Kohn–Sham Spectral Embedding on Sparse Graphs at the Nishimori Temperature for Image Classification

Классификация изображений с помощью спектрального вложения Кона-Шэма на разряженных графах при температуре Нисимори

Usatyuk V., Researcher, South-West State University
Sapoznikov D., Senior Research Engineer, T8 LTD
Egorov S., Professor, South-West State University

The 9th International Conference on Deep Learning in Computational Physics

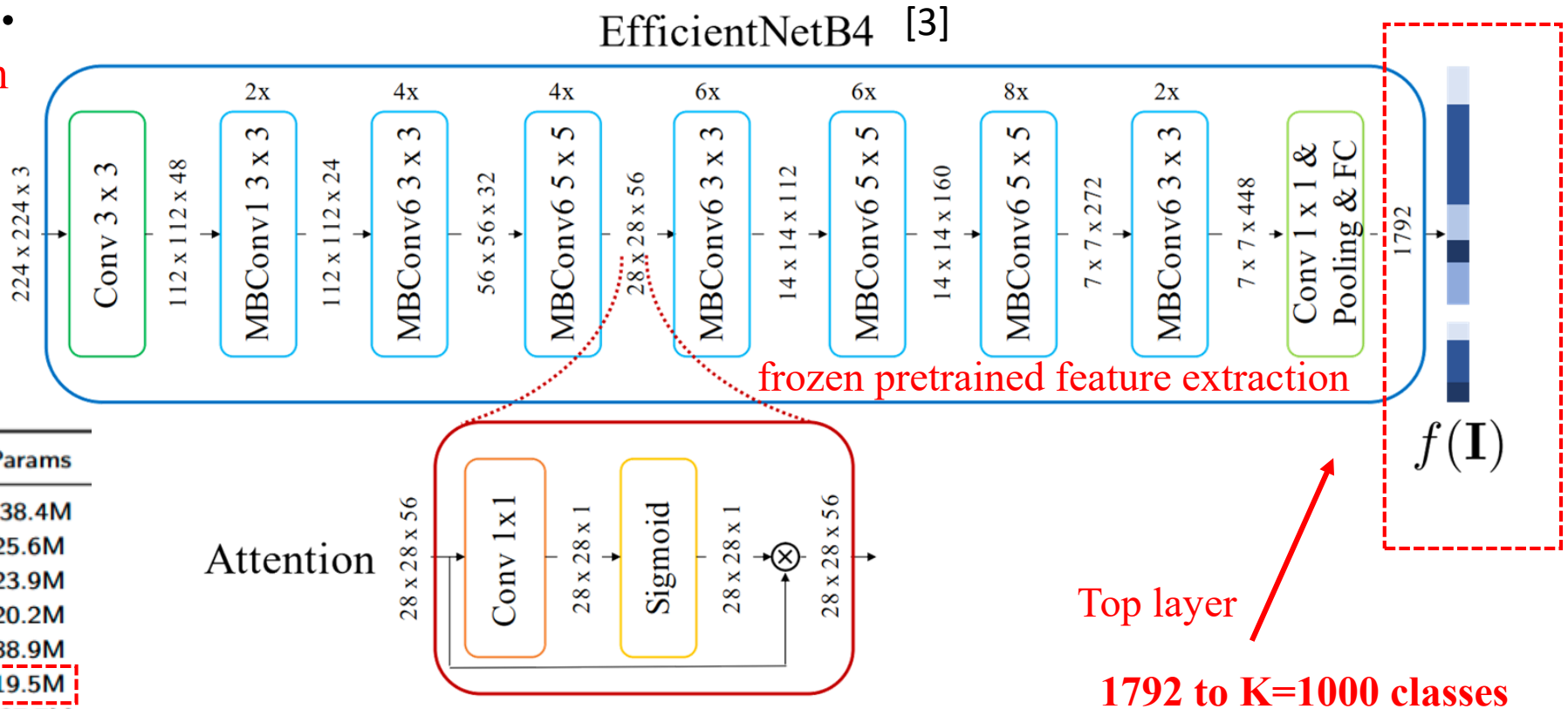
SINP MSU, Moscow, Russia
6 July, 2026

Goal of Research:

Improve of DLCP 2025 results[1]: from Natural Images ImageNet-100 with Arnoldi $O(m \times nnz(G)+m^2N)$ Nishimori temp. est. to ImageNet-1K and beyond with large test batch[2], less complexity, classification accuracy better than VT.

Object of research:

top layer (MLP, last layer) which replaced by Energy model.
Result: +5.9% Top-1 classif. accuracy with cost FFT+Relay projection $O(Nlog(N)+k_{mode}^2 N)$



CNN baselines (Keras ImageNet-1K)

[4]

Model	Size(MB)	Top-1(%)	Params
VGG16	528	71.3	138.4M
ResNet50V2	98	76.0	25.6M
InceptionV3	92	77.9	23.9M
DenseNet201	80	77.3	20.2M
NASNetLarge	343	82.5	88.9M
EfficientNet-B4	75	82.9 88.83%	19.5M
ConvNeXt-Large	755	86.3	197.7M

1.Usatyuk V.S., Sapozhnikov D.A., Egorov S.I. Natural Image Classification via Quasi-Cyclic Graph Ensembles and Random-Bond Ising Models with Enhanced Nishimori Temperature. Moscow Univ. Phys. 80 (Suppl. 3), S1039–S1053 (2025).
2. Reach high speed of classification hardware (FPGA/TPU/NPU/GPU/CPU)
3. Tan, M., & Le, Q. V. (2019). EfficientNet: Rethinking Model Scaling for Convolutional Neural Networks. In *International Conference on Machine Learning* (pp. 6105–6114). PMLR.
4. https://keras.io/api/applications/

Energy Based Model (EBM) proposition:

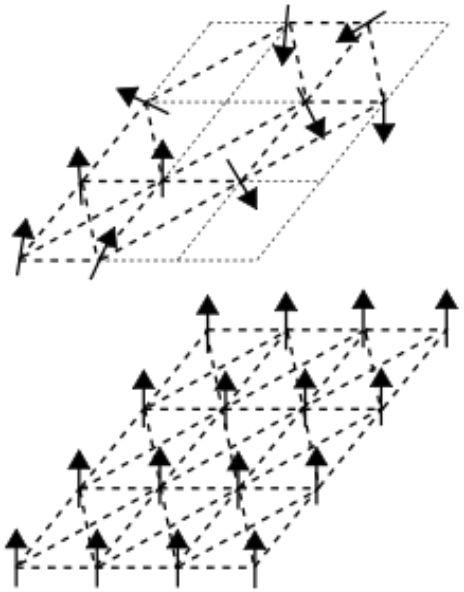
1. Construct a physics-inspired EBM of artificial massive quasi-crystalline lattices(non-regular clusters) with 1,000 (classes) distinct atomic nuclei centers (centroids), encoded using ImageNet-1K dataset features from EfficientNet-B4 and mapped to Energy Landscape via Trapping Set Ontology Fractal–optimized sparse graph spectral embedding's.

2. Demonstrate that under (heavy atomic nuclei) adiabatic evolution driven by Bethe energy and Nishimori temperature, the system converges to a self-consistent (Kohn-Sham*) equilibrium, which is equivalent to parallel (test batch), low complexity FFT-based (with Relay perturbation projection) Sparse Graph Spectral Embedding classification on ImageNet-1K.

*Kohn, Walter; Sham, Lu Jeu (1965). "Self-Consistent Equations Including Exchange and Correlation Effects". *Physical Review*. **140** (4A): A1133–A1138. Bibcode:1965PhRv..140.1133K. doi:10.1103/PhysRev.140.A1133.

Random Bond Ising Model (RBIM) - 2D Ising model on Graph $\mathcal{G} = (\mathcal{V}, \mathcal{E})$

$$\mathcal{H}(\sigma) = -\sum_{(i,j)\in\mathcal{E}} J_{ij}\sigma_i\sigma_j,$$



Each node $i \in \mathcal{V}$ is associated with a spin variable, $\sigma_i \in \{-1, +1\}$, $\mathcal{E}$ - is the set of edges (bonds)
J_{ij}- bond strengths, which are random assigned according to a specified probability distribution.
In the random bond Ising model, J_{ij} can take different values, typically following a distribution such as

$$P(J_{ij}) = p\delta(J_{ij} - J) + (1-p)\delta(J_{ij} + J),$$

where δ is the Dirac delta function, J is a positive constant representing the bond strength, and p is the probability of a bond being ferromagnetic ($J_{ij} = J$) versus antiferromagnetic ($J_{ij} = -J$).

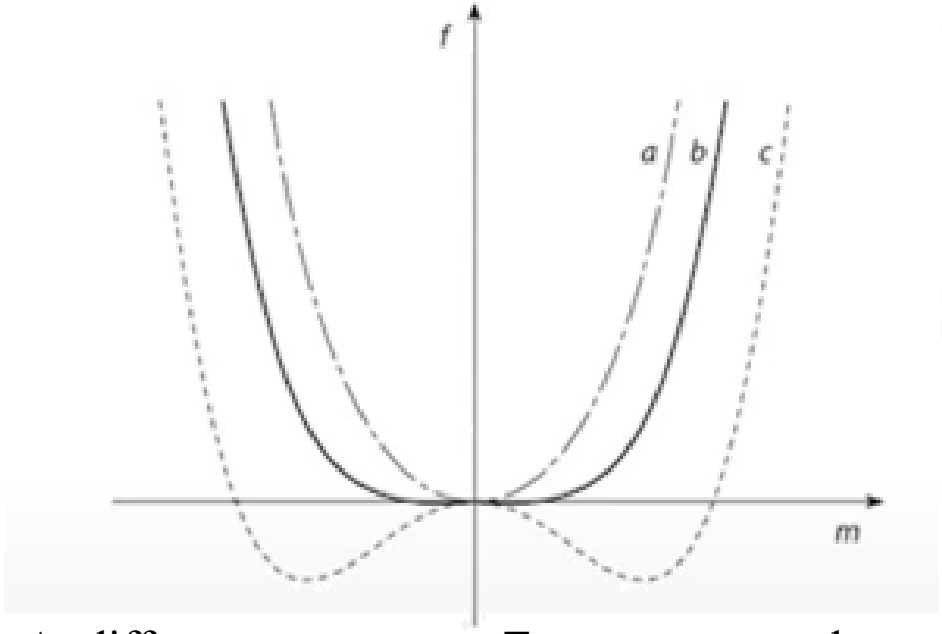
The partition function Z of the system, which is used to study the statistical properties, is given by

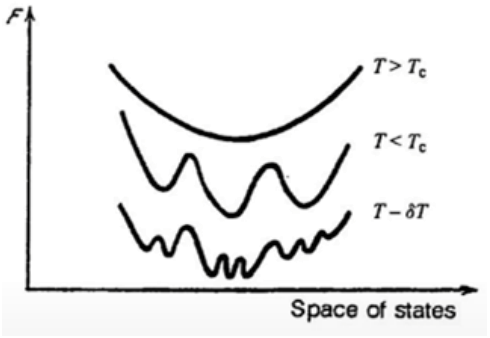
$$Z = \sum_{\sigma} e^{-\beta\mathcal{H}(\sigma)},$$

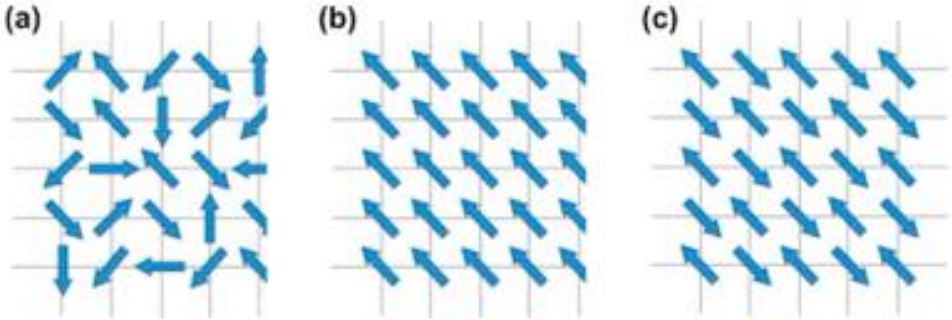
where $\beta = \frac{1}{k_B T}$ is the inverse temperature, with k_B being the Boltzmann constant and T the temperature.

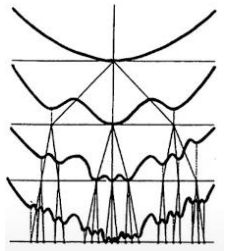
Rapaport D. C.. , *J. of Phys.* , 1972 , vol. C5 1830[IoP STACKS]; D. C. Rapaport, J. of Phys. **C5** (1972),2813[IoP STACKS], K. I. Grozdev and A. M. Kosevich, Phys. Stat. Sol. **b66** (1974), 77; K. I. Grozdev and A. M. Kosevich, Soviet Phys.-JETP **66** (1974), 2269

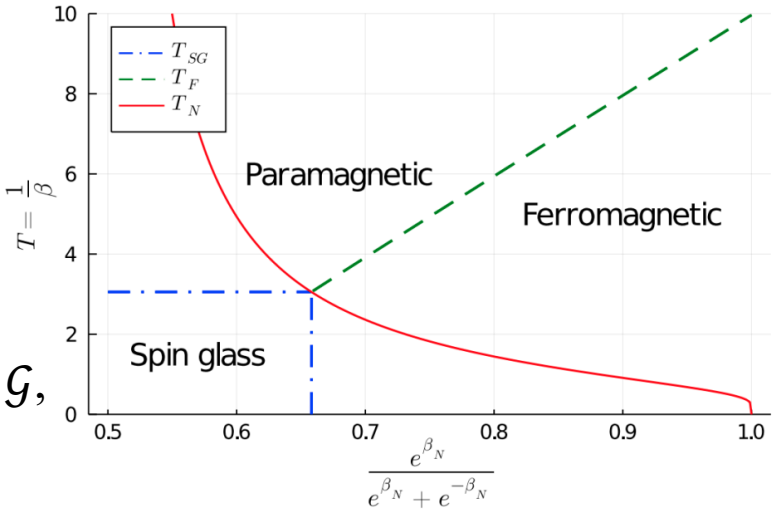
RBIM phase transaction under different temperature











At different temperature T we can see a phase transition

Under Nishimori Temp. Let $A \in \{0,1\}^{n\times n}$ be the symmetric adjacency matrix of $\mathcal{G}$, defined by $A_{ij} = 1$ if $(ij) \in \mathcal{E}$, and $A_{ij} = 0$ otherwise. Let $D \in \mathbb{N}^{n\times n}$ be the diagonal degree matrix $D = diag(A\mathbf{1}_n)$.

$$\mathbb{E}\left[H_{\beta_N,J}\right] = I_n + \mathbb{E}\left[\frac{th(\beta J_{ij})}{1-th^2(\beta J_{ij})}\right](D - A).$$

A $-$ *non diagonal element completly defined* by graph $\mathcal{G} = (\mathcal{V}, \mathcal{E})$

Viktor Dotsenko, Introduction to the Replica Theory of Disordered Statistical Systems *Landau Institute for Theoretical Physics, Moscow, 2001, 219 p*

Mezard M . Parisi G., Virasoro M. Spin Glass Theory and Beyond An Introduction to the Replica Method and Its Applications.1986, 476 p.

H. Nishimori, J. Phys. C: Solid State Phys. 13(21), 4071 (1980) , H. Nishimori, Prog. Theor. Phys. 66(4), 1169 (1981), H. Nishimori, Prog. Theor. Phys. 76, 305 1986

Lorenzo Dall'Amico et al Nishimori meets Bethe: a spectral method for node classification in sparse weighted graphs J. Stat. Mech. (2021) 093405

$$\mathbb{E}[H_{\beta_N,J}] = I_n + \mathbb{E}\left[\frac{th(\beta J_{ij})}{1-th^2(\beta J_{ij})}\right](D-A).$$

Quasi-Cyclic Sparse Graphs of spin interaction in RBIM

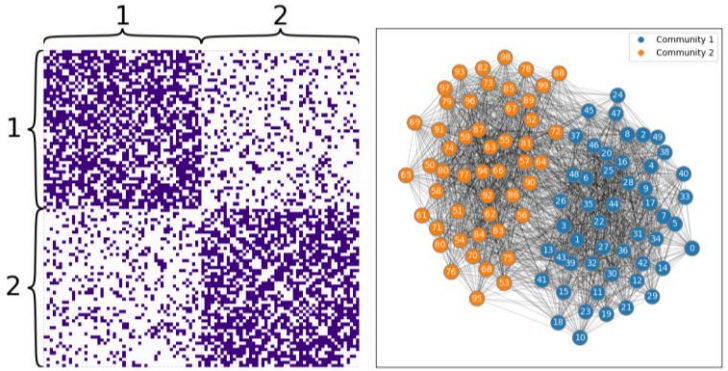
$A\ in\ Bethe-Hessian\ \ defined$ by graph $\mathcal{G} = (\mathcal{V}, \mathcal{E})$

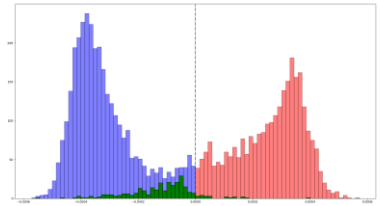
A **quasi-cyclic Low Density parity-check (QC-LDPC) code** has a sparse parity-check matrix H composed of m×n blocks, each of size $p \times p$, *[1,2]*. Each block is either:
A **zero matrix**, or a **cyclic permutation matrix (CPM)** P^a ($a \in \mathbb{Z}$), defined by cyclically shifting the rows of the identity matrix I_p by a positions.
Exponents a of CPM are element of the cyclic group $\mathbb{Z}/p\mathbb{Z}$ (Toric Structure).
Such matrix represented by bipartite graph (Tanner-graph).

$p = 2$

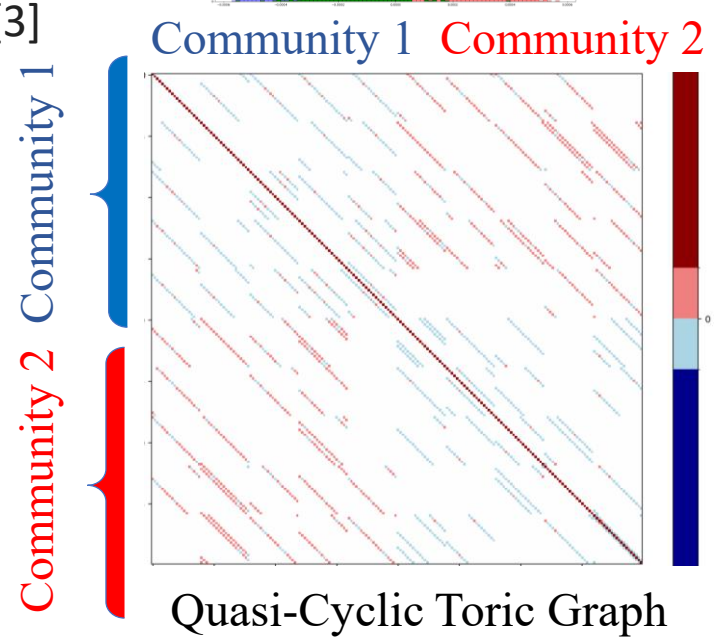
$$H = \begin{array}{c|cccccc} & v_1 & v_2 & v_3 & v_4 & v_5 & v_6 \\ \hline c_1 & 1 & 0 & 0 & 1 & 0 & 1 \\ c_2 & 0 & 1 & 1 & 0 & 1 & 0 \\ \hline c_3 & 1 & 0 & 0 & 0 & 1 & 0 \\ c_4 & 0 & 1 & 0 & 0 & 0 & 1 \end{array}$$

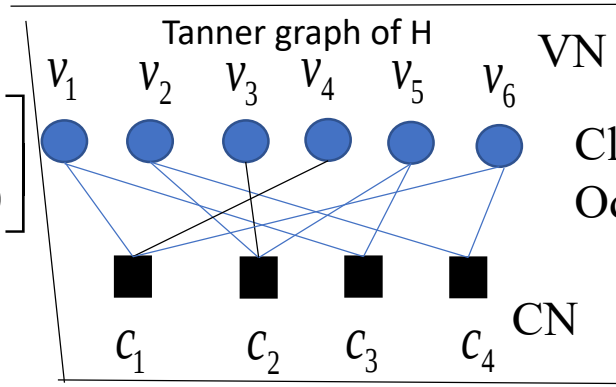
$$H_{QC} = \begin{bmatrix} I^0 & I^1 & I^1 \\ I^0 & I^{-1} & I^0 \end{bmatrix} = \begin{bmatrix} 0 & 1 & 1 \\ 0 & -1 & 0 \end{bmatrix}$$





[3]





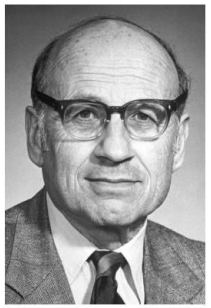
Closed path from variable node form cycle.
Odd degree cycle (imbalance) form basin in energy lands.

Cycle 8:

$$v_1 \to c_3 \to v_5 \to c_2 \to v_2 \to c_4 \to v_6 \to c_1 \to v_1$$

[1] R. M. Tanner, D Sridhara, and T. Fuja, "A class of group-structured LDPC codes," Proc. ISCTA 2001, 2001. [2] Fossorier M.P.C., "Quasi-cyclic low-density parity-check codes from circulant permutation matrices", *IEEE Trans. Inf. Theory*, vol. 50, no. 8, pp. 1788–1793, 2004.
[3] 1.Usatyuk V.S., Sapozhnikov D.A., Egorov S.I. Natural Image Classification via Quasi-Cyclic Graph Ensembles and Random-Bond Ising Models with Enhanced Nishimori Temperature. Moscow Univ. Phys. 80 (Suppl. 3), S1039–S1053 (2025).

Why Kohn–Sham? The Many-Body Problem



Kohn-Sham([1] 1965, [2] Nobel Prize in Chemistry 1998, Foundation of DFT): An N-electron system has a many-body wavefunction $\Psi(\mathrm{r}_1, \ldots \mathrm{r}_N)$ governed by the full Schrodinger equation — intractable for large N.

Kohn & Sham replaced it with N non-interacting single-particle equations that reproduce the same ground-state density $\rho(r)$:

Single-particle Hamiltonian

$$\left(-\frac{\hbar^2}{2m}\nabla^2 + v_{eff}[\rho](r)\right)\varphi_i(r) = \varepsilon_i\varphi_i(r),$$

***r* 3*N* coordinate became 3**

$\vec{r} = (x, y, z)$

where $-\frac{\hbar^2}{2m}\nabla^2$-kinetic energy operator, $\varphi_i(r)$ - Kohn-Sham orbit, ε_i - orbit energy, v_{eff} depends on ρ, $\rho(r) = \sum_{i=1}^{N}|\varphi_i(r)|^2$.

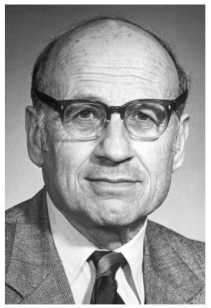
In Kohn–Sham DFT the effective potential decomposes as:

$$v_{eff}(r) = v_{ext}(r) + e^2\int\frac{\rho(r')}{|r-r'|}dr' + \frac{\delta E_{xc}[p]}{\delta p(r)},$$

where $v_{ext}(r)$- external field created by the nuclei of atoms, $e^2\int\frac{\rho(r')}{|r-r'|}dr'$-Hartree (classical Coulomb repulsion of electrons), $\frac{\delta E_{xc}[p]}{\delta p(r)}$-exchange-correlation potential (contains all quantum effects).

[1] *Kohn, Walter; Sham, Lu Jeu (1965). "Self-Consistent Equations Including Exchange and Correlation Effects". *Physical Review*. **140** (4A): A1133–A1138. Bibcode:1965PhRv..140.1133K. doi:10.1103/PhysRev.140.A1133.
[2] https://www.nobelprize.org/prizes/chemistry/1998/summary/

Why Kohn–Sham? The Many-Body Problem in Spectral Embedding



Kohn-Sham([1] 1965, [2] Nobel Prize in Chemistry 1998, Foundation of DFT): An N-electron system has a many-body wavefunction $\Psi(\mathrm{r}_1, \ldots \mathrm{r}_N)$ governed by the full Schrodinger equation — intractable for large N.

Kohn & Sham replaced it with N non-interacting single-particle equations that reproduce the same ground-state density $\rho(r)$:

Single-particle Hamiltonian

$$\left(-\frac{\hbar^2}{2m}\nabla^2 + v_{eff}[\rho](r)\right)\varphi_i(r) = \varepsilon_i\varphi_i(r),$$

***r* 3*N coordinate* became 3**

$$\vec{r} = (x, y, z)$$

where $-\frac{\hbar^2}{2m}\nabla^2$-kinetic energy operator, $\varphi_i(r)$ - Kohn-Sham orbit, ε_i - orbit energy, v_{eff} depends on ρ, $\rho(r) = \sum_{i=1}^{N}|\varphi_i(r)|^2$.

Due to Pontyagin Duality [3] continuous single-particle Hamiltonian [4] maps onto a Bethe-Hessian (Regularized (deformed), [5]) Laplacian under QC Torical Graph.

A sparse QC graph with D = 1792 feature channels has a coupled many-feature affinity structure. Instead of solving one giant coupled problem, Kohn–Sham Spectral Embedding solves D independent single-feature eigenproblems — each with its own effective Hamiltonian $L(k)(\beta(k)N)$ — and concatenates the results, exactly as Kohn–Sham concatenates orbitals.

[1] *Kohn, Walter; Sham, Lu Jeu (1965). "Self-Consistent Equations Including Exchange and Correlation Effects". *Physical Review*. **140** (4A): A1133–A1138. Bibcode:1965PhRv..140.1133K. doi:10.1103/PhysRev.140.A1133.
[2] https://www.nobelprize.org/prizes/chemistry/1998/summary/
[3] L. S. Pontrjagin. 1934. The theory of topological commutative groups. Annals of Mathematics 35, 2 (1934), 361–388.
[4] Penz, M., & van Leeuwen, R. 2021. Density-functional theory on graphs. The Journal of chemical physics, 155(24), 244111
[5] Alireza Saade, Florent Krzakala, and Lenka Zdeborová. 2014. Spectral Clustering of Graphs with the Bethe Hessian. In Proceedings of the 27th International Conference on Neural Information Processing Systems (NIPS'14). MIT Press, Cambridge, MA, USA, 406–414.

Bethe-Hessian (Regularized Graph Laplacian) — Kohn-Sham Analogy

For fixed β and feature k with affinity weights J_{ij} effective interaction and variance:

$$W_{ij} = \frac{th(\beta J_{ij})}{1 - th^2(\beta J_{ij})} = \frac{1}{2}\sinh(2\beta J_{ij}), \qquad \Lambda_{ij} = \sinh^2(\beta J_{ij}),.$$

Diagonal regularized and scaling:

$$D_{ii} = 1 + \sum_{j \in \partial i} \Lambda_{ij}, \qquad s_i = \frac{1}{\sqrt{D_{ii}}}, \qquad S = diag(s_1, \dots, s_N).$$

Regularized Laplacian to detect community boundaries in complex networks under noisy conditions:

$$L(\beta) = I - SWS,$$

where W- scales down strong physical bonds that act as structural outliers, S - down-weights highly volatile nodes by tracking their localized spin variances

[1] *Kohn, Walter; Sham, Lu Jeu (1965). "Self-Consistent Equations Including Exchange and Correlation Effects". *Physical Review*. **140** (4A): A1133–A1138. Bibcode:1965PhRv..140.1133K. doi:10.1103/PhysRev.140.A1133.
[2] https://www.nobelprize.org/prizes/chemistry/1998/summary/

Kohn-Sham DFT and Spectral Embedding Self-Consistency cycle

Kohn–Sham DFT cycle

1. Guess initial density $\rho_0(\mathbf{r})$.
2. Build $v_{\text{eff}}[\rho_0]$.
3. Solve single-particle Schrödinger eq.:

$$\left(-\frac{\hbar^2}{2m}\nabla^2 + v_{\text{eff}}\right)\varphi_i = \varepsilon_i\,\varphi_i$$

4. Compute new density $\rho_1 = \sum |\varphi_i|^2$.
5. If $|\rho_1 - \rho_0| > \epsilon$, set $\rho_0 \leftarrow \rho_1$, go to 2.
6. At convergence: ground-state energy $E = \sum \varepsilon_i - E_H + E_{xc} - \int v_{xc}\,\rho$.

KSSE cycle (per feature k)

1. Build affinity $\mathrm{J}^{(k)}$ from feature channel k.
2. Compute spin-glass bound $\beta_{sg}^{(k)}$ (graph statistics: degree, ϕ).
3. For trial $\beta = i \cdot \beta_{sg}$:

$$\mathrm{L}^{(k)}(\beta) = \mathrm{I} - \mathrm{SW}^{(k)}\mathrm{S}$$

4. Compute $\lambda_{\min}(\mathrm{L}^{(k)}(\beta))$ via FFT.
5. If $\lambda_{\min} \neq 0$, update β (quadratic interpolation / bisection).
6. At convergence: $\beta_N^{(k)}$ where $\lambda_{\min} = 0$; embedding $\mathrm{e}^{(k)} = \mathrm{S}\,v_{\min}$.

Final embedding (analogue of the Slater determinant):

$$\mathbf{E} = [\,\mathbf{e}^{(1)} \mid \mathbf{e}^{(2)} \mid \cdots \mid \mathbf{e}^{(D)}\,] \in \mathbb{R}^{N\times D}, e_i^{(k)} = s_i^{(k)}\, v_{\min}^{(k)}[i].$$

Classification: logistic regression trained on frozen training portion of E.

Both cycles seek a self-consistent fixed point:

ensity ρ such that $\mathrm{v}_{eff}\,[\rho]$ reproduces ρ.

temperature β_N such that the regularised Laplacian is at criticality ($\lambda_{\min} = 0$).

[1] *Kohn, Walter; Sham, Lu Jeu (1965). "Self-Consistent Equations Including Exchange and Correlation Effects". *Physical Review*. **140** (4A): A1133–A1138. Bibcode:1965PhRv..140.1133K. doi:10.1103/PhysRev.140.A1133.
[2] https://www.nobelprize.org/prizes/chemistry/1998/summary/

Spectral Embedding Nuclei (Frozen) and Valence Electron (Moving)

Graph size is fixed N_{graph}:

$$N_{\text{graph}} = N_{\text{frozen}} + N_{thawed}$$

Frozen representation (training)

- One representative subset of training examples.
- Selected by proximity to class centroids in spectral space.
- Remains constant across all test batches.
- E.g. 40,000 nodes (∼40/class for 1K classes).

Moving representations (testing)

- Test set split into $M = N_{\text{test}}/N_{\text{thawed}}$ uniform random slices.
- Each slice: $\sim N_{\text{thawed}}/K$ examples per class.
- E.g. 5,000 nodes $\times 10$ batches for ImageNet-1K.

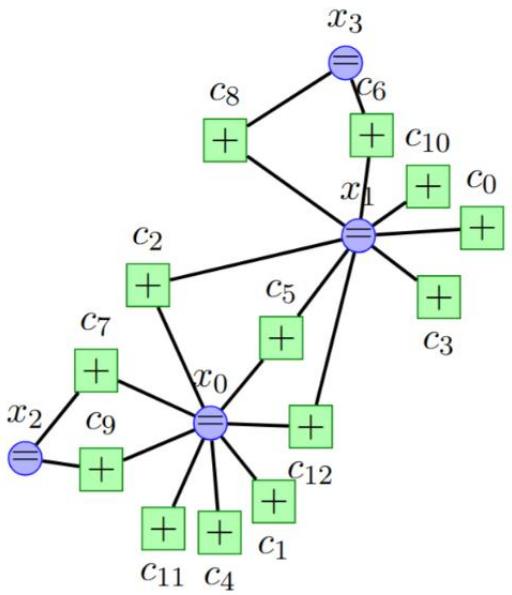
Atomic Nuclei quasi-lattice (40 per class)
High-mass, stationary centers that establish the fundamental geometric scaffolding and the global **Energy Landscape**

Valence Electron Cloud (thawed)
Low-mass, highly dynamic particles injected into the lattice to probe, settle, and find equilibrium within the field.

Trapping Set in QC Sparse Graph

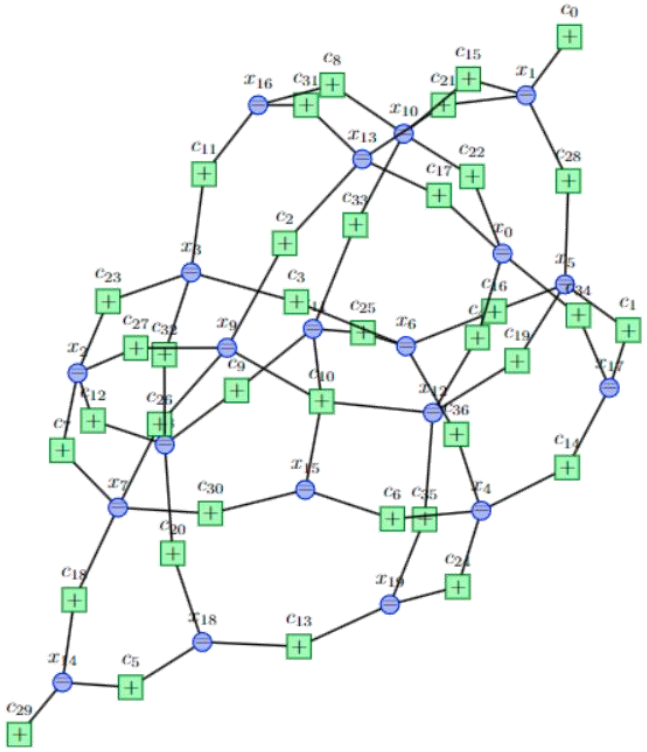
Trapping sets is a subgraph formed by cycles or it overlap with a variable nodes, b odd degree check.

TS(4,6)

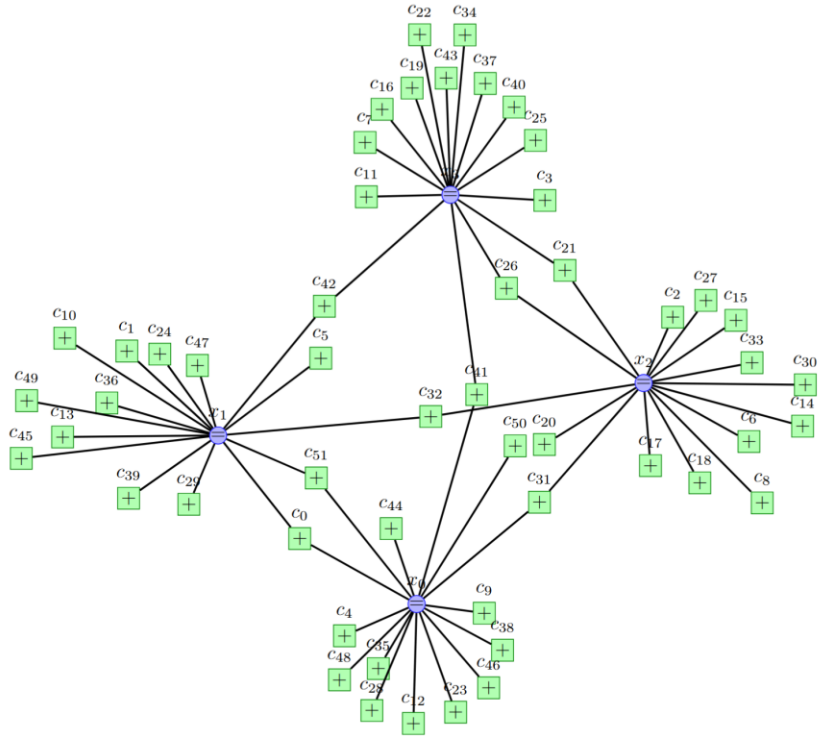


$$H_{TS(4,6)} = \begin{bmatrix} 0 & 1 & 0 & 0 \\ 1 & 0 & 0 & 0 \\ 1 & 1 & 0 & 0 \\ 0 & 1 & 0 & 0 \\ 1 & 0 & 0 & 0 \\ 1 & 1 & 0 & 0 \\ 0 & 1 & 0 & 1 \\ 1 & 0 & 1 & 0 \\ 0 & 1 & 0 & 1 \\ 1 & 0 & 1 & 0 \\ 0 & 1 & 0 & 0 \\ 1 & 0 & 0 & 0 \\ 1 & 1 & 0 & 0 \end{bmatrix}$$

TS(20,2)



TS(4,44)

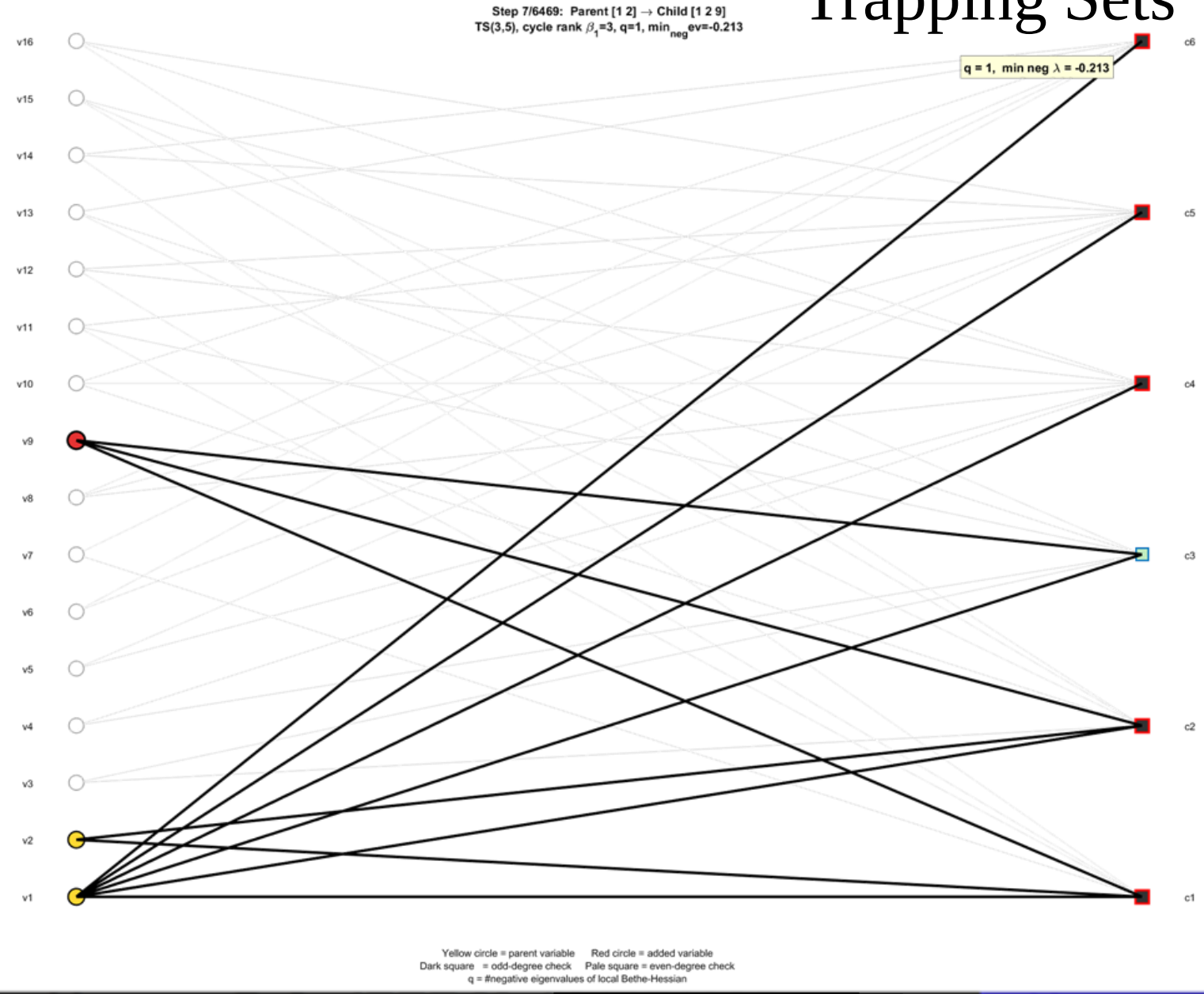


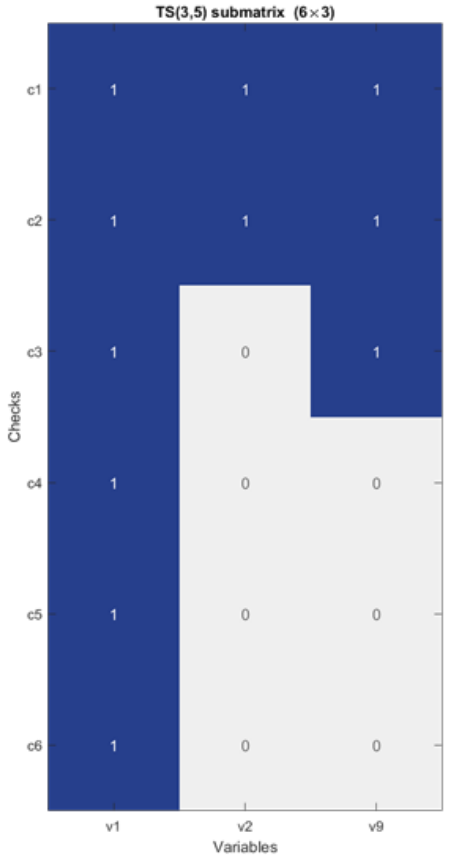
TS(a,0) form (low energy) solutions, large a can increase gap between such solution. HOMO-LUMO gap, (Highest Occupied Molecular Orbital (HOMO) and LUMO (Lowest Unoccupied Molecular Orbital) - A large gap means the molecule is highly stable and unreactive.

TS(a,b≠ 0) submatrix and related graph introduce negative eigen values.

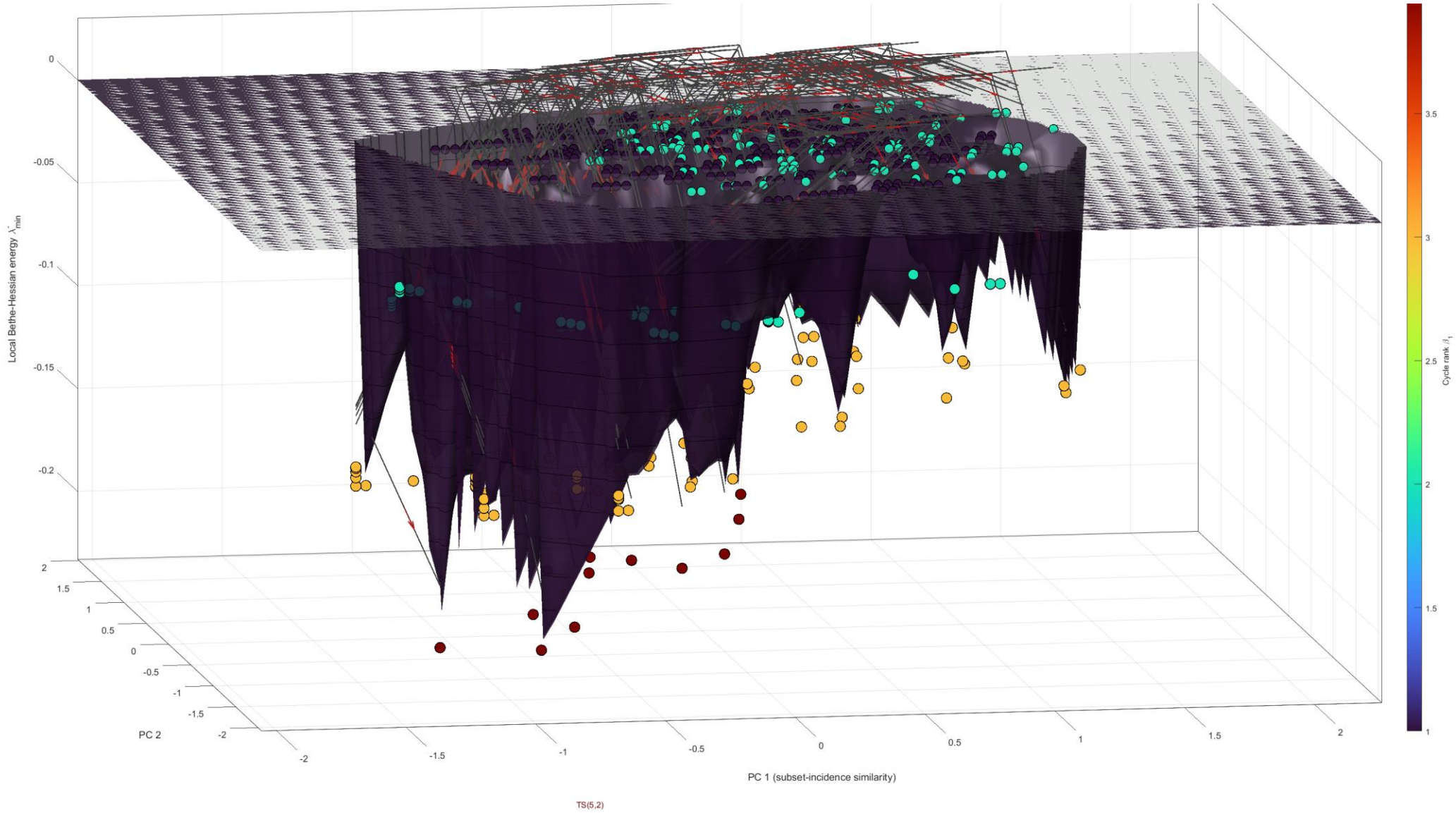
Graph embedding must be optimized by TS sugraph.

Trapping Sets

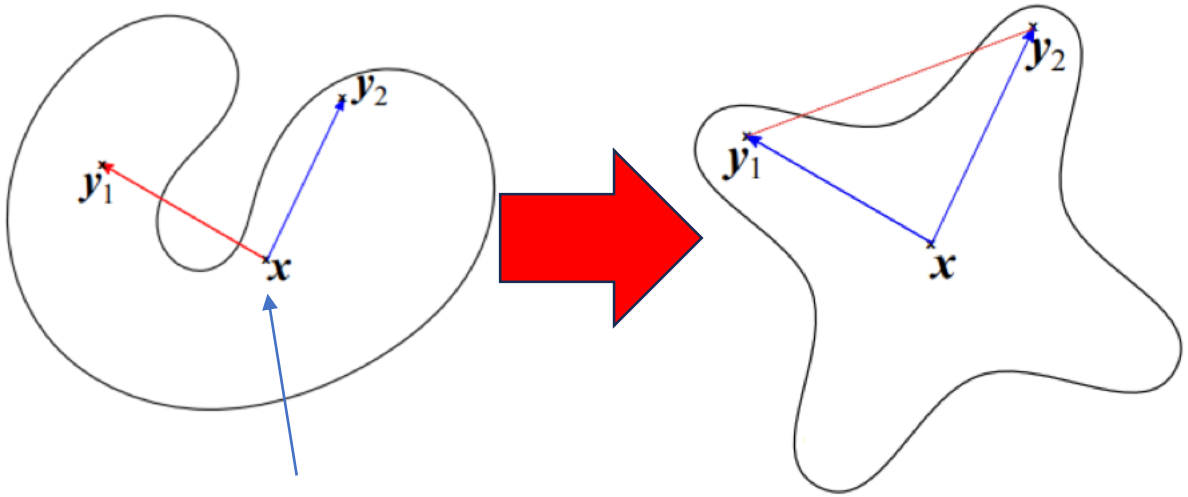




Trapping Set Landscape



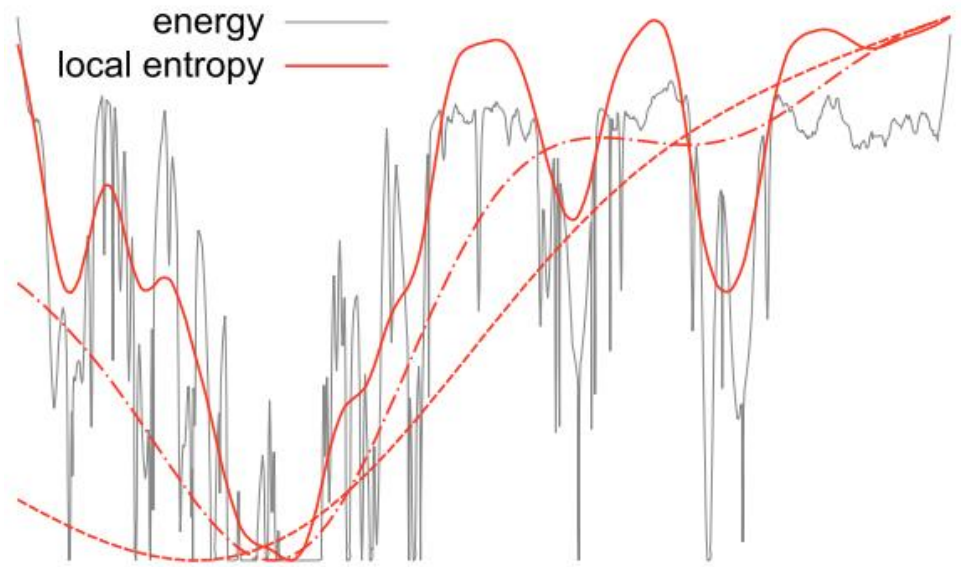
Trapping Set Landscape optimization goal



x is TS(a,0) – energy minima

y_i TS(a,0) destroy star domain ‘local’ convexity

Star Domain
‘local’ convexity



Wide Basin

Minimizing this free energy involves adjusting the neural network parameters θ to ensure that $Q_\theta(\sigma)$ closely approximates the true distribution $P(\sigma)$. This process is analogous to learning the weights in the neural network, where the energy function $\mathcal{H}(\sigma)$ can be interpreted as a parameterized function (such as a neural network) that captures the dependencies between the variables. Local entropy[1] $\Phi(\sigma, \beta, \theta(\gamma)) = log \sum_{\{\sigma'\}} e^{-\beta E(\sigma') - \gamma d(\sigma,\sigma'),}$, where $d(\cdot,\cdot)$ – distance between configuration, γ –weight, σ – reference spin configuration.

Global extremum represented by symmetrical cycle in graph – TS(a,0). Local extremum around represented by odd degree connected cycle in graph, TS(a,b).

1. Carlo Baldassia, Christian Borgsc , et a/ Unreasonable effectiveness of learning neural networks: From accessible states and robust ensembles to basic algorithmic schemes *Proceedings of the National Academy of Sciences* (PNAS), 113 (48), **November 15, 2016,** E7655-E7662 https://www.pnas.org/doi/full/10.1073/pnas.1608103113

Trapping Set Landscape Fractal Analysis

The correlation dimension D_2 describes how the number of nearby point-pairs scales as you expand your search radius r, [1] :

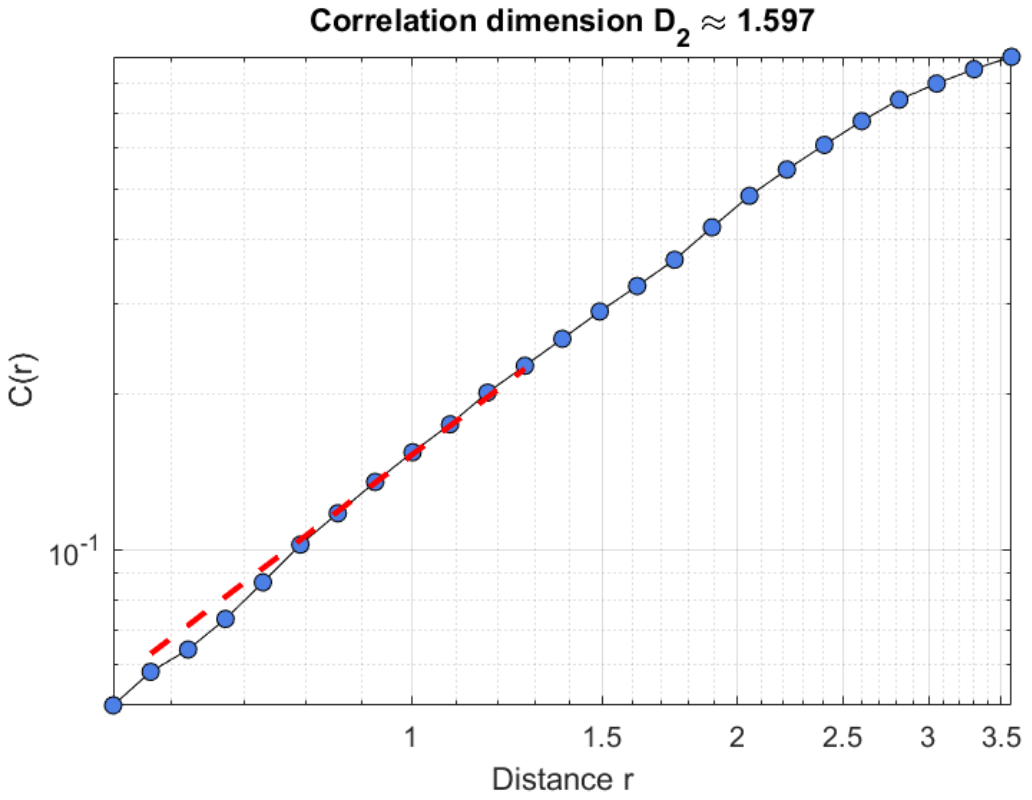
$$C(r) = \frac{2}{N(N-1)} \sum_{i=1}^{N} \sum_{j=i+1}^{N} \Theta\left(r - \| X_i - X_j \|\right),$$

where N −total number of points, Θ- Heaviside step function (equals 1 if the distance between points i and j is less than r, and 0 otherwise), $\| X_i - X_j \|$ is the Euclidean distance on your PCA plane.

Correlation dimension describe measure stability of Energy landscape

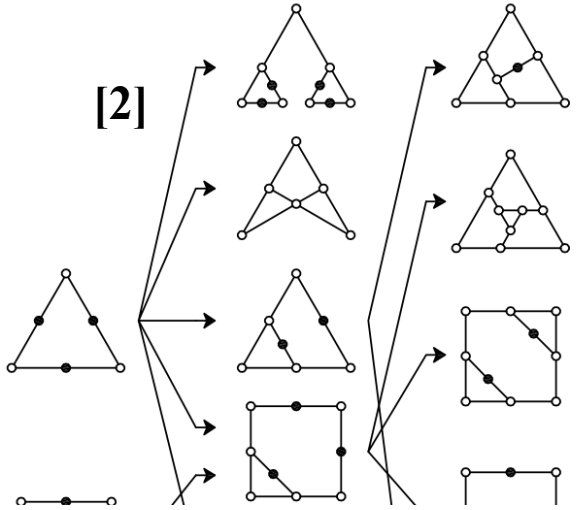
Smaller is fewer active degree of freedom.

Highly Fractional - More Chaotic.



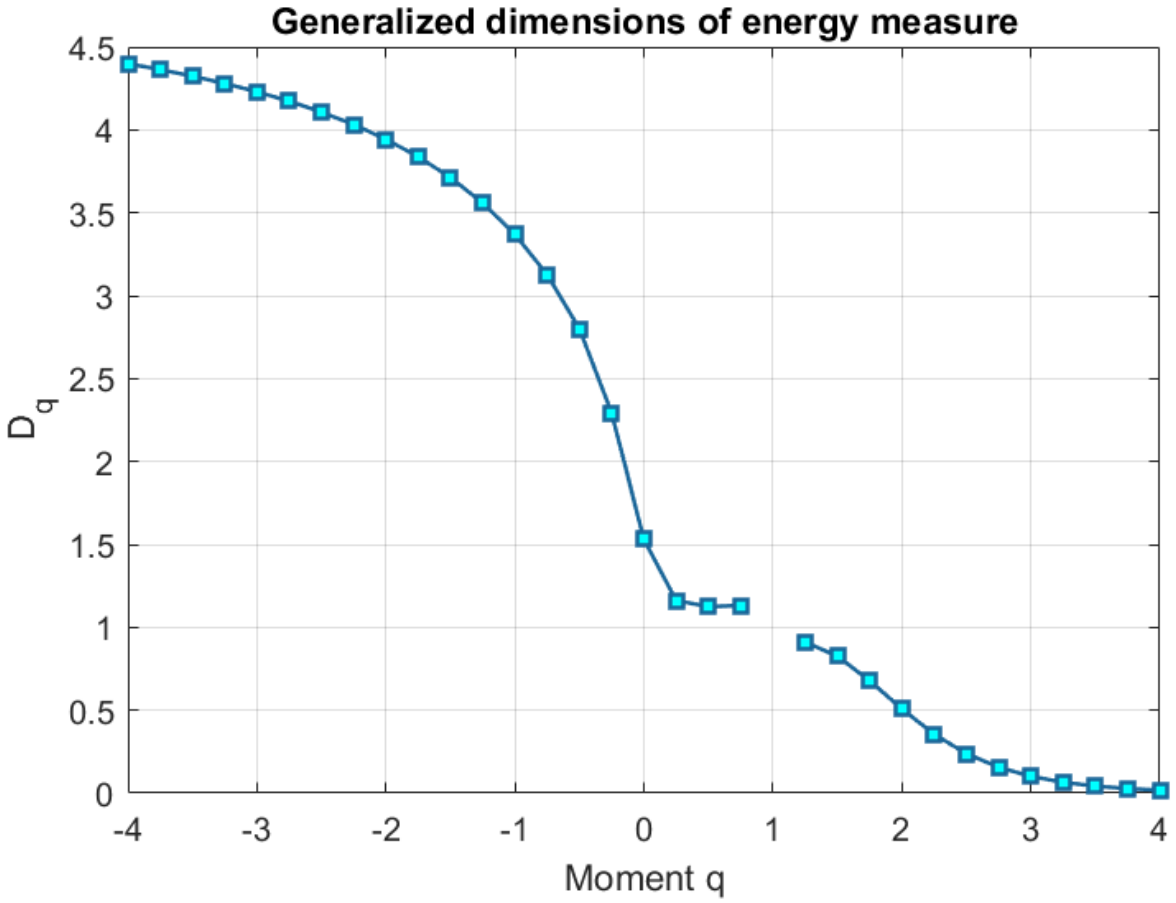
Trapping Set Ontology
Nested families

[2]



1. Peter Grassberger and Itamar Procaccia (1983). "Measuring the Strangeness of Strange Attractors". Physica D: Nonlinear Phenomena. **9** (1–2): 189–208. Bibcode:1983PhyD....9..189G. doi:10.1016/0167-2789(83)90298-1.
2.B. Vasić, S. K. Chilappagari, D. V. Nguyen and S. K. Planjery, "Trapping set ontology," 2009 47th Annual Allerton Conference on Communication, Control, and Computing (Allerton), Monticello, IL, USA, 2009, pp. 1-7, doi: 10.1109/ALLERTON.2009.5394825

Trapping Set Landscape Fractal Analysis



The generalized dimension spectrum D_q as a function of the statistical moment order $q \in [-4,4]$ computed for the local Bethe-Hessian energy measure, [1]. The sigmoidal curve demonstrates pronounced multifractal scaling behavior, transitioning from high-dimensional sparse fluctuations ($D_q \approx 4.4$ at $q = -4$) to low-dimensional highly-concentrated energy peaks ($D_q \approx 0$ at $q = 4$). The central gap explicitly isolates the numerical indeterminacy at the information dimension boundary ($q = 1$).

Highly Complex Landscape: The clean sigmoidal shape confirms that the Bethe-Hessian energy distribution is strictly **multifractal**, rather than a simple uniform structure. **Structurally Unstable**: The transition from a high dimension D=4.4 to zero proves that the energy landscape is highly fragmented and **heterogeneous**, filled with rare, deep energy traps mixed with widely scattered, unstable transition states.

1. Grassberger, P. (1983). *Generalized dimensions of strange attractors*. Physics Letters A, 97(6), 227-230.

Kohn–Sham Spectral Embedding (KSSE) under optimized sparse graph

Frozen EfficientNet-B4 (1792-D features, no retraining). Logistic regression classifier on spectral embedding.

	Best config	Worst config
Top-1 accuracy (50K test)	88.83%	79.20%
Graph size (nodes)	45,000	25,000
Frozen (train nodes)	40,000 (∼40/class)	15,000 (∼15/class)
Moving (test nodes/batch)	5,000 ×10	10,000 ×5
Quasi-stationarity	Yes (< 1% variation)	Yes
EfficientNet-B4 baseline	82.9%	

Model	Pre-train	Parameters	Top-1 Accuracy
Swin-L [1]	ImageNet-21K	197M	86.4%
Swin-L [1]	ImageNet-21K	197M	87.3%
ViT-H/14 [2]	ImageNet-21K	632M	88.0%
Proposed EffB4 feature [3]+KSSE	ImageNet-1K	≈19.56	88.83%
EfficientNet-B7		66M	84.3

CNN baselines (Keras ImageNet-1K) [4]

Model	Size(MB)	Top-1(%)	Params
VGG16	528	71.3	138.4M
ResNet50V2	98	76.0	25.6M
InceptionV3	92	77.9	23.9M
DenseNet201	80	77.3	20.2M
NASNetLarge	343	82.5	88.9M
EfficientNet-B4	75	82.9	19.5M
ConvNeXt-Large	755	86.3	197.7M

KSSE achieves a Top-1 accuracy of 88.83% on ImageNet-1K using only ≈19.56M parameters, significantly outperforming much larger VT architectures such as Swin-L (197M params, up to 87.3%) and ViT-H/14 (632M params, 88.0%).

1. Ze Liu, Yutong Lin, et al. 2021. Swin Transformer: Hierarchical Vision Transformer using Shifted Windows. In Proceedings of the IEEE/CVF International Conference on Computer Vision (ICCV). IEEE, Montreal, QC, Canada, 10012–10022. https://doi.org/10.1109/ICCV48922.2021.00986
2. Alexey Dosovitskiy, Lucas Beyer, et al. 2021. An Image is Worth 16x16 Words: Transformers for Image Recognition at Scale. In Proceedings of the International Conference on Learning Representations (ICLR). OpenReview.net, Virtual Event.
3. Tan, M., & Le, Q. V. (2019). EfficientNet: Rethinking Model Scaling for Convolutional Neural Networks. In *International Conference on Machine Learning* (pp. 6105–6114). PMLR.
4. https://keras.io/api/applications/

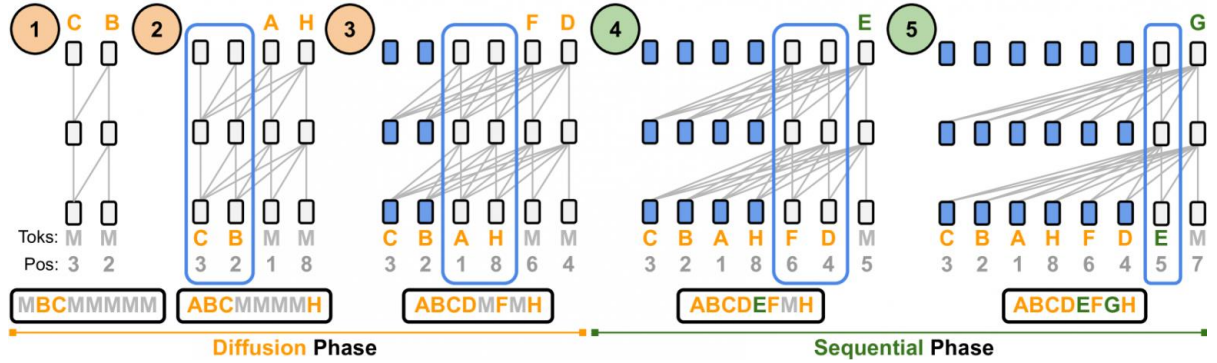
Quasi-Stationary - Nishimori temp similar per each feature due heavy nuclear

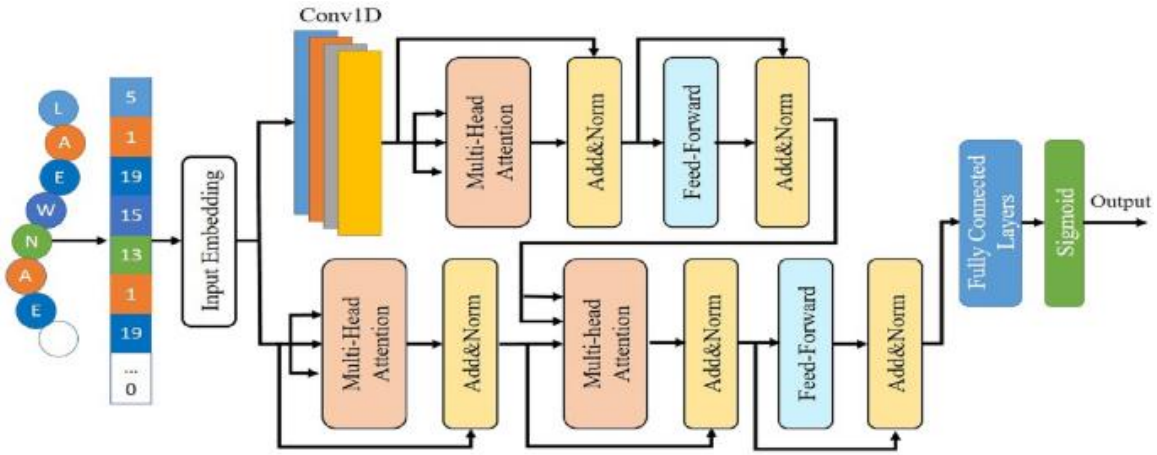
Top layer 1792*1000=1.792Mparam EffficentNetB4

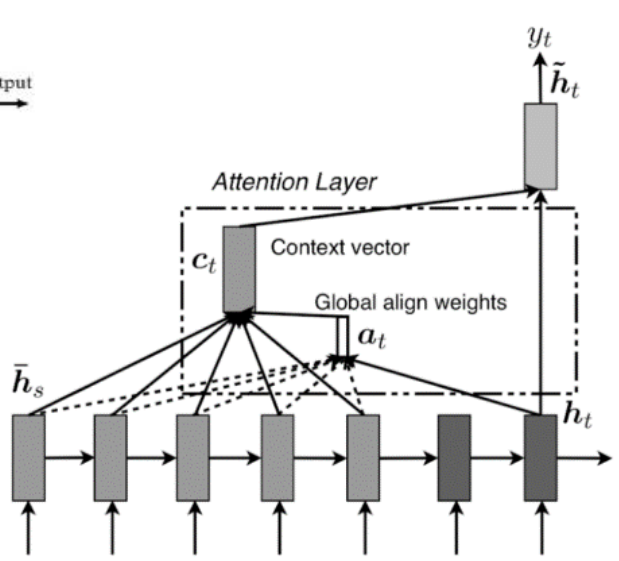
Summary

We proposed Kohn–Sham Spectral Embedding establishes a new Energy Model Based approach for parameter-efficient vision modeling by delivering state-of-the-art accuracy with drastically reduced computational overhead. As demonstrated in our experiments, KSSE achieves a Top-1 accuracy of 88.83% on ImageNet-1K using only ≈19.56M parameters, significantly outperforming much larger architectures such as Swin-L (197M params, up to 87.3%) and ViT-H/14 (632M params, 88.0%). This represents a reduction in model size of over 10× compared to Swin-L and more than 30× compared to ViT-H/14, while simultaneously delivering superior classification performance. The results validate KSSE's core design principles, highlighting its ability to extract highly discriminative features without relying on massive parameter counts or expensive pre-training regimes.

Thank You For Your Attention!







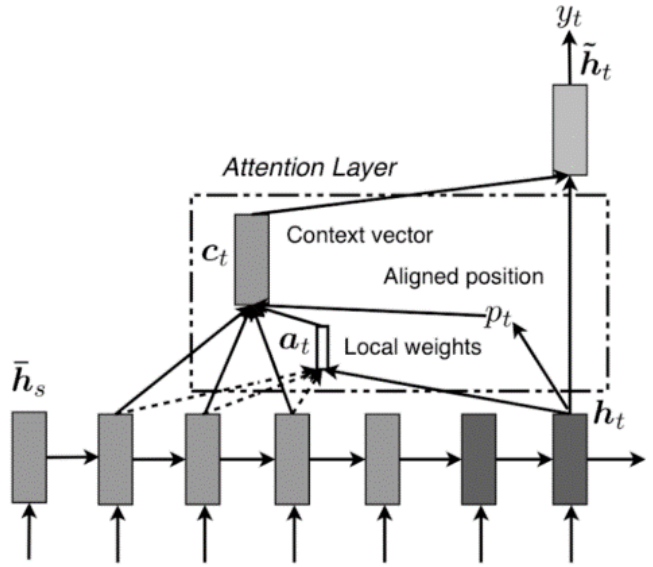


Таблица 1. 1000 классов. Точность в % в зависимости от веса, размера графа и числа образов на класс в представлении. У 15000 НУМЕРОЛОГИЯ ДРУГАЯ, там веса 26, 35, 38. ПОЭТОМУ СТОЛБЕЦ ПОМЕЧЕН КРАСНЫМ

РАЗМЕР ГРАФА	15000	20000	20000	25000	25000	30000	30000	30000	40000	40000	40000	45000	45000	45000
ОБРАЗЦОВ В ПОДВИЖНОМ ПРЕДСТАВЛЕНИИ	5	5	10	5	10	5	10	15	5	10	15	5	10	15
Вес 28	79.21	80.35	80.2	81.5	81.1	82.1	81.95	81.7	84.4	84.15	83.8	86.24	85.9	85.45
Вес 34	79.2	80.55	80.3	81.95	81.4	82.8	82.5	82.1	85.15	84.7	84.2	87.44	86.95	86.55
Вес 48	79.3	80.8	80.5	82.2	81.7	83.5	83	82.45	85.95	85.3	84.65	88.93	88.3	87.7

FFT Spectral Decomposition + Rayleigh Refinement

1. $\ell_0 = \mathsf{L}_0$ extract circulant shift
2. Approximate eigenvalues: $\tilde{\lambda}_m = \mathrm{Re}\big[\mathrm{FFT}(\ell_0)\big]_m$
3. Locate minimum: $\mathrm{idx}_{\min} = \arg\min(\tilde{\lambda})$
4. Build Fourier submatrix F with k_{mode} modes around $\mathrm{idx}_{\min}$:

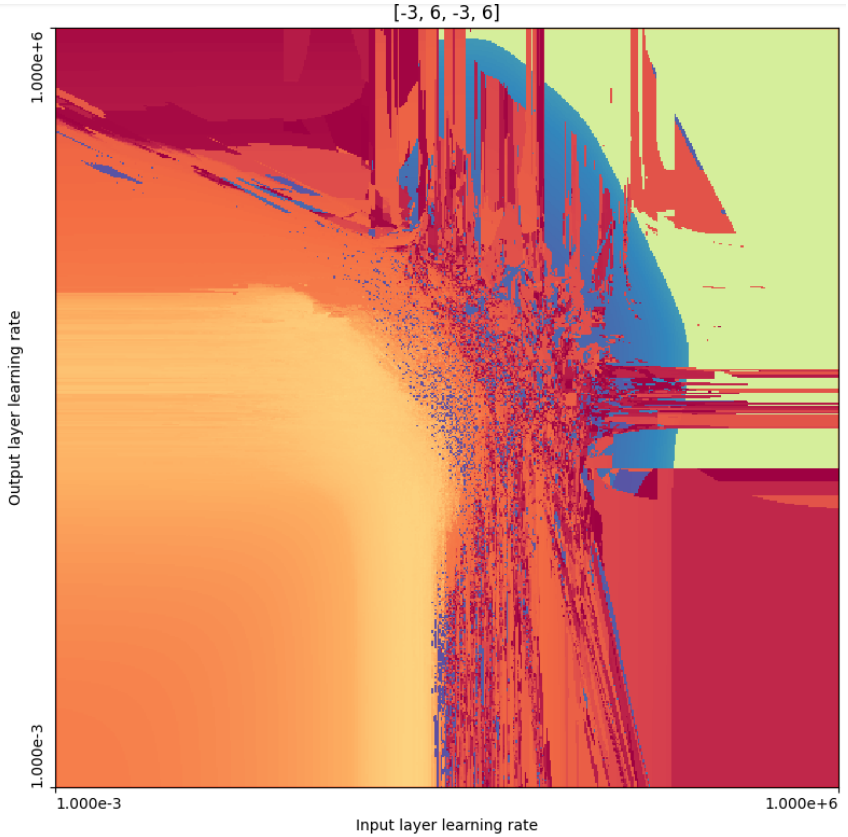
$$F_{n,m} = \frac{1}{\sqrt{N}} \exp\left(\frac{2\pi i\, n\, l_m}{N}\right)$$

5. Rayleigh refinement: $\rho_m = \mathsf{f}_m^{\dagger}\, \mathsf{L}\, \mathsf{f}_m$
6. Select $m^* = \arg\min_m \rho_m$; $\lambda_{\min} = \rho_{m^*}$, $v_{\min} = \mathrm{Re}(\mathsf{f}_{m^*})$

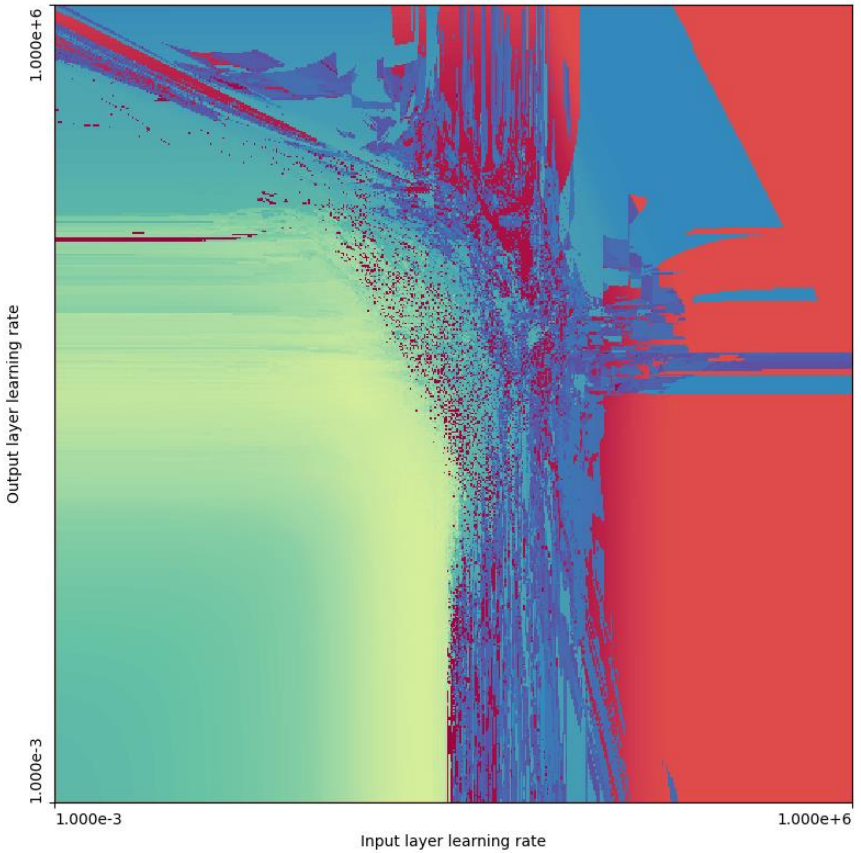
TS and learning rate fractal

$$f(x) = f(x_n + \Delta x) = f(x_n) + f'(x_n)\Delta x + f''(x_n)\Delta x^2,\ f(x) = 0, \Delta x = -\frac{f'(x_n)}{f''(x_n)},\ f''$$ – это Гессиан

before TS



After TS elimination



learning rate $\triangle x$

In an energy-based model, the probability of a configuration σ is defined in terms of an energy function, similar to the Hamiltonian of the Ising model. For the random bond Ising model, the probability $P(\sigma)$ of a spin configuration σ is given by the Boltzmann distribution:

$$P(\sigma) = \frac{e^{-\beta \mathcal{H}(\sigma)}}{Z},$$

where $\mathcal{H}(\sigma)$ is the Hamiltonian and Z is the partition function.

Similarly, deep neural networks can be used to model complex probability distributions. Consider a neural network with parameters θ that models a probability distribution $Q_\theta(\sigma)$. The goal is to approximate the true distribution $P(\sigma)$.

Variational inference involves approximating the true posterior distribution $P(\sigma)$ with a parameterized distribution $Q_\theta(\sigma)$ by minimizing the KL divergence between them. The KL divergence is defined as:

$$\mathrm{KL}(Q_\theta \parallel P) = \sum_\sigma Q_\theta(\sigma) \log \frac{Q_\theta(\sigma)}{P(\sigma)}.$$

Substituting $P(\sigma)$ with its Boltzmann form, we get:

$$\mathrm{KL}(Q_\theta \parallel P) = \sum_\sigma Q_\theta(\sigma)(\log Q_\theta(\sigma) + \beta \mathcal{H}(\sigma) + \log Z).$$

Since log Z is constant with respect to θ, it can be omitted from the optimization process. The optimization objective becomes:

$$\mathrm{KL}(Q_\theta \parallel P) \propto \sum_\sigma Q_\theta(\sigma)(\log Q_\theta(\sigma) + \beta \mathcal{H}(\sigma)).$$

Quasi-Instatones– local minimum TS(a,b) under local entropy Φ

solving an incorrectly posed operator problem*

$A \subset B, A - submatrices$ (local solution) contained in B connected through variable y_1, y_2, y_3.

A subgraph with cycle 6

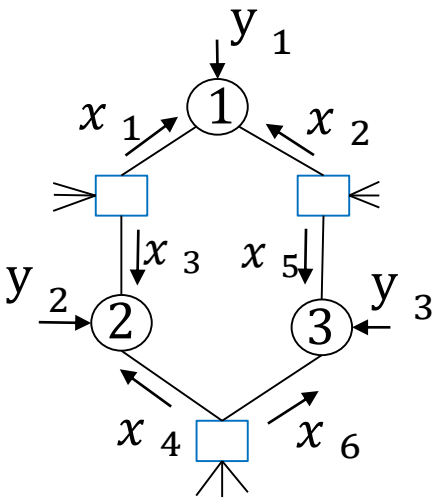
TS(a,b)

$x_1, x_2, x_3, x_4, x_5, x_6$

$$Ax = y,$$

$$\begin{cases} x_1 = (y_2 + x_4)/\alpha \\ x_2 = (y_3 + x_6)/\alpha \\ x_3 = (y_1 + x_2)/\alpha \\ x_4 = (y_3 + x_5)/\alpha \\ x_5 = (y_1 + x_1)/\alpha \\ x_6 = (y_2 + x_3)/\alpha \end{cases}$$

$$A = \begin{pmatrix} \alpha & 0 & 0 & -1 & 0 & 0 \\ 0 & \alpha & 0 & 0 & 0 & -1 \\ 0 & -1 & \alpha & 0 & 0 & 0 \\ 0 & 0 & 0 & \alpha & -1 & 0 \\ -1 & 0 & 0 & 0 & \alpha & 0 \\ 0 & 0 & -1 & 0 & 0 & \alpha \end{pmatrix}$$



Choise of normalize α influences the convergence to global minima

For $\alpha > 1, A$ – has full rank and, accordingly, a unique solution

The selection of α allows you to find a solution to an ill-posed problem (the inverse of Tikhonov's regularization**, "balances" the variables in the cycle with the subgraph external to them***)

*Тихонов А. Н. О некорректных задачах линейной алгебры и устойчивом методе их решения // Доклады Академии наук СССР. — 1965. — Т. 163, № 3. — С. 591—594.

**М. И. Сумин, "О роли множителей Лагранжа и двойственности в некорректных задачах на условный экстремум. К 60-летию метода регуляризации Тихонова", Вестник российских университетов. Математика, 28:144 (2023), 414–435

***M. Chertkov et al. Interaction screening: efficient and sample-optimal learning of ising models. NIPS'16 pp. 2603-2611

Training Neural Networks

The training of a neural network to approximate the distribution $P(\sigma)$ can thus be viewed as minimizing the KL divergence. This is equivalent to minimizing the variational free energy:

$$\mathcal{F}(\theta) = \sum_{\sigma} Q_\theta(\sigma)\left(\mathcal{H}(\sigma) + \frac{1}{\beta}\log Q_\theta(\sigma)\right).$$

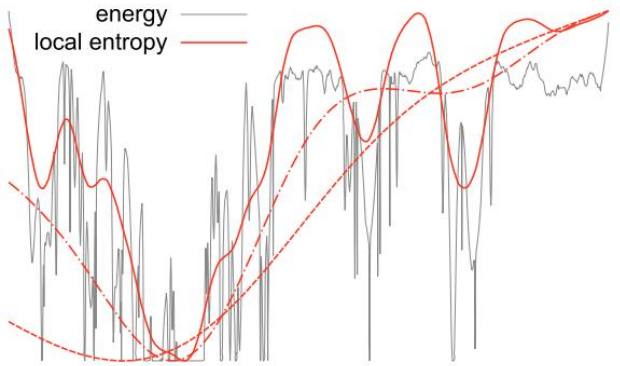
Minimizing this free energy involves adjusting the neural network parameters θ to ensure that $Q_\theta(\sigma)$ closely approximates the true distribution $P(\sigma)$. This process is analogous to learning the weights in the neural network, where the energy function $\mathcal{H}(\sigma)$ can be interpreted as a parameterized function (such as a neural network) that captures the dependencies between the variables.

Local entropy*

$$\Phi(\sigma, \beta, \theta(\gamma)) = log \sum_{\{\sigma\prime\}} e^{-\beta E(\sigma') - \gamma d(\sigma,\sigma')},$$

where $d(\cdot,\cdot)$ – distance between configuration, γ –weight, σ – reference spin configuration

Global extremum represented by symmetrical cycle in graph.
Local extremum around represented by odd degree connected cycle in graph.



Energy landscape compared with local entropy landscape

*Carlo Baldassia, Christian Borgsc , Jennifer T. Chayesc , Alessandro Ingrosso, Carlo Lucibello, Luca Saglietti , and Riccardo Zecchina Unreasonable effectiveness of learning neural networks: From accessible states and robust ensembles to basic algorithmic schemes *Proceedings of the National Academy of Sciences* (PNAS), 113 (48), **November 15, 2016,** E7655-E7662
https://www.pnas.org/doi/full/10.1073/pnas.1608103113
Carlo Baldassi, Alessandro Ingrosso, Carlo Lucibello, Luca Saglietti, and Riccardo Zecchina Subdominant Dense Clusters Allow for Simple Learning and High Computational Performance in Neural Networks with Discrete Synapses. Physical Review Letters 115, 128101, p. 1-5., 2015
Usatyuk V., Sapozhnikov D., Egorov S. Spherical and Hyperbolic Toric Topology-Based Codes On Graph Embedding for Ising MRF Models: Classical and Quantum Topology Machine Learning https://ui.adsabs.harvard.edu/abs/2023arXiv230715778U

Hierarchical tree (Negative Curvature topology of Graph) of spin glass state

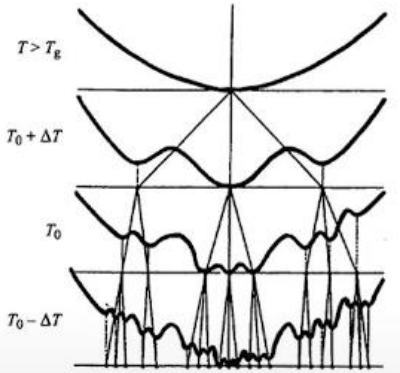
Def. 1 (Weighted non-backtracking matrix) *Given a graph $\mathcal{G}(\mathcal{V},\mathcal{E})$ and a function $f:\mathcal{E}\to\mathbb{R}$, so that $\forall e\in\mathcal{E}$, $f(e)=\omega_e$ is the weight corresponding to the edge e, the weighted non backtracking matrix $B\in\mathbb{R}^{2|\mathcal{E}|\times 2|\mathcal{E}|}$ is defined on the set of directed edges of $\mathcal{G}$ as*

$$B_{(ij),(k\ell)}=\delta_{jk}(1-\delta_{i\ell})\omega_{k\ell}.$$

Local (hierarchical) Hessian

Def. 2 (Bethe-Hessian matrix) *Given a graph $\mathcal{G}(\mathcal{V},\mathcal{E})$, a function $f:\mathcal{E}\to\mathbb{R}$ so that $\forall e\in\mathcal{E}$, $f(e)=\omega_e$ and a parameter $x\in\mathbb{C}\backslash\{\pm\omega_{ij}\}_{(ij)\in\mathcal{E}}$, the Bethe-Hessian matrix $H(x)\in\mathbb{C}^{n\times n}$ is defined as*

$$H_{ij}(x)=\left(1+\sum_{k\in\partial i}\frac{\omega_{ik}^2}{x^2-\omega_{ik}^2}\right)\delta_{ij}-\frac{x\omega_{ij}}{x^2-\omega_{ij}^2}.$$



The hyerarchical tree of spin-glass states

Since $\mathcal{G}$ is an undirected graph, $H(x)$ is symmetric but not Hermitian, unless $x\in\mathbb{R}$. The relation between the spectra of the matrices B and $H(x)$ is given by the Watanabe-Fukumizu formula.

Prop. 1 (Watanabe-Fukumizu) *Let $H(x)$ and B on the same graph $\mathcal{G}$ and for the same weighting function f. $x\in\mathbb{C}\backslash\{\pm\omega_{ij}\}_{(ij)\in\mathcal{E}}$:*

$$\det[xI_{2|\mathcal{E}|}-B]=\det[H(x)]\prod_{(ij)\in\mathcal{E}}(x^2-\omega_{ij}^2),$$

x in the spectrum of B, $\det[H(x)]=0$.

Yusuke Watanabe and Kenji Fukumizu. Loopy belief propagation, bethe free energy and graph zeta function. arXiv preprint arXiv:1103.0605, 2011
Iwao Sato, Hideo Mitsuhashi, and Hideaki Morita. A matrix-weighted zeta function of a graph. Linear and Multilinear Algebra, 62(1):114–125, 2014.
A. Saade, F. Krzakala, and L. Zdeborova, ``Spectral clustering of graphs with the Bethe Hessian,'' NIPS, 2014.
F. Krzakala, et al., Spectral redemption in clustering sparse networks, Proc. Natl. Acad. Sci. U.S.A. 110 (52) 20935-20940

Nishimori Temperature-Based Node Classification Algorithm

To perform accurate node clustering, knowledge of the Nishimori temperature β_{N} is crucial for obtaining a precise estimate of the true node classes σ. A powerful estimator of σ is derived from the signs of the entries of the eigenvector $\mathbf{x}$ of the Bethe-Hessian matrix $H_{\beta_{\mathrm{N}},\tilde{J}}$, associated with its smallest amplitude eigenvalue, which is close to zero.First, consider $\mathbf{y}$, the eigenvector associated with the smallest eigenvalue of $H_{\beta_{\mathrm{N}},J}$. Let $A \in \{0,1\}^{n\times n}$ be the symmetric adjacency matrix of $\mathcal{G}$, defined by $A_{ij} = 1$ if $(ij) \in \mathcal{E}$, and $A_{ij} = 0$ otherwise. Let $D \in \mathbb{N}^{n\times n}$ be the diagonal degree matrix $D = \mathrm{d}iag(A\mathbf{1}_n)$.

$$\mathbb{E}\left[H_{\beta_{\mathrm{N}},J}\right] = I_n + \mathbb{E}\left[\frac{\mathrm{th}(\beta J_{ij})}{1-\mathrm{th}^2(\beta J_{ij})}\right](D - A).$$

Alg. 2 The Nishimori-Bethe relation for node classification

Input: Weighted adjacency matrix of a graph $\tilde{J} \in \mathbb{R}^{n\times n}$, precision error $\varepsilon \in \mathbb{R}$

Return:Value of $\hat{\beta}_{\mathrm{N}} \in \mathbb{R}^{+}$, estimated label vector $\hat{\sigma} \in \{-1,1\}^n$

- Shift the non-zero $\tilde{J}_{ij}$ as: $\tilde{J}_{ij} \leftarrow \tilde{J}_{ij} - \frac{1}{2|\mathcal{E}|}\mathbf{1}_n^T\tilde{J}\mathbf{1}_n$
- Compute $\hat{\beta}_{\mathrm{N}} \leftarrow$ Compute_$\hat{\beta}_{\mathrm{N}}$
- Compute $H_{\beta_t,J} = \delta_{ij}\left(1 + \sum_{k\in\partial i}\frac{\mathrm{th}^2(\beta J_{ik})}{1-\mathrm{th}^2(\beta J_{ik})}\right) - \frac{\mathrm{th}(\beta J_{ij})}{1-\mathrm{th}^2(\beta J_{ij})}$.
- Compute $\mathbf{x} \leftarrow$ the eigenvector associated to $\gamma_{min}(H_{\hat{\beta}_{\mathrm{N}},\tilde{J}})$
- Estimate $\hat{\sigma}$ as the output of 2-class *k-means* on the entries of $\mathbf{x}$

return: β_t, $\hat{\sigma}$.

Lorenzo Dall'Amico et al Nishimori meets Bethe: a spectral method for node classification in sparse weighted graphs J. Stat. Mech. (2021) 093405

Simulated Annealing method for Quasi-Cyclic LDPC bipartite graph construction

Algorithm 1 Simulated Annealing method for construction of QC-LDPC codes

Require: $M(\mathbf{H})$−mother matrix, L− circulant size, g-girth of lifted matrix, EMD-minimal EMD value, $Iter$− maximal number of iterations, $seed$- a seed to be used in a pseudo-random number generator, $Temp$− initial value of temperature

$Nstep = 0$

$i, j = rnd(seed)$

for $it = 0; it <= Iter; it = it + 1$ **do**

  while $M_{ij}(\mathbf{H}) = 0$ **do**

    $i, j = rnd(seed)$

  end while

  for $k = 0; k <= L - 1; k = k + 1$ **do**

    $\Theta_k = enumcirccycles(i, j, k, g, EMD)$

$$P(k) = w(k) / \sum_{m=0}^{L-1} w(m),$$

$$w(k) = e^{\frac{-\Theta_k}{Temp}},$$

where Θ_k-number of cycles through $E_{ij}(\mathbf{H})$-CPM with shift value k, $P(k)$− probability of k-shift CPM value choice, $w(k)$−probability weight function;

  end for

$$\Phi = enumcycles(E(\mathbf{H}), g, EMD),$$

where Φ-total number of cycles in exponent matrix $E(\mathbf{H})$;

$$E_{ij}(\mathbf{H}) = rndshift(P, Temp)$$

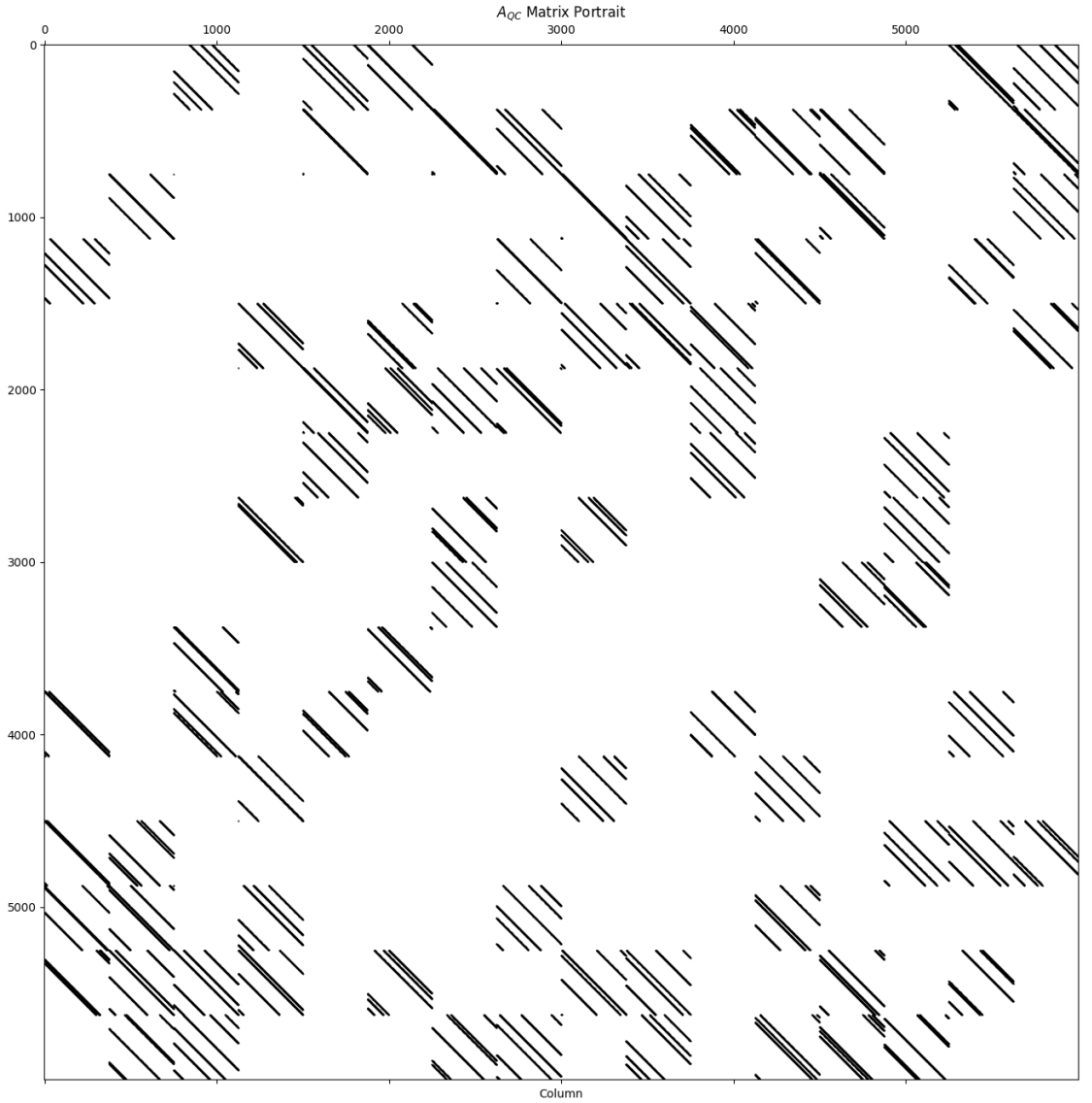
Nstep=Nstep+1

$$Temp = \eta \frac{\Phi}{Nstep^2},$$

where η-some constant value;

end for

return $E(\mathbf{H})$



V. Usatyuk and I. Vorobyev, "Simulated Annealing Method for Construction of High-Girth QC-LDPC Codes," *2018 41st International Conference on Telecommunications and Signal Processing (TSP)*, Athens, Greece, 2018, pp. 1-5

The Weighted Laplacian Matrix (Laplacian Method)

A classical spectral clustering method for weighted graphs uses the *weighted Laplacian* matrix $L = \overline{D} - \tilde{J}$, where $\overline{D} = \mathrm{d}iag(|\tilde{J}|\mathbf{1}_n)$. The eigenvector associated with the smallest eigenvalue of L provides a relaxed solution to the NP-hard optimization problem of signed ratio-cut graph clustering.

For *signed graphs*, where $\tilde{J}$ entries are ± 1, the Bethe-Hessian matrix $H^{signed}_{\beta,\tilde{J}}$ is defined as:

$$H^{signed}_{\beta,\tilde{J}} = (1 - \mathrm{t}h^2(\beta))I_n + \mathrm{t}h^2(\beta)D - \mathrm{t}h(\beta)\tilde{J}.$$

As $\beta_{\mathrm{N}} \to \infty$, $H^{signed}_{\beta_{\mathrm{N}},\tilde{J}} \to L$. The signed Laplacian can be seen as the zero temperature limit of the Bethe-Hessian matrix.

The Mean Field Approximation (Naïve Bayes)

Using a naive mean field approximation, the probability distribution is:

$$p_{\mathbf{m}}(\mathbf{s}) = \prod_{i \in \mathcal{V}} \frac{1+m_i s_i}{2},$$

where m_i is the average of s_i over the distribution.

The associated variational free energy is:

$$\tilde{F}^{\mathrm{MF}}_{\tilde{J},\beta}(\mathbf{m}) = -\sum_{(ij)\in\mathcal{E}} \beta \tilde{J}_{ij} m_i m_j + \sum_{i \in \mathcal{V}} \sum_{s_i} \frac{1+m_i s_i}{2} log\left(\frac{1+m_i s_i}{2}\right).$$

The Hessian of the free energy at the paramagnetic point $\mathbf{m} = \mathbf{0}$ is:

$$H^{\mathrm{MF}}_{\beta,\tilde{J}} = I_n - \beta \tilde{J}.$$

Thus, the eigenvectors of $H^{\mathrm{MF}}_{\beta,\tilde{J}}$ are simply the eigenvectors of $\tilde{J}$, making β irrelevant.

The Spin Glass Bethe-Hessian

The spin glass Bethe-Hessian matrix, with $\beta = \beta_{SG}$, can also achieve non-trivial clustering as soon as theoretically possible. The temperature β_{SG} is estimated by solving:

$$c\mathbb{E}[\mathrm{t}h^2(\beta_{SG}\tilde{J}_{ij}\sigma_i\sigma_j)] = c\mathbb{E}[\mathrm{t}h^2(\beta_{SG}\tilde{J}_{ij})] = 1.$$

However, for realistic heterogeneous graphs, this choice may be suboptimal.

Alaa Saade, Marc Lelarge, Florent Krzakala, and Lenka Zdeborov´a. Clustering from sparse pairwise measurements. In 2016 IEEE International Symposium on Information Theory (ISIT), pages 780–784. IEEE, 2016.

Квазициклические Низкоплотностные коды и их ключевые свойства

$$H = \begin{matrix} v_1 & v_2 & v_3 & v_4 & v_5 & v_6 \end{matrix} \\ \begin{bmatrix} 1 & 1 & 0 & 0 & 1 & 0 \\ 1 & 0 & 0 & 1 & 0 & 1 \\ 1 & 1 & 1 & 0 & 0 & 1 \end{bmatrix} \begin{matrix} c_1 \\ c_2 \\ c_3 \end{matrix}$$

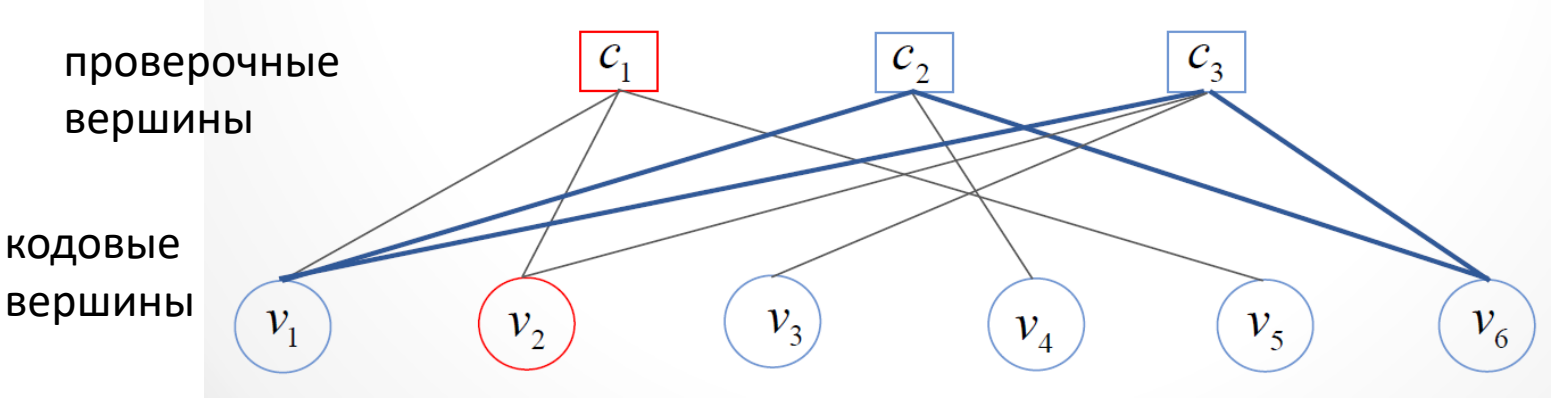
Низкоплотностные-коды (Low Density Parity Check,LDPC-коды), – это блочный линейный код размерностью k и длиной кодового слова n, задаваемый проверочной матрицей H размерностью $(n\text{-}k)\cdot n$, имеющей небольшую плотность отличных от нуля символов. Матрица размера $(n\text{-}k)\cdot n$ содержит порядка n проверок.

Каждая строка проверочной матрицы H задает уравнение проверки на четность:

$$\sum_{l=0}^{n-1} v_l \cdot h_{\mathrm{j},l} = 0$$

где j – номер строки проверочной матрицы (номер проверочного уравнения); l – номер символа кодового слова; $h_{j,l}$ – элемент проверочной матрицы.

Граф Таннера –это двудольный граф, соответствующей проверочной матрице LDPC- кода



Замкнутый простой путь в Таннер графе с узлами $v_1 \rightarrow c_2 \rightarrow v_6 \rightarrow c_3 \rightarrow v_1$ образуют цикл длины 4

Квазициклические низкоплотностные коды (Quasi Cyclic-LDPC)

$$H = \begin{array}{c} \begin{matrix} v_1 & v_2 & v_3 & v_4 & v_5 & v_6 \end{matrix} \\ \left[\begin{array}{cccccc} 1 & 0 & 0 & 1 & 0 & 1 \\ 0 & 1 & 1 & 0 & 1 & 0 \\ \hline 1 & 0 & 0 & 0 & 1 & 0 \\ 0 & 1 & 0 & 0 & 0 & 1 \end{array}\right] \end{array} \begin{matrix} c_1 \\ c_2 \\ c_3 \\ c_4 \end{matrix}$$

$$H_{QC} = \begin{bmatrix} I^0 & I^1 & I^1 \\ I^0 & I^{-1} & I^0 \end{bmatrix} = \begin{bmatrix} 0 & 1 & 1 \\ 0 & -1 & 0 \end{bmatrix}$$

Традиционно используются матрицы веса 1.

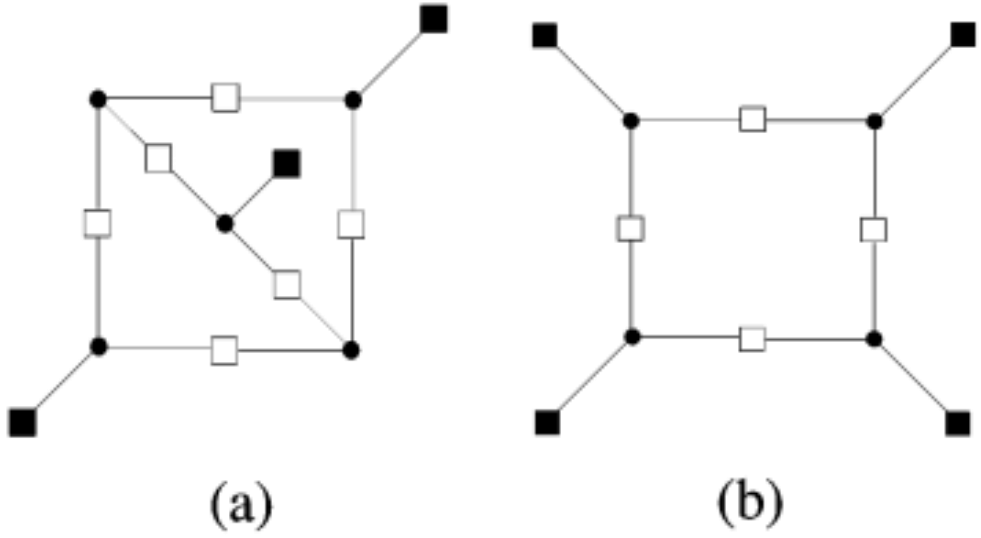
Квазициклический LDPC (коды QC-LDPC) - LDPC-коды с матрицей проверки на четность, определяемой структурированной блочной подматрицей - матрицей циклических перестановок.

Матрица циклических перестановок размера 2x2, веса 1 и ее сжатая форма представления

$$I^0 = \begin{bmatrix} 1 & 0 \\ 0 & 1 \end{bmatrix}, I^1 = \begin{bmatrix} 0 & 1 \\ 1 & 0 \end{bmatrix}, I^{-1} = \begin{bmatrix} 0 & 0 \\ 0 & 0 \end{bmatrix}$$

Свойства низкоплотностных кодов. Трэппин-сеты (a,b), a,b∈ ℤ

На свойства низкоплотностых кодов отрицательно влияют циклы в графе Таннера, образующих «Трэппин-сеты». (Trapping set, TS,) или (*a*,*b*)-подграфы (подграфы в графе Таннера, состоящие из *a* символьных узлов, *b* из которых инцидентны проверочным узлам с нечетными степенями). На рисунке ниже изображены два (a,b)-подграфа. Левый подграф образован пересечением трех циклов длины 8, правый одним циклом длины 8.



Trapping sets: (a) $(5,3)$ and (b) $(4,4)$

●-символьные узлы,

□ - проверочные узлы с четной степенью инцидентности

■ - проверочные узлы с нечетной степенью инцидентности

NP-сложная задача

A. McGregor and O. Milenkovic, "On the Hardness of Approximating Stopping and Trapping Sets in LDPC Codes," *2007 IEEE Information Theory Workshop*, Tahoe City, CA, USA, 2007, pp. 248-253

Поиск Трэппин-сетов (a,b), b≠ 0

*

Полный перебор требует анализ

$\binom{4896}{48} \approx 8.2812e+115$

Gurobi 7.5.1 solver линейного программирования 32 threads

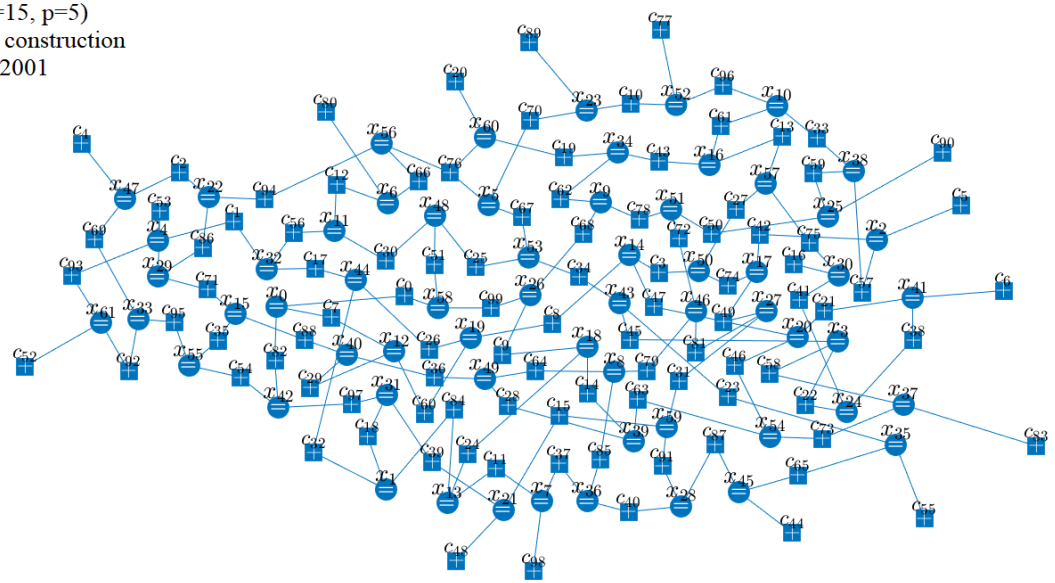
SS 48 в коде Маргулиса (4896,2474) Margulis за 700 451 с.

*Velasquez A., Subramani K. , Wojciechowski P. , On the complexity of and solutions to the minimum stopping and trapping set problems,Theoretical Computer Science, Volume 915, 2022, Pages 26-44,

Полный перебор требует анализ

$\binom{4896}{62} \approx 1.2701E+143$

TS(62, 16) at Margulis(4896, 2448)
Ramanujan graph(q=15, p=5)
Vontobel-Rosenthal construction
Allerton 2000, ISIT 2001
true_K=2474



Предложенный метод**

TS(62,16) with
TS variable nodes, (x_0)...(x_end) on Fig: 6 41 65 159 364 546 719 769 880 970 981 1097 1132 1140 1164 1274 1279 1385 1508 1561 1625 1681 1716 1819 1823 1979 1988 2194 2216 2369 2460 2506 2738 2765 2795 2855 2950 2976 3024 3154 3182 3271 3444 3566 3575 3836 3940 4012 4109 4141 4162 4355 4396 4401 4463 4547 4581 4671 4710 4800 4859 4864

CPLEX 12.8 LP solver 32 threads
Total (root+branch&cut) = 3.67 секунды. (292.92 ticks)

V. S. Usatyuk, "Low Error Floor QC-LDPC Codes Construction Using Modified Cole's Trapping Sets Enumerating," *2023 25th International Conference on Digital Signal Processing and its Applications (DSPA)*, Moscow, Russian Federation, 2023, pp. 1-6

Кодовые слова - Трэппин-сеты TS(a,0), поиск слов малого(минимального веса)

Задача оценки кодового расстояния

Пусть матрица H, размера $(n-k) \times n$, заданная над F_q определяет линейное отображение $F_q^n \to F_q^{n-k}$ такое, что $x \to Hx$, $\ker(H) = \left\{ c \in F_q^n \middle| Hc = 0 \right\}$.

Оценка кодового расстояния линейного блочного кода: Найти вектор с минимальным число отличных от нуля координатных компонент, $c \in ker(H) \setminus \{0\}$?

d_{min}-кодовое расстояние

Пример поиска кодового расстояния - Трэппин-сета (d_{min},0)

$$GF(2)$$

$$Hc = \begin{pmatrix} 1 & 1 & 1 & 0 & 0 \\ 0 & 1 & 0 & 1 & 1 \\ 0 & 0 & 1 & 1 & 1 \end{pmatrix} \begin{bmatrix} c_1 \\ c_2 \\ c_3 \\ c_4 \\ c_5 \end{bmatrix} = \begin{bmatrix} 0 \\ 0 \\ 0 \end{bmatrix}$$

$$\bar{c}_0^{\,T} = [0 \quad 0 \quad 0 \quad 0 \quad 0] \qquad w(\bar{c}_0^{\,T}) = \left\| \bar{c}_0^{\,T} \right\| = 0$$

$$\bar{c}_1^{\,T} = [0 \quad 0 \quad 0 \quad 1 \quad 1] \qquad w(\bar{c}_1^{\,T}) = \left\| \bar{c}_1^{\,T} \right\| = 2$$

$$\bar{c}_2^{\,T} = [0 \quad 1 \quad 1 \quad 0 \quad 1] \qquad w(\bar{c}_2^{\,T}) = \left\| \bar{c}_2^{\,T} \right\| = 3$$

$$\bar{c}_3^{\,T} = [0 \quad 1 \quad 1 \quad 1 \quad 0] \qquad w(\bar{c}_3^{\,T}) = \left\| \bar{c}_3^{\,T} \right\| = 3$$

Полином весового спектра кода

$$f(z) = 1 + z^2 + 2z^3$$

Слово веса 0

Весовой спектра кода $W(c) = [1, 0, 1, 2, 0, 0]$

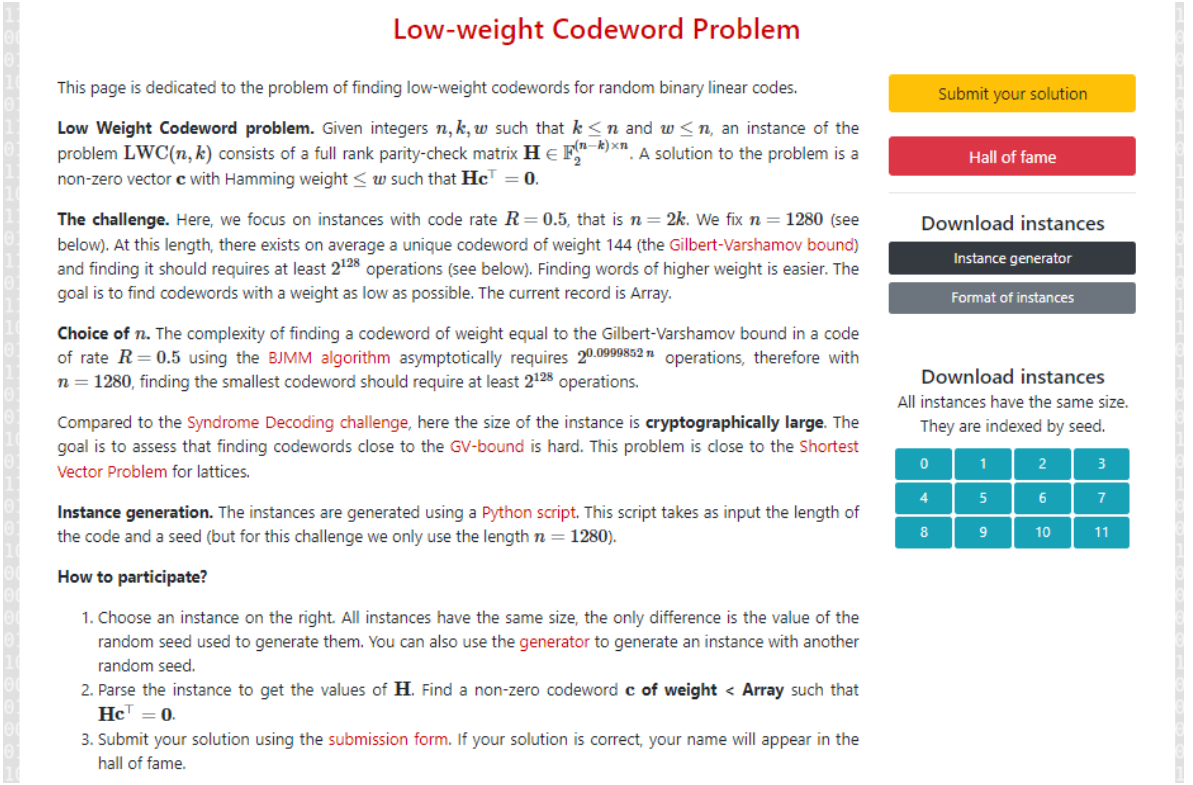
Слово веса 2

Слова веса 3

$$\mathrm{d}_{\min} = 2;$$

Спектр связности можно улучшать и в частности TS(d_{min},0): $d_{min} \rightarrow max$

Международный конкурс по оценки кодового расстояния от объединения: Французского национального центра научных исследований (CNRS), Национального института исследований в области цифровых наук и технологий (Inria Paris), национального исследовательский институт математики и информатики в Нидерландах (CWI)



Low-weight Codeword Problem

This page is dedicated to the problem of finding low-weight codewords for random binary linear codes.

Low Weight Codeword problem. Given integers n, k, w such that $k \leq n$ and $w \leq n$, an instance of the problem $\mathbf{LWC}(n, k)$ consists of a full rank parity-check matrix $\mathbf{H} \in \mathbb{F}_2^{(n-k)\times n}$. A solution to the problem is a non-zero vector $\mathbf{c}$ with Hamming weight $\leq w$ such that $\mathbf{Hc}^\top = \mathbf{0}$.

The challenge. Here, we focus on instances with code rate $R = 0.5$, that is $n = 2k$. We fix $n = 1280$ (see below). At this length, there exists on average a unique codeword of weight 144 (the Gilbert-Varshamov bound) and finding it should requires at least 2^{128} operations (see below). Finding words of higher weight is easier. The goal is to find codewords with a weight as low as possible. The current record is Array.

Choice of n. The complexity of finding a codeword of weight equal to the Gilbert-Varshamov bound in a code of rate $R = 0.5$ using the BJMM algorithm asymptotically requires $2^{0.0999852\,n}$ operations, therefore with $n = 1280$, finding the smallest codeword should require at least 2^{128} operations.

Compared to the Syndrome Decoding challenge, here the size of the instance is **cryptographically large**. The goal is to assess that finding codewords close to the GV-bound is hard. This problem is close to the Shortest Vector Problem for lattices.

Instance generation. The instances are generated using a Python script. This script takes as input the length of the code and a seed (but for this challenge we only use the length $n = 1280$).

How to participate?

1. Choose an instance on the right. All instances have the same size, the only difference is the value of the random seed used to generate them. You can also use the generator to generate an instance with another random seed.
2. Parse the instance to get the values of $\mathbf{H}$. Find a non-zero codeword **c of weight < Array** such that $\mathbf{Hc}^\top = \mathbf{0}$.
3. Submit your solution using the submission form. If your solution is correct, your name will appear in the hall of fame.

Submit your solution

Hall of fame

Download instances

Instance generator

Format of instances

Download instances

All instances have the same size. They are indexed by seed.

0	1	2	3
4	5	6	7
8	9	10	11

Best solutions

Weight	Authors	Algorithm	Details
214	Vasiliy Usatyuk	Lattice: Kannan emb, SBP (SBP), SVP	See details
215	Samuel Neves	-	See details
220	Valentin Vasseur	Dumer	See details

https://decodingchallenge.org/low-weight/

NP-сложная задача

Best solutions

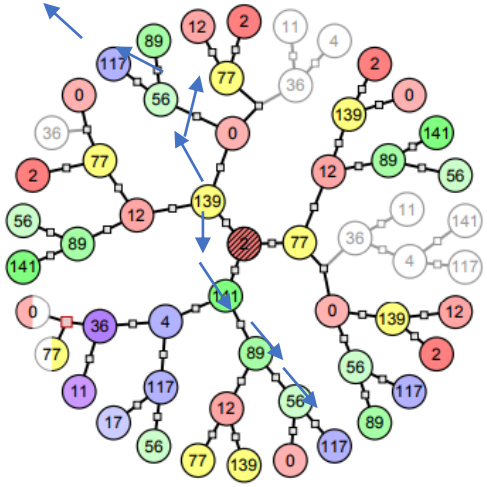
Weight	Authors	Algorithm	Details
211	Leo Ducas, Marc Stevens	Hybrid Wagner-Babai using Code Reduction [eprint:2020/869]	See details
212	Vasiliy Usatyuk	Lattice:SBP(BKZ), SVP	See details
214	Vasiliy Usatyuk	Lattice: Kannan emb, SBP (SBP), SVP	See details

I. Dumer, D. Micciancio and M. Sudan, "Hardness of approximating the minimum distance of a linear code," in *IEEE Transactions on Information Theory*, vol. 49, no. 1, pp. 22-37, Jan. 2003

Модель Шеррингтона-Киркпатрика (Sherrington-Kirkpatrick)

$$H_E = -\sum_{i=1,\dots,N}\ \sum_{a=1,\dots,M} C_{ij} J_{ij} \sigma_i \sigma_j$$

C_{ij} - матрица связности, элемент которого равен 1, если два спина взаимодействуют, и 0 в противном случае,
J_{ij}- вес (потенциал) взаимодействия между спинами, σ_i- спины модели Изинга



H_E можно преобразовать:

$$H_{SK} = -\sum_p \sum_{i_{1,\dots,i_p}=1,\dots,n} C^{(p)}_{i_{1,\dots,i_p}} J_{i_{1,\dots,i_p}} \sigma_{i_1} \dots \sigma_{i_p},$$

где $J_0^{(p)}, \Delta\ J^2_{(p)}$ М.О. и дисперсия J_{ij}.

Проверочная матрица квазициклического кода H определяет $C^{(p)}_{i_{1,\dots,i_p}}$.

x_i n-битное кодовое слов кодирующее спин $\sigma_i = 2x_i - 1$.

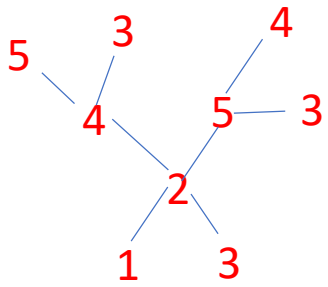
Взаимодействие пар спинов (без кратных ребер) происходит на глубину p.

$J_{i_{1,\dots,i_p}}$ - двух спинового взаимодействия и принимаются как независимые случайные величины с известным распределением вероятностей.

Число слоев p –это число ***Пар спинов соединённых Проверочной матрицей кода. Число итераций декодирования***

$$\bar{x}_0^T = [0\quad 0\quad 0\quad 0\quad 0], \bar{x}_1^T = [0\quad 0\quad 0\quad 1\quad 1]$$
$$\bar{x}_2^T = [0\quad 1\quad 1\quad 0\quad 1], \bar{x}_3^T = [0\quad 1\quad 1\quad 1\quad 0]$$

Множество спинов соответствующих кодовым словам является состояниями минимальной энергии Гамильтониана



$$Hx = \begin{matrix} 1 & 2 & 3 & 4 & 5 \end{matrix}\begin{pmatrix} 1 & 1 & 1 & 0 & 0 \\ 0 & 1 & 0 & 1 & 1 \\ 0 & 0 & 1 & 1 & 1 \end{pmatrix}\begin{bmatrix} x_1 \\ x_2 \\ x_3 \\ x_4 \\ x_5 \end{bmatrix} = \begin{bmatrix} 0 \\ 0 \\ 0 \end{bmatrix}$$

Sourlas, N.. Spin Glasses, Error-Correcting Codes and Finite-Temperature Decoding. EPL (Europhysics Letters). 25. 159. (2007)

Граф Таннер (мультиграф, в случае наличия не тривиального X_1) Квазициклического низкоплотностного кода соответствует калибровочному поля

Кодовые вершины | Ребра Графа | Проверочные узлы

$$X = (X_0, X_1, X_2)$$

$$X_1 \subset X_0 \times X_2$$

$$(j,\alpha) \in X_1 \Leftrightarrow j \in \alpha \qquad \alpha = 1, \cdots, M$$

$$|X_0| = N \quad |X_2| = M$$

$$|X_1| = \sum_{j \in X_0} m_j = \sum_{\alpha \in X_2} n_\alpha$$

Кодовые слов (закодированные в спин)

$$\sigma = \{\sigma_j\}_{j \in X_0}; \sigma_j = \pm 1$$

Эквивалентные коды (калибровочная инвариантность)

$$X \sim X' \Leftrightarrow C_X = C_{X'}$$

$$C_X = \left\{ \sigma \middle| \prod_{j \in \alpha} \sigma_j = 1, \forall \alpha \in X_2 \right\}$$

Калибровочная группа

$$G = GL(M, F_2)$$

Эквивалентные коды (калибровочная инвариантность)

$$Hx = \begin{array}{c} \begin{matrix} 1 & 2 & 3 & 4 & 5 \end{matrix} \\ \begin{pmatrix} 1 & 1 & 1 & 0 & 0 \\ 0 & 1 & 0 & 1 & 1 \\ 0 & 0 & 1 & 1 & 1 \end{pmatrix} \end{array} \begin{bmatrix} x_1 \\ x_2 \\ x_3 \\ x_4 \\ x_5 \end{bmatrix} = \begin{bmatrix} 0 \\ 0 \\ 0 \end{bmatrix}$$

$$H'x = \begin{array}{c} \begin{matrix} 1 & 2 & 3 & 4 & 5 \end{matrix} \\ \begin{pmatrix} 1 & 0 & 1 & 1 & 1 \\ 0 & 1 & 0 & 1 & 1 \\ 0 & 0 & 1 & 1 & 1 \end{pmatrix} \end{array} \begin{bmatrix} x_1 \\ x_2 \\ x_3 \\ x_4 \\ x_5 \end{bmatrix} = \begin{bmatrix} 0 \\ 0 \\ 0 \end{bmatrix}$$

Графы не эквивалентны, мы добавили цикл

Статистическая модель независимых ошибок спина

Намагниченности спинового стекла

Вероятность результата измерения

$$P_j(\zeta) = \left(2\pi s^2\right)^{-N/2} \int dh \delta\left(m_j(h) - \zeta\right) \exp\left(-F(h)\right)$$

Вероятность индивидуальных ошибок спина

$$B_j = \int_{-\infty}^{0} d\zeta P_j(\zeta)$$

Для симметрично распределенного Гауссова шума

$$F(h) = \left(2s^2\right)^{-1} \sum_{j \in X_0} \left(h_j - s^2\right)^2$$

Отношения сигнал-шум (ОСШ), Температура

В функции плотности вероятности преобладает наиболее вероятная конфигурация шума инстантон (седловая точка)

$$\frac{\partial}{\partial h}\left(F(h) - \lambda\left(m_j(h) - \zeta\right)\right)\Big|_{h=\bar{h}} = 0$$

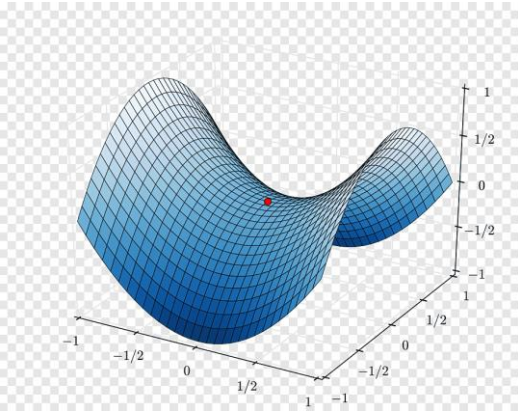
$$m_j\left(\bar{h}\right) - \zeta = 0$$

Множитель Лагранжа

$$P_j(\zeta) \approx \kappa^{-1}(\zeta) \exp\left(-F\left(\bar{h}\right)\right)$$

$$B_j \approx \kappa_0^{-1} \exp\left(-F\left(\bar{h}\right)\right)$$

TS(a,0)
Кодовые слова



V. Chernyak, M. Chertkov, M. Stepanov, and B. Vasić, "Local theory of BER for LDPC codes: instantons on a tree," *Los Alamos Workshop on Applications of Statistical Physics to Coding Theory,* Jan. 2005.

Топологический инварианты и кривизна Изинг модели полученной из графов Таннера (мультиграфа в случае веса циркулянта больше 1)

Локальная система:

Каждый узел имеет древовидную “окрестность”

$$j \in U_j^{(l)} \subset X$$

Универсальное покрывающее дерево

(аналогично римановым поверхностям)

$$p : \overline{X} \to X$$

Фундаментальная группа

свободная группа с генератором

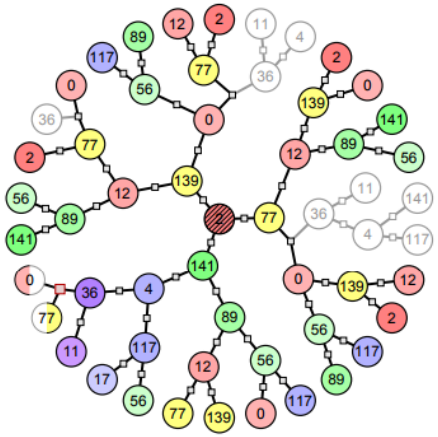
$$\overline{X}(a;l) \subset \overline{X}; a \in p^{-1}(j) \cong \pi_1(X) \cong F(g)$$

Теорема Черна-Гаусса-Бонне

Род поверхности (genus)

$$\sum_{j \in X_0} \frac{2 - m_j}{2} + \sum_{\alpha \in X_2} \frac{2 - n_\alpha}{2} = 1 - g$$

Локальная кривизна



Род поверхности 3 (genus)

Графы с постоянной кривизной

$$m_j = m \qquad mN = nM$$

$$n_\alpha = n \qquad \overline{X} \cong Y(m,n)$$

Покрывающее дерево универсально и обладает высокой симметрией

Блочной структурой

Wiberg Codes and iterative decoding on general graphs 1995

Weiss Correctness of Local Probability Propagation in Graphical Models with Loops 2000

Knill A GRAPH THEORETICAL GAUSS-BONNET-CHERN THEOREM arXiv:1111.5395 2011

Forney, "Codes on Graphs: Models for Elementary Algebraic Topology and Statistical Physics 2018

Meshulam, Roy. (2018). Graph codes and local systems.

Квази-инстантоны – псевдокодовые слова TS(a,b)

Дискретные значения спина

$$A(\bar{\eta}) \subset \bar{X}(0;l)$$

$$\sigma_j = \prod_{i \in \gamma_j}^{i \neq j} \operatorname{sgn}(\bar{\eta}_i); j \in A_0(\bar{\eta}) \qquad (A, \sigma)$$

Намагниченность спинового стекла

$$\bar{\eta}_0 = \bar{h}_0 + \sum_{k \in A_0}^{k \neq 0} s_k \bar{h}_k = \sum_{i \in X_0} n_i \bar{h}_i \qquad s_k = \prod_{a \in C_k^{(0)}}^{a \neq 0} \sigma_a$$

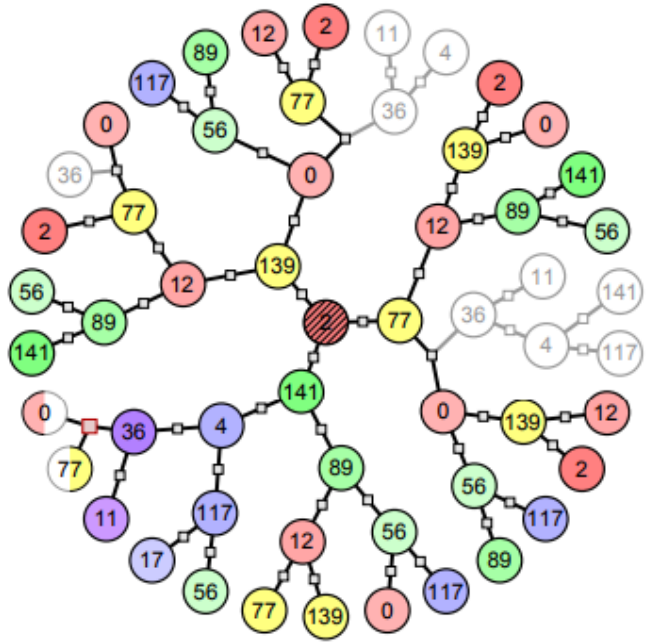
TS(a,b), b≠ 0

Квази-инстантоны (субоптимальные решения в результате неполноты ранга *решаемой системы, циклов в графе)*

$$\frac{\partial}{\partial \bar{h}_j} \sum_{i \in X_0} \left(\frac{s^2 (\bar{h}_i - 1)^2}{2} - \lambda n_i \bar{h}_i \right) = 0; \sum_{i \in X_0} n_i \bar{h}_i = 0 \qquad m_j = s^2 \bar{\eta}_0$$

$$\bar{h}_j = 1 - n_j \left(\sum_{k \in X_0} n_k \right) \left(\sum_{k \in X_0} n_k^2 \right)^{-1} \qquad \bar{S}(n) = \frac{1}{2} \left(\sum_{k \in X_0} n_k \right)^2 \left(\sum_{k \in X_0} n_k^2 \right)^{-1}$$

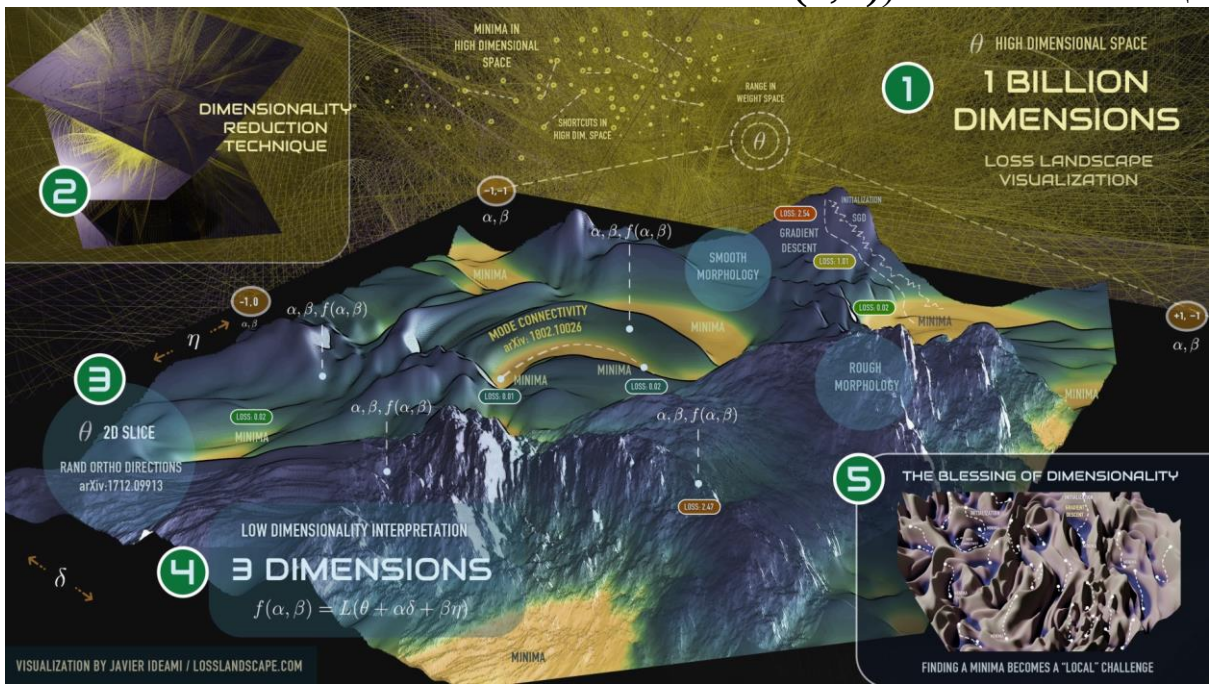
Вместо глобального минимума множество локальных



Дерево квазициклического Низкоплотносного кода Таннера [155,64,20] для p=4 слоев (после 4 итераций)

Beom Jun Kim* Performance of networks of artificial neurons: The role of clustering. PHYSICAL REVIEW E 69, 045101(R) (2004)

Ландшафт потерь квантовых моделей на калибровочном поле заданном Низкоплотностным кодом.
“Гладкость” (хорошая обусловленность), возникает при улучшение спектров связности (* устранения треппин-сетов TS(a,b)) и максимизации кодового расстояния TS(a,0)



*Ландшафты поверхностей образуются в результате формирования ячеек вокруг кодовых слов TS(a,0) и псевдокодовых слов (TS(a,b)) образованных циклами

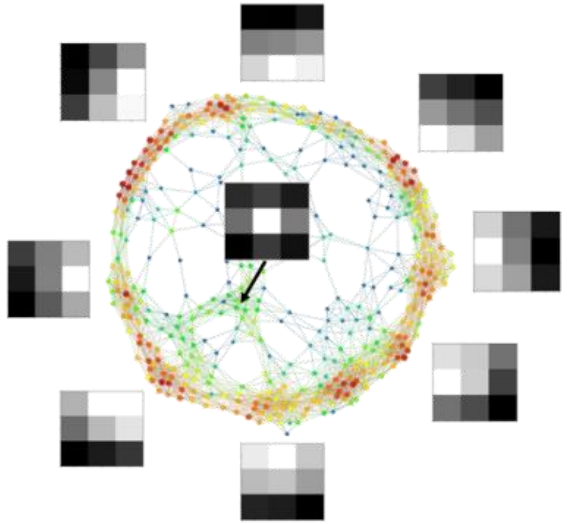


Figure 6: CIFAR-10 layer 2, gray scale

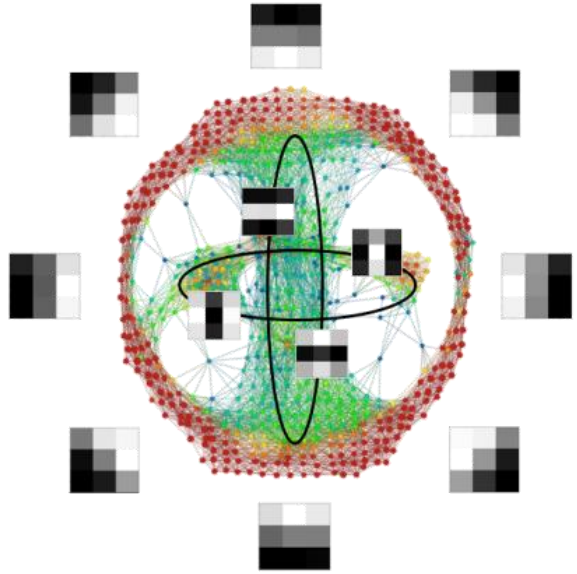


Figure 7: First layer, CIFAR-10

** Топология структуры (комплекс) активации тензоров глубокой нейронной сети (CNN, 3x3), обученной на CIFAR-10

*Beom Jun Kim Performance of networks of artificial neurons: The role of clustering. PHYSICAL REVIEW E 69, 045101(R) (2004)

** [1712.09913] Visualizing the Loss Landscape of Neural Nets (arxiv.org)

***Carlsson, G., Gabrielsson, R.B. (2020). Topological Approaches to Deep Learning. In: Baas, N., Carlsson, G., Quick, G., Szymik, M., Thaule, M. (eds) Topological Data Analysis. Abel Symposia, vol 15. Springer