer-valued.
- Architecture explored is plain GEBL stacks; no residuals, recurrence, or attention.
- Generalization to other Lie groups untested.

Next steps

- Exact, gauge-invariant discretisations for higher Chern classes.
- Deeper or transformer-style equivariant blocks for greater expressivity.
- Broader benchmarks: other symmetry groups, experimental data.

Thank you!