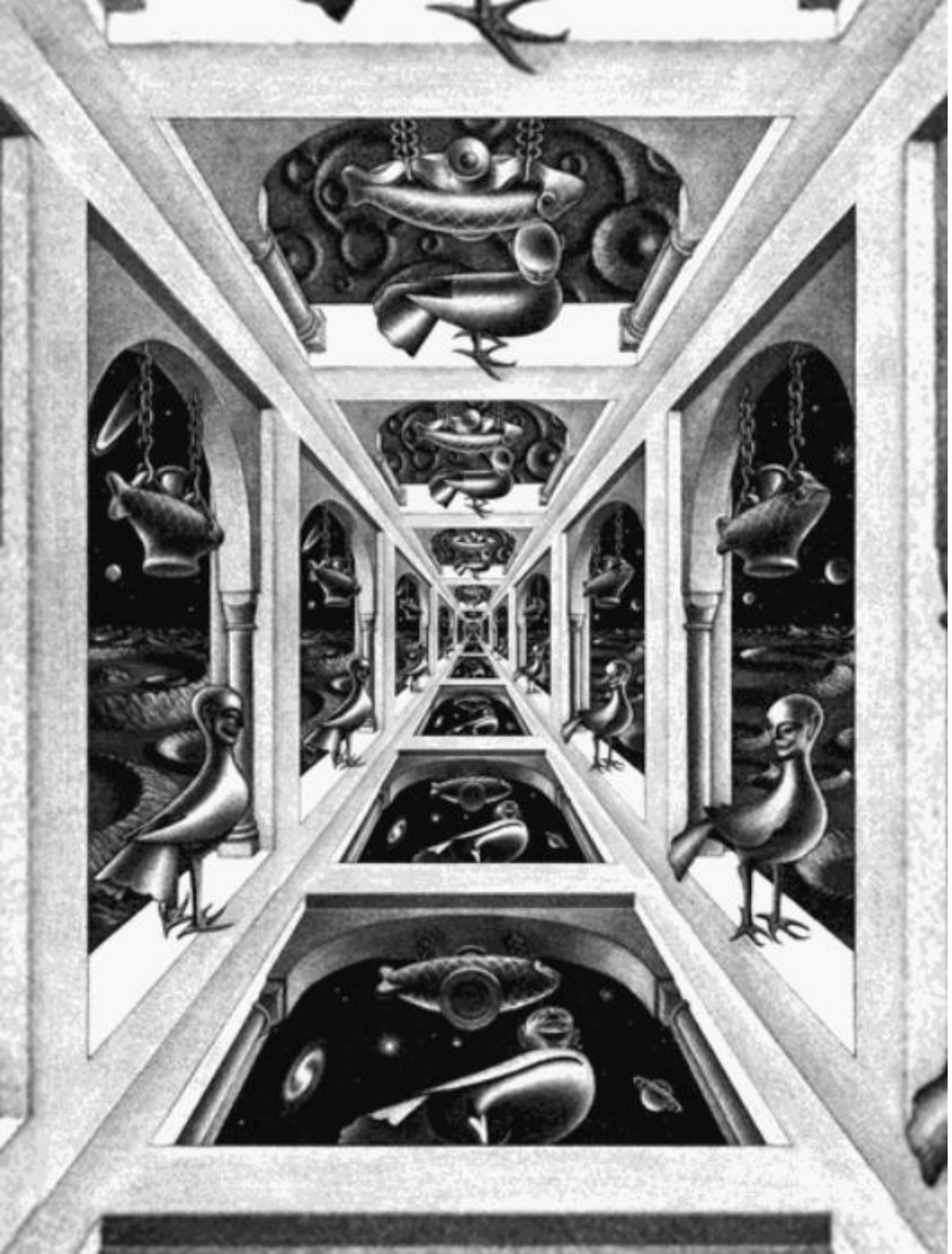


AI goes MAD^2

In Search of Cosmic Topology with AI

Andrius Tamošiūnas
Case Western Reserve University
IFT UAM-CSIC
COMPACT Collaboration

COMPACT CWRU ift

COMPACT Collaboration


Mikel Martin Barandiaran


Valeri Vardanyan


Yashar Akrami


Glenn Starkman


Deyan Mihaylov


Fernando Cornet-Gomez


Javier Carrón Duque


Samanta Saha


Johannes Eskilt


Özenç Güngör


Stefano Anselmi


Quinn Taylor


James Mertens


Thiago Pereira


Amirhossein Samandar


Andrew Jaffe


Arthur Kosowksy


Craig Copi

Cosmic Topology

- A key goal of cosmic topology: to measure the **shape** of the Universe.
- I.e. is the Universe:
  - ❖ Finite or infinite?
  - ❖ Open or closed?
  - ❖ Simply or multiply-connected?
  - ❖ Orientable or not?
- If we model space-time as a **manifold**, what is the topology of that manifold?
- If the Universe is **flat**: 17 allowed non-trivial topologies ($\boldsymbol{E_1}$-$\boldsymbol{E_{17}}$)(Riazuelo et al. 2004).

Image: Wolfram MathWorld

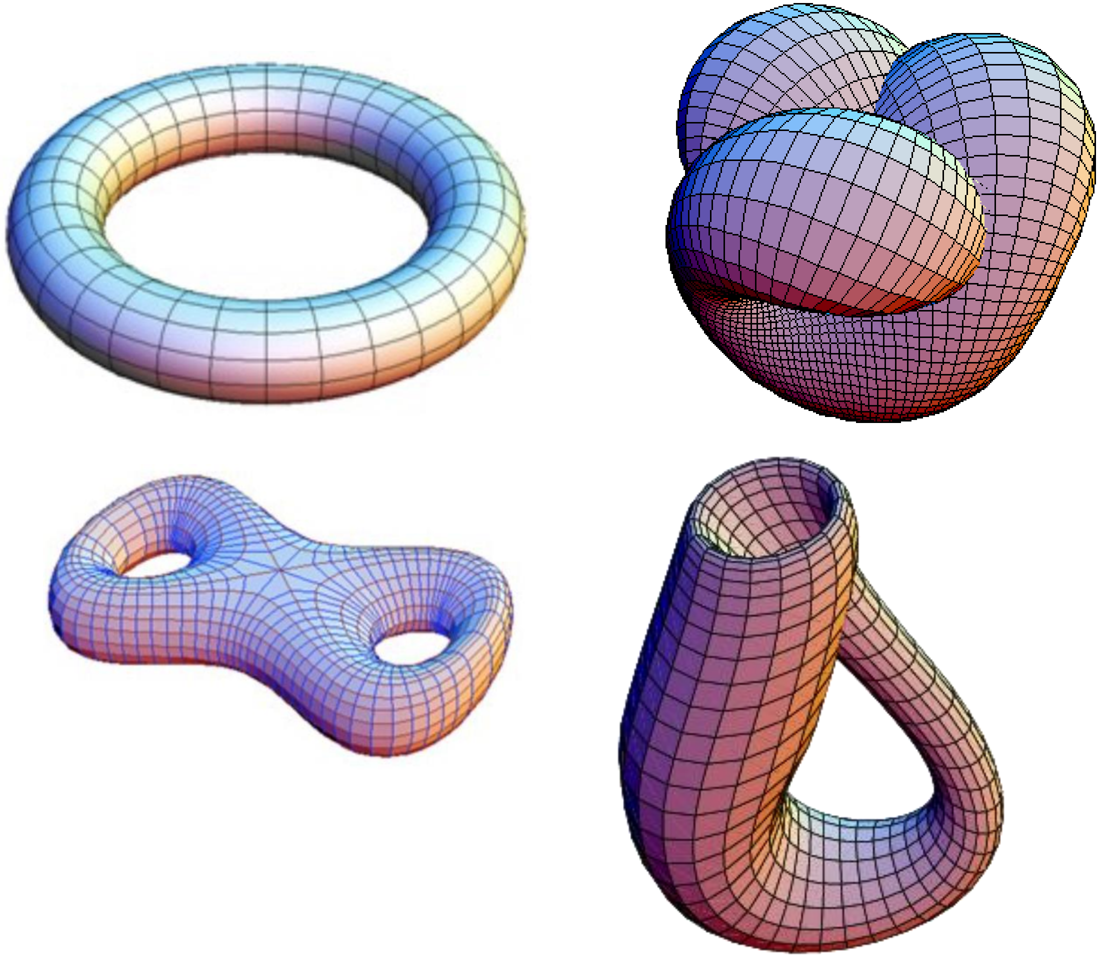
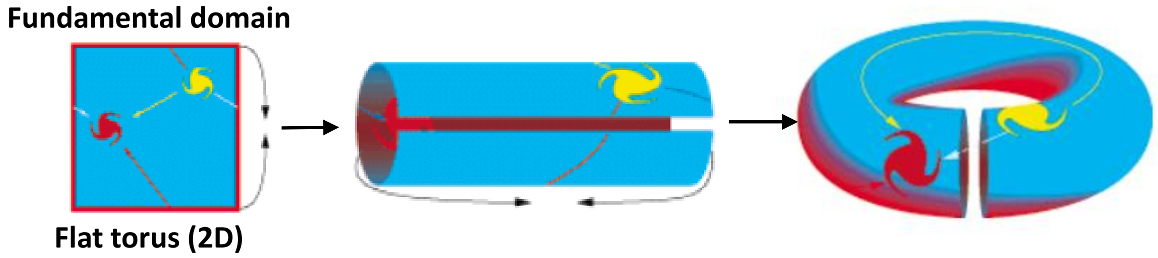


Figure 1: examples of manifolds with different topological properties.

Observational Signatures

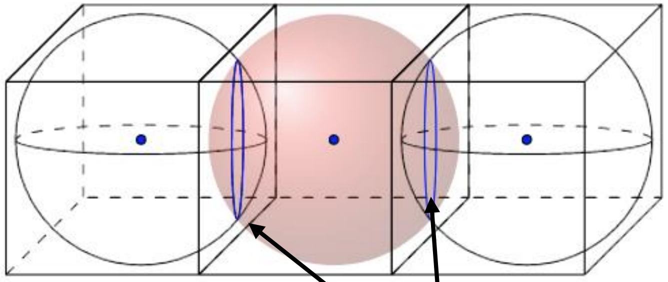


- Non-trivial topology leads to multiple observational effects:
  - ❖ ***Clone* images** of astronomical sources;
  - ❖ Observational signatures in the CMB (***circles in the sky***);
  - ❖ **Non-diagonal** correlations in harmonic space;
  - ❖ Others? E.g. observables in **polarization** data.

Image: Luminet (2015)

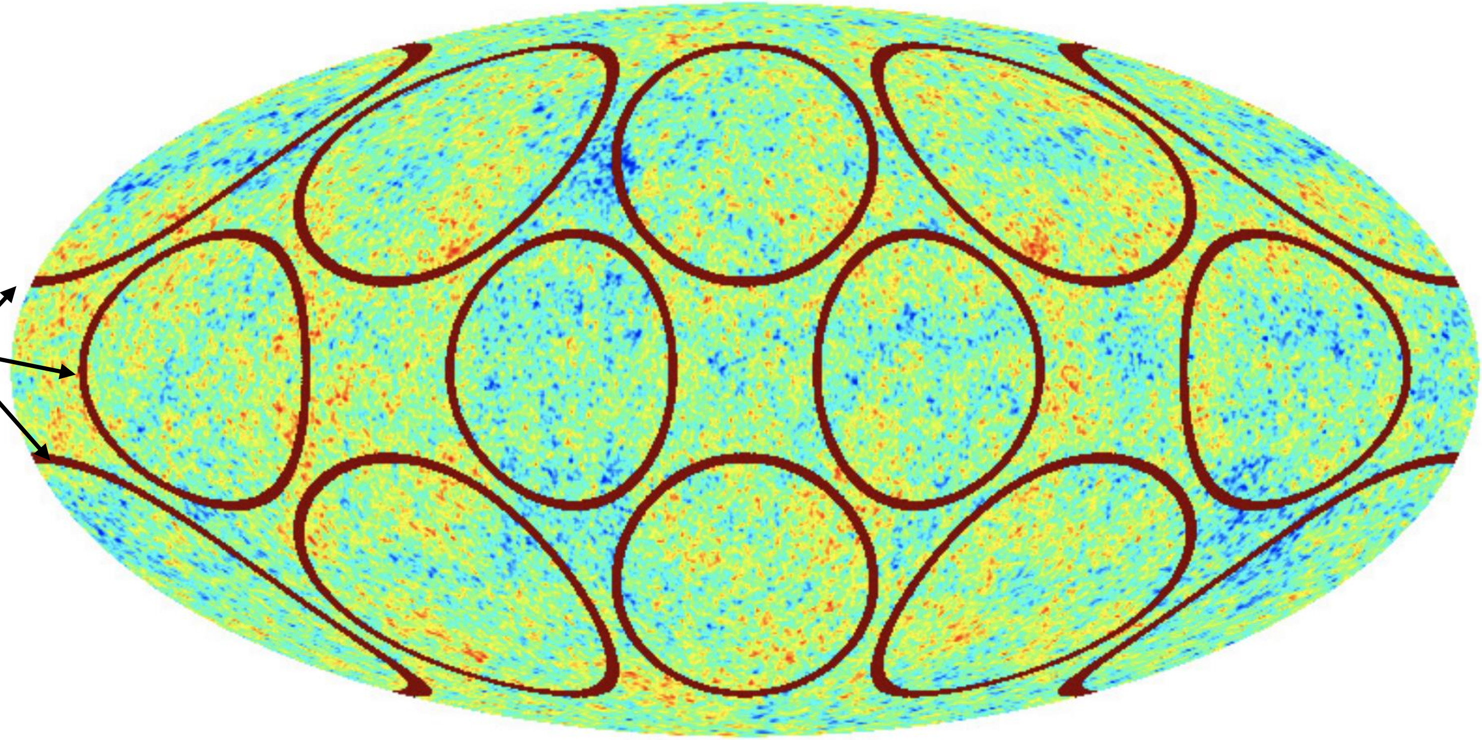
Figure 2: the mirror hall effect (image: **Rebecca Dale**).

Circles in The Sky



Intersection of the last scattering surface and the fundamental domain

Locations of circle pairs with matched CMB patterns



Other features encoded in the $a_{\ell m}$'s:

$$T(x) = \sum_{\ell,m} a_{\ell m} Y_{\ell m}(x)$$

$$\langle a_{\ell m} a^*_{\ell' m'} \rangle = C_\ell \delta_{\ell\ell'} \delta_{mm'}$$

Figure 3: Last scattering surface intersecting itself in a 3-torus Universe (**top right**). The dark circles show the locations of the matched-circle pairs in the CMB (**center**) (Planck Collaboration 2013).

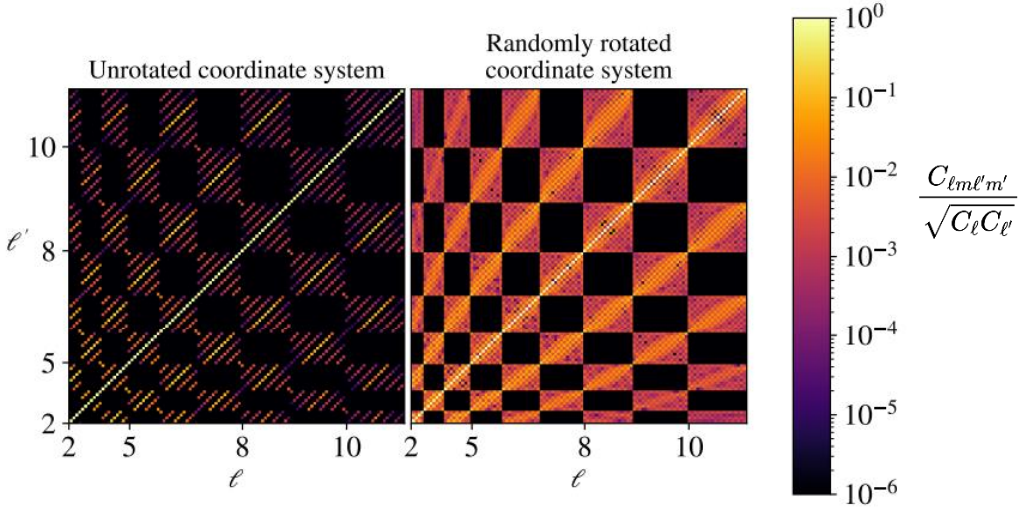
Detectability of Cosmic Topology

- **Non-diagonal elements** in the covariance matrix encode information about topology.
- **KL divergence** measures the detectability of these features:

$$D_{\mathrm{KL}}(p\|q) = \int \mathrm{d}\{a_{\ell m}\} p(\{a_{\ell m}\}) \ln\left[\frac{p(\{a_{\ell m}\})}{q(\{a_{\ell m}\})}\right]$$

$p(\{a_{\ell m}\})$ **- probability of non-trivial topology**

$q(\{a_{\ell m}\})$ **- probability of trivial topology**



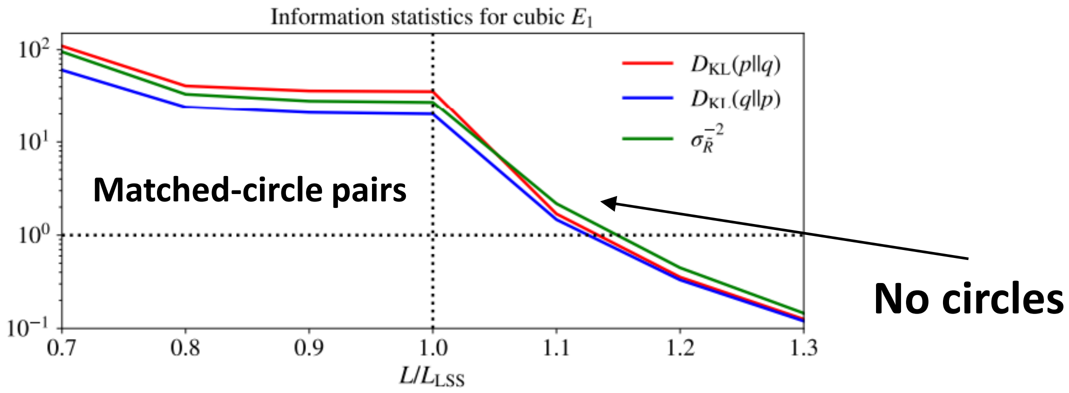


Figure 4: temperature covariance matrix (**top**). KL divergence in E_1 topology (**bottom**).

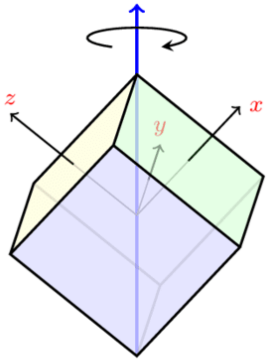
Detecting Topology with AI

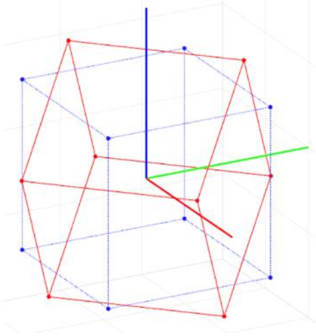
$a_{\ell m}$ realizations $\longrightarrow$ Algorithm $\longrightarrow$ Class 1 or Class 2 … Class N

$C_{\ell m \ell' m'} = a_{\ell m} a^*_{\ell' m'}$ realizations $\longrightarrow$ Algorithm $\longrightarrow$ Class 1 or Class 2 … Class N

- The goal: an algorithm to classify **harmonic space** realizations from different topologies.
- Start with a single topology: 3-torus (***E*$_1$**) of different sizes.
- Two datasets: **rotated and non-rotated**.
- Algorithms to try:
  - ❖ Random forests and XGBoost;
  - ❖ 1D convolutional neural networks;
  - ❖ 2D convolutional neural networks;
  - ❖ Complex neural networks.

Fundamental domain for a given topology





Rotating the coordinate system (multiplying by Wigner D-matrix)

4 classes: 40,000 realizations

- E1 with $L_x = L_y = L_z = 0.05 \times L_{LSS}$
- E1 with $L_x = L_y = L_z = 0.1 \times L_{LSS}$
- E1 with $L_x = L_y = L_z = 0.5 \times L_{LSS}$
- Trivial topology $L_x = L_y = L_z = L_\infty$

$$a^{E1}_{\ell m} = \frac{4\pi}{\sqrt{V_{E1}}} i^l \sum_{\vec{n}} \delta_{\vec{k}_{\vec{n}}} e^{-i\vec{k}_{\vec{n}}\cdot\vec{x}_0} Y^\star_{\ell m}(\hat{k}) \Delta_\ell(k)$$

The Dataset

Realizations of E_1 topology of 2 different sizes

Complex harmonic coefficients

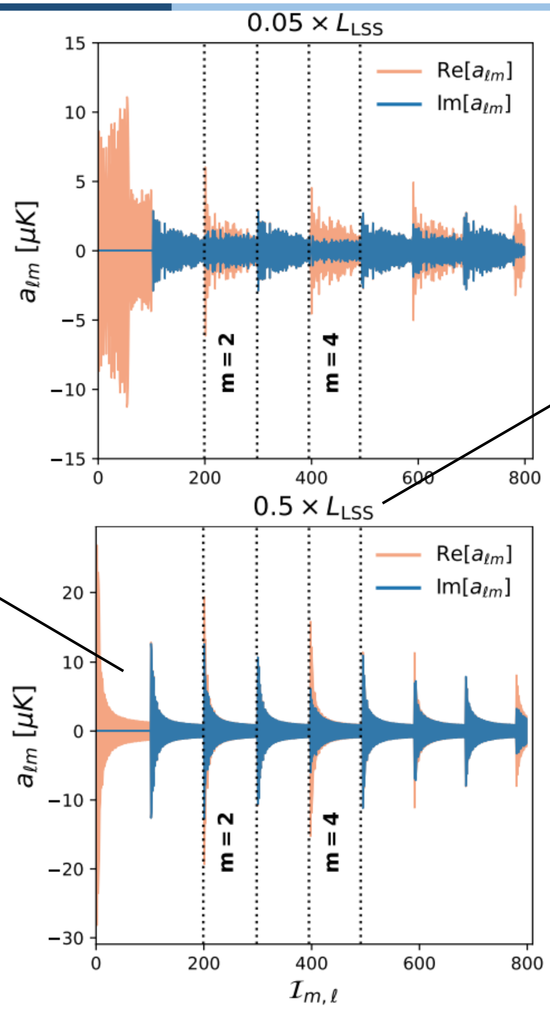


Figure 5: Real vs imaginary components of the $a_{\ell m}$ realizations.

$$C_{\ell m \ell' m'} = a_{\ell m} a^*_{\ell' m'}$$

Correlations between different angular scales

$$C_{\ell m \ell' m'} = \langle a_{\ell m} a^*_{\ell' m'} \rangle$$

Corresponding (m, m') values

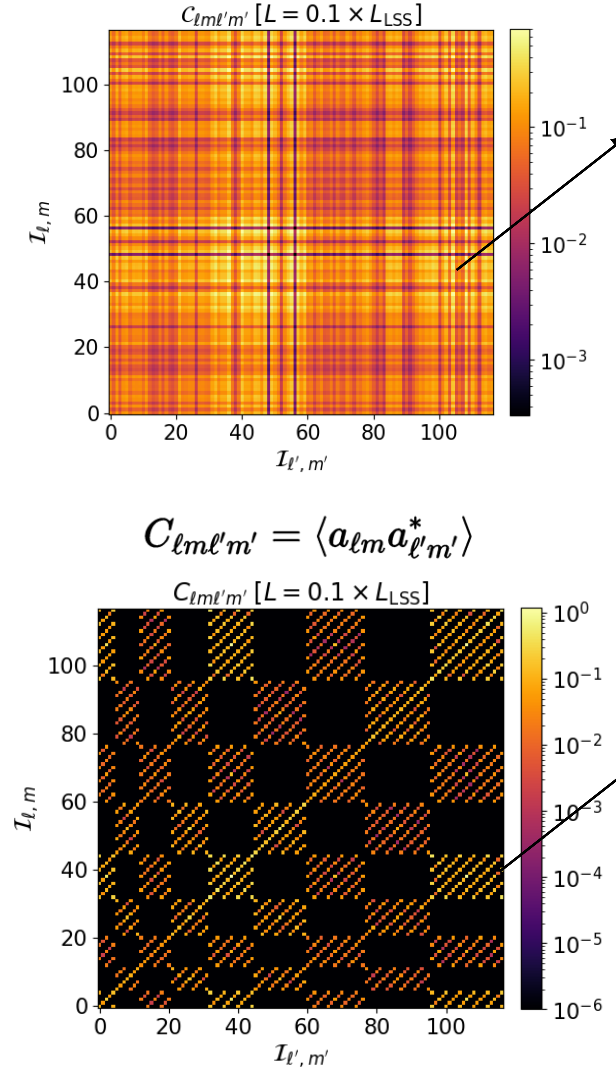
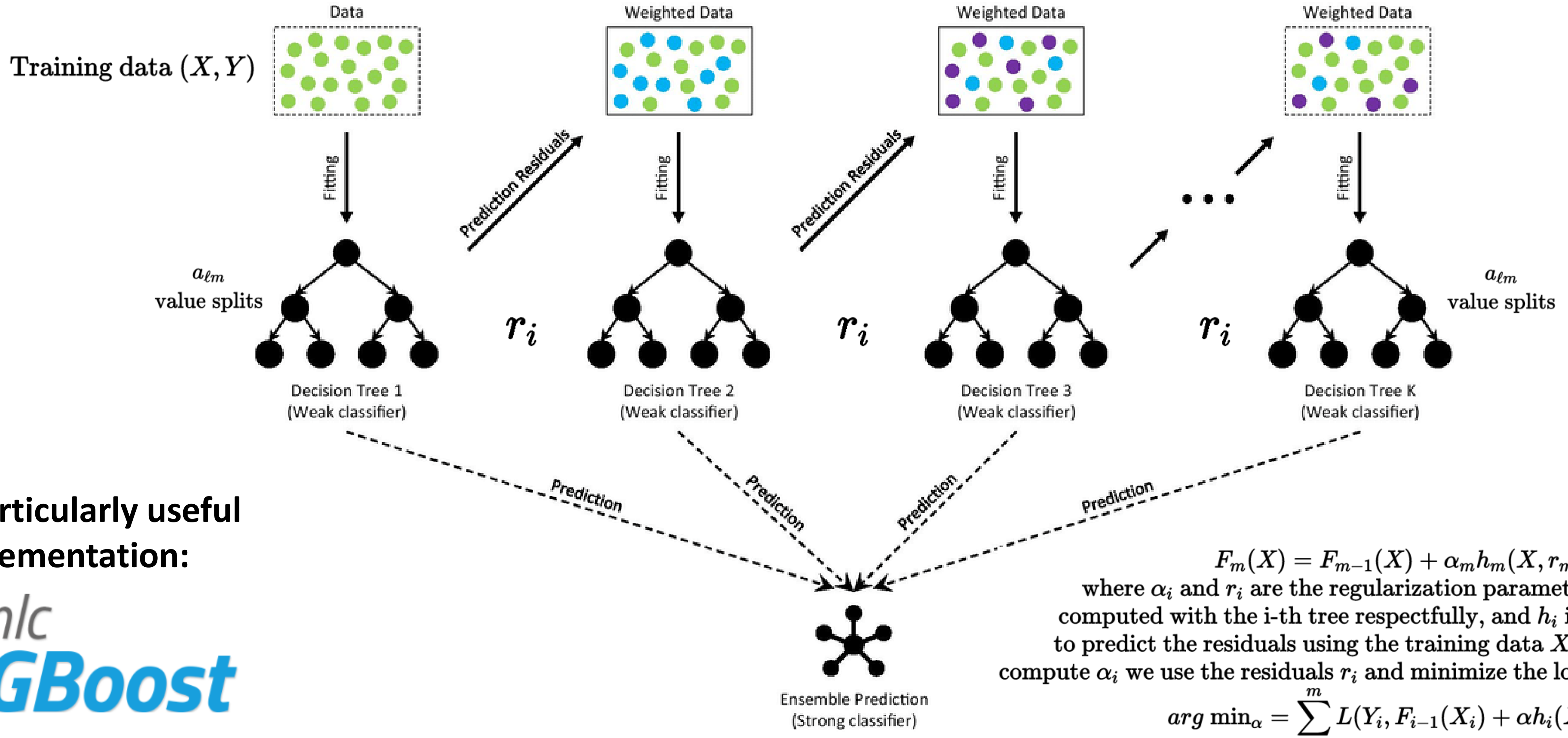


Figure 6: Harmonic space covariance matrices in E_1 topology.

The Algorithms: Extreme Gradient Boosting



$$F_m(X) = F_{m-1}(X) + \alpha_m h_m(X, r_{m-1}),$$

where α_i and r_i are the regularization parameters and residuals computed with the i-th tree respectfully, and h_i is a function trained to predict the residuals using the training data X for the i-th tree. To compute α_i we use the residuals r_i and minimize the loss function $L(Y, F(X))$:

$$arg\ \min_\alpha = \sum_{i=1}^{m} L(Y_i, F_{i-1}(X_i) + \alpha h_i(X_i, r_{i-1}))$$

A particularly useful implementation:

dmlc XGBoost

Figure 7: The gradient boosting classification algorithm (Deng et al. 2021).

The Algorithms: NN Architectures

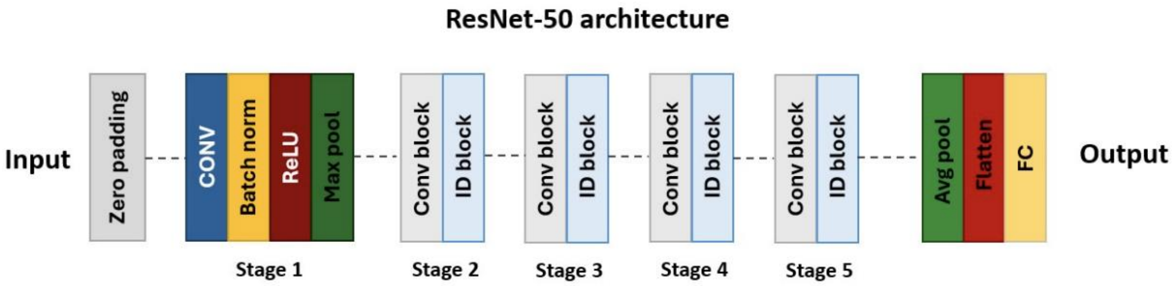


Figure 8: ResNet-50 architecture.

	Activation	Output shape	Parameters
Input map	-	(None, 2652, 32)	128
Conv1D	LReLU	(None, 2652, 64)	10304
Conv1D	LReLU	(None, 2652, 128)	57472
Conv1D	LReLU	(None, 2652, 256)	295 K
Max Pooling 1D	-	(None, 1326, 256)	-
Dropout	-	(None, 1326, 256)	-
Flatten	-	(None, 339456)	-
Dense	LReLU	(None, 512)	173 M
Dense	LReLU	(None, 256)	131 K
Dense	LReLU	(None, 128)	32 K
Output layer	Softmax	4	516
Total trainable parameters:			**174 M**

Table 1: The architecture of the 1D CNN used in this work.

	Activation	Output shape	Parameters
Input map	-	(207,207,1,32)	-
ComplexConv2D	cart_relu	(None, 205, 205, 32)	640
ComplexConv2D	cart_relu	(None, 203, 203, 32)	18 K
ComplexAvgPooling2D	-	(None, 101, 101, 32)	-
ComplexConv2D	cart_relu	(None, 99, 99, 64)	37 K
ComplexConv2D	cart_relu	(None, 98, 98, 64)	33 K
ComplexAvgPooling2D	-	(None, 49, 49, 64)	-
ComplexConv2D	cart_relu	(None, 47, 47, 128)	148 K
ComplexConv2D	cart_relu	(None, 47, 47, 64)	17 K
ComplexConv2D	cart_relu	(None, 45, 45, 128)	148K
ComplexAvgPooling2D	-	(None, 22, 22, 128)	-
ComplexConv2D	cart_relu	(None, 20, 20, 256)	590 K
ComplexConv2D	cart_relu	(None, 20, 20, 128)	66 K
ComplexConv2D	cart_relu	(None, 18, 18, 256)	590K
ComplexAvgPooling2D	-	(None, 9, 9, 256)	-
ComplexConv2D	cart_relu	(None, 7, 7, 512)	2.4 M
ComplexConv2D	cart_relu	(None, 7, 7, 256)	263 K
ComplexConv2D	cart_relu	(None, 5, 5, 512)	2.4 M
ComplexConv2D	cart_relu	(None, 5, 5, 256)	263 K
ComplexConv2D	cart_relu	(None, 3, 3, 512)	2.4 M
ComplexAvgPooling2D	-	(None, 1, 1, 512)	-
ComplexFlatten	-	(None, 512)	-
ComplexDense	cart_relu	(None, 64)	66 K
ComplexDense	softmax_real_with_abs	4	520
Total trainable parameters:			**590 K**

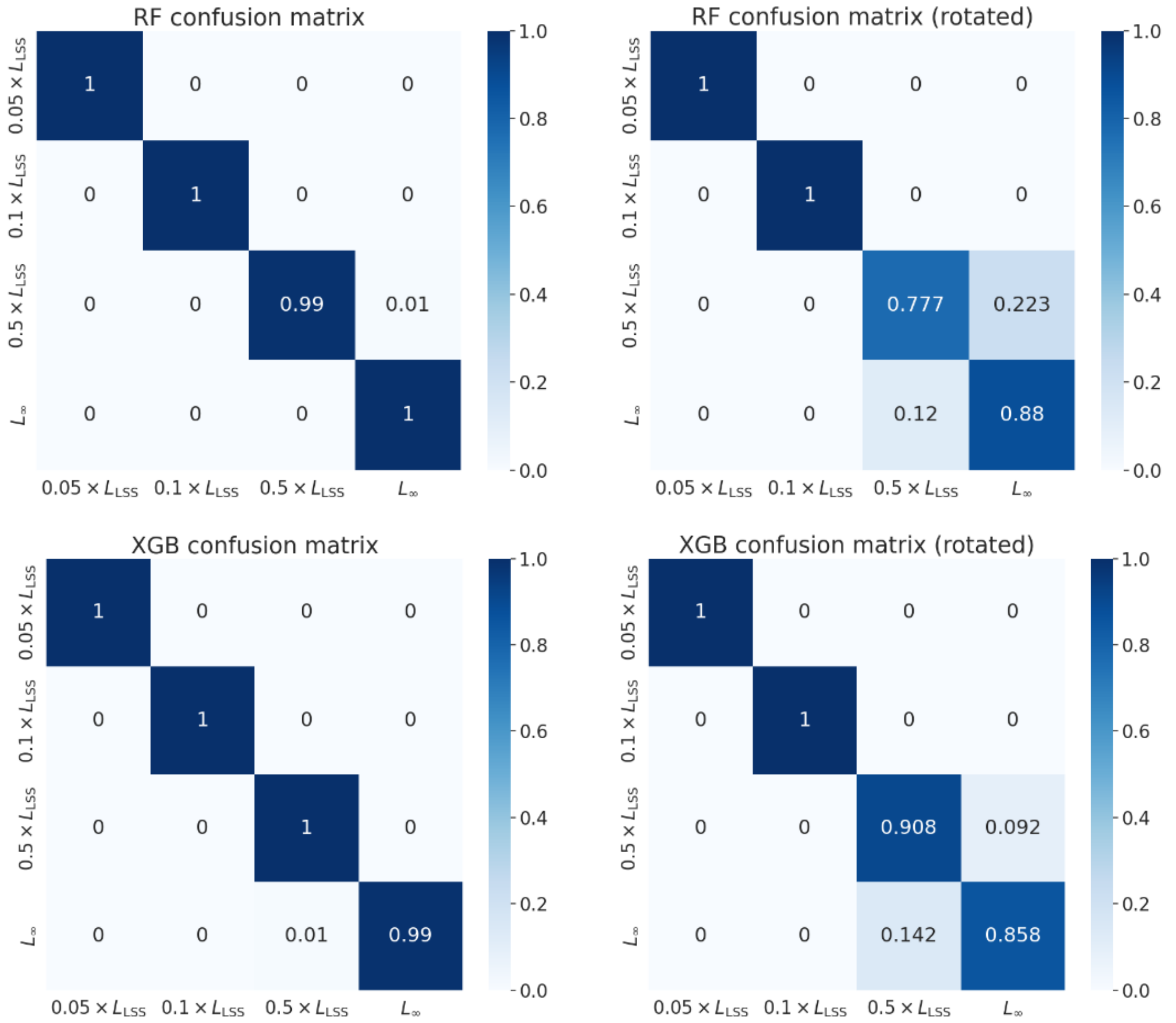
Table 2: 2D complex CVNN architecture.

Results: Random Forests + XGBoost

- Initial test: RF + XGBoost algorithms trained on the $a_{\ell m}$'s.
- Results depend on the coordinate system **orientation**/rotations.

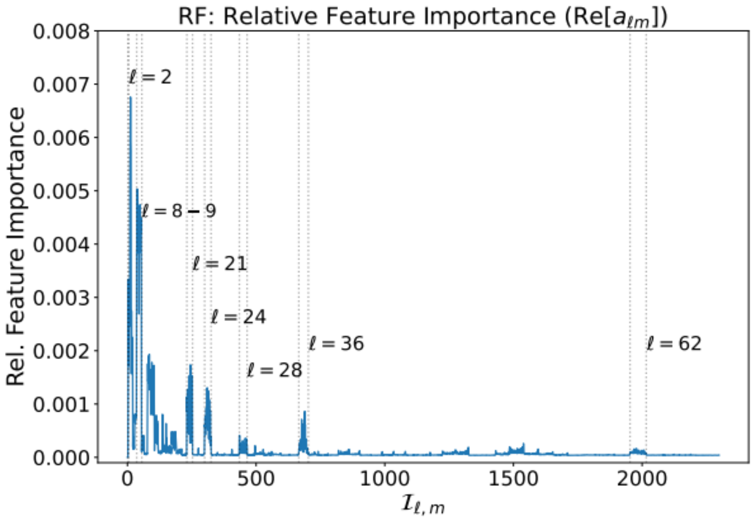
Algorithm:	Random forests	XGBoost
Training (unrotated):	100	100
Training (rotated):	**100**	**100**
Test (unrotated):	99.8	99.8
Test (rotated):	**91.4**	**94.2**

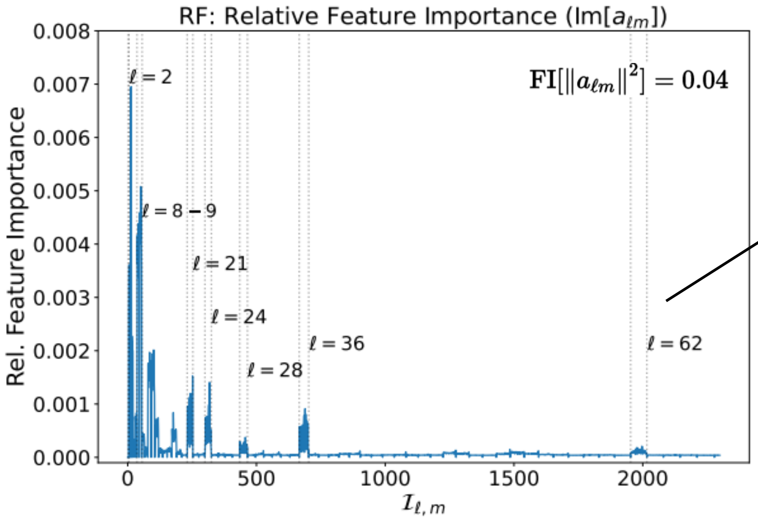
Table 3: summary of the decision tree-based algorithm classification results.



Results: Feature Importance Analysis

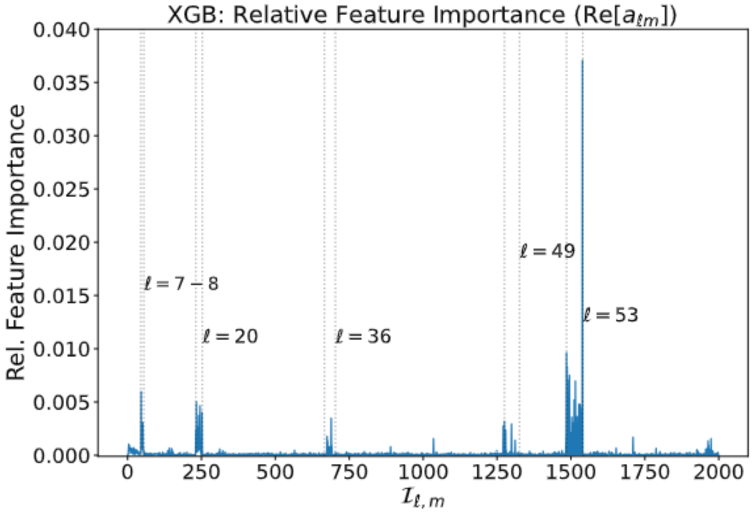
Random forests:

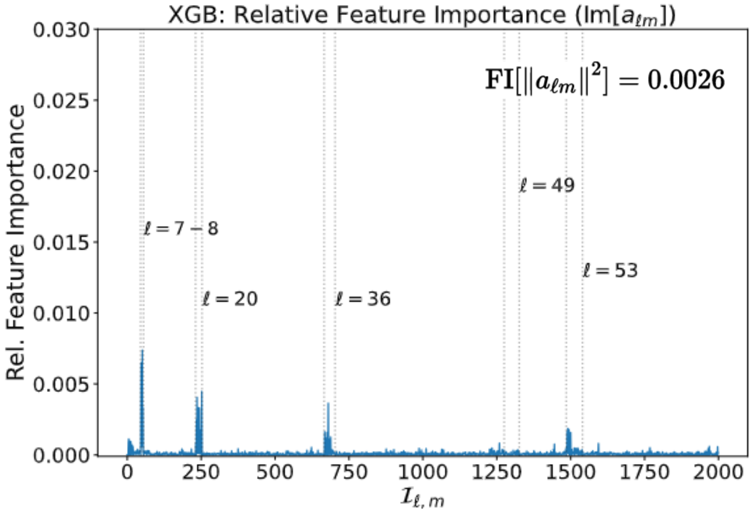




most important $a_{\ell m}$'s
$\rightarrow$ important angular scales
in the CMB

XGBoost classifier:





Results: NN Classification

- Neural network results show a similar trend.
- Results depend on the coordinate system orientation/Wigner **rotations**.
- The 2D Results are slightly worse (for the randomly-rotated data).
- Complex NNs do not perform better.

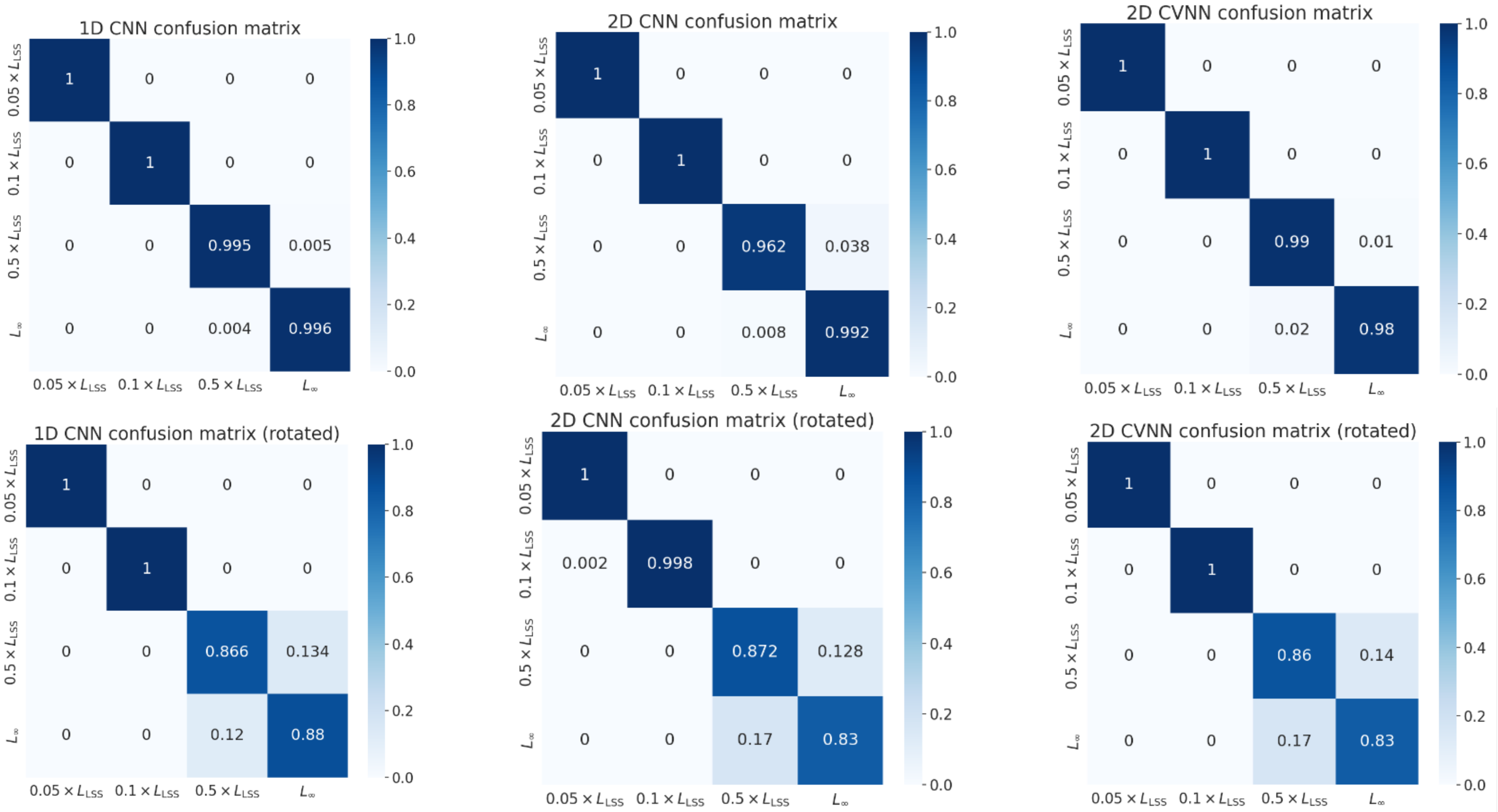
High accuracies for all non-rotated topology classes

Algorithm:	1D CNN	2D CNN (ResNet)	2D Complex CNN (CVNN)
Training (unrotated):	**100**	**99.9**	**100**
Training (rotated):	**98.8**	**98.4**	**100**
Test (unrotated):	**99.8**	**99.4**	**99.3**
Test (rotated):	**93.7**	**92.5**	**92.2**

Table 4: summary of the neural network algorithm classification results.

Significant reduction in accuracy when trained on rotated data

Results: NN Classification



Results: $L \approx L_{LSS}$

- Next challenge: classify realizations with $L > L_{\text{LSS}}$.
- We expect this to be more challenging (**smaller KL divergence, no circles**).
- Our techniques work well on non-rotated data.
- Key challenge: **classifying randomly rotated data**.

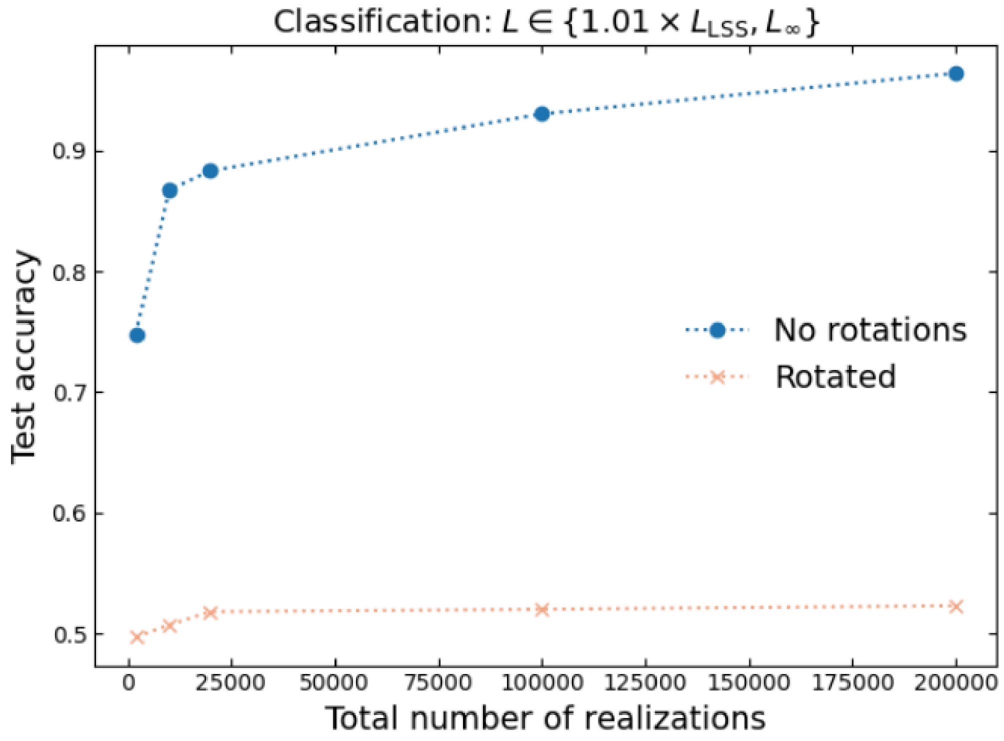
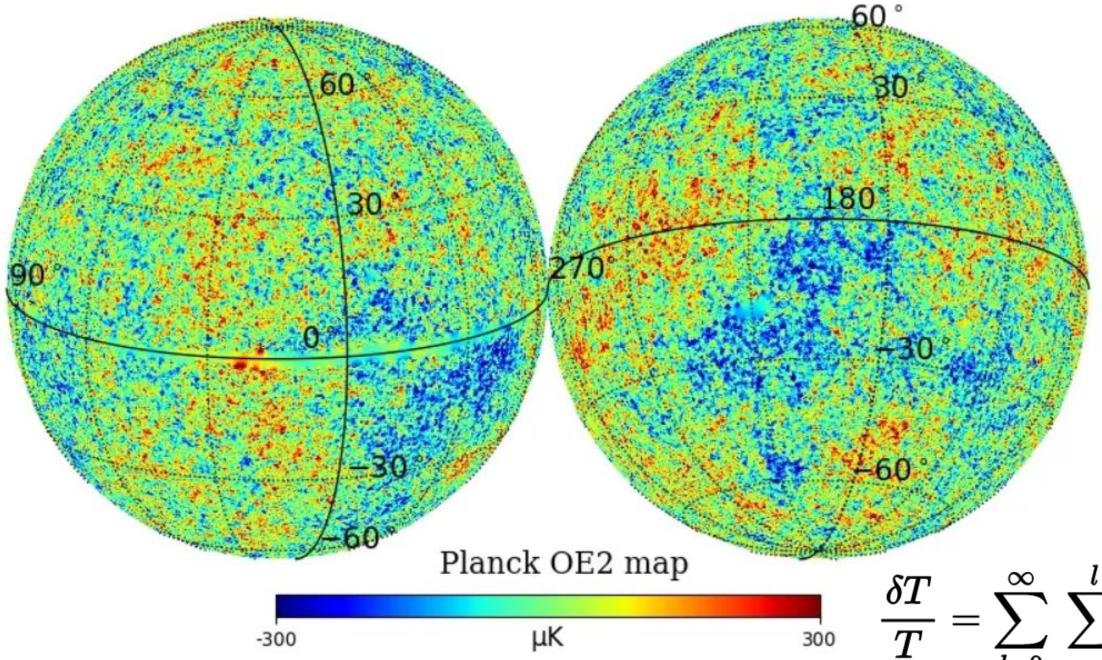


Figure 9: Classification results for realizations with $L \approx L_{\text{LSS}}$.

Rotation Problem: The Multipole Vector Formalism

CMB in spherical and harmonic space



$$\frac{\delta T}{T} = \sum_{l=0}^{\infty} \sum_{m=-l}^{l} a_{lm} Y_{lm}$$

Figure 10: CMB temperature anisotropies as a field on a sphere (Planck data).

Features in the MV data characteristic to non-trivial topology realizations

CMB in multipole vector formalism

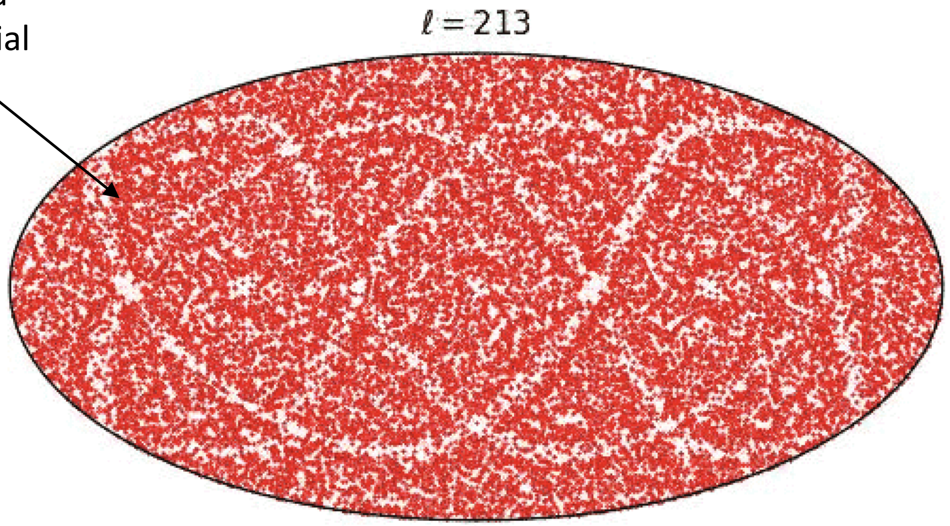


Figure 11: CMB temperature data represented as ℓ unit vectors on a sphere (i.e., *multipole vectors*).

Harmonic space rotations: $C'^{E_i}_{\ell m \ell' m'} = \sum_{\bar{m}=-\ell}^{\ell} \sum_{\bar{m}'=-\ell'}^{\ell'} D^{\ell}_{m\bar{m}}(\theta_0, \phi_0) D^{*\ell'}_{m'\bar{m}'}(\theta_0, \phi_0) C^{E_i}_{\ell \bar{m} \ell' \bar{m}'}$

MV rotations: rigid rotations (by 3 Euler angles)

Rotation Problem: MVs + E-mode Polarization Data

- The *signal* of non-trivial topology is stronger in **E-mode polarization data**.
- In E_1 topology, multipole vectors align along the **coordinate axes**.
- This characteristic feature can be used to orient randomly rotated multipole vectors.
- Approach:
  - ❖ Find the cross product of $\ell = 2$ MVs (Frechet vector). Find the rotation that orients it along the **z-axis**;
  - ❖ Use the obtained rotation to orient all the MVs;
  - ❖ Find the rotation that orients the $\ell = 2$ MVs along the **x-axis**;
  - ❖ Further rotate all the MVs using the obtained rotation.

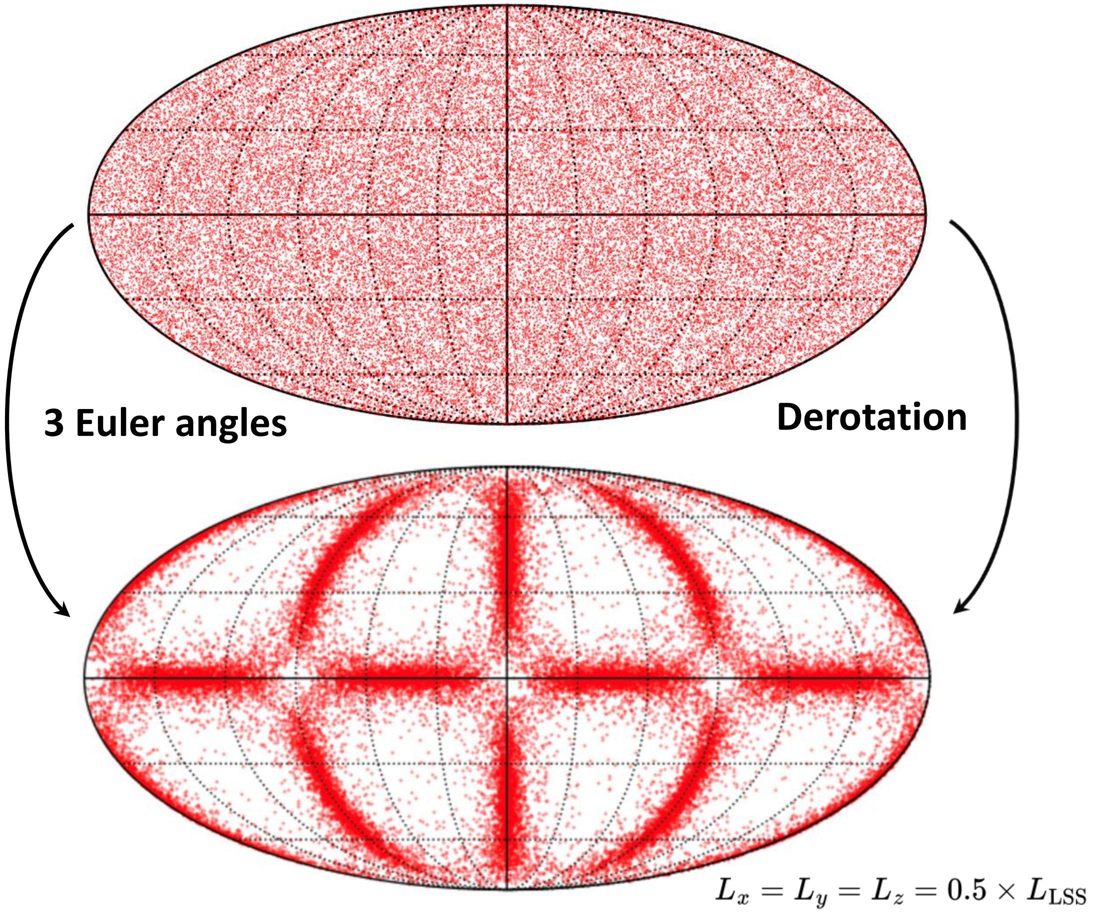
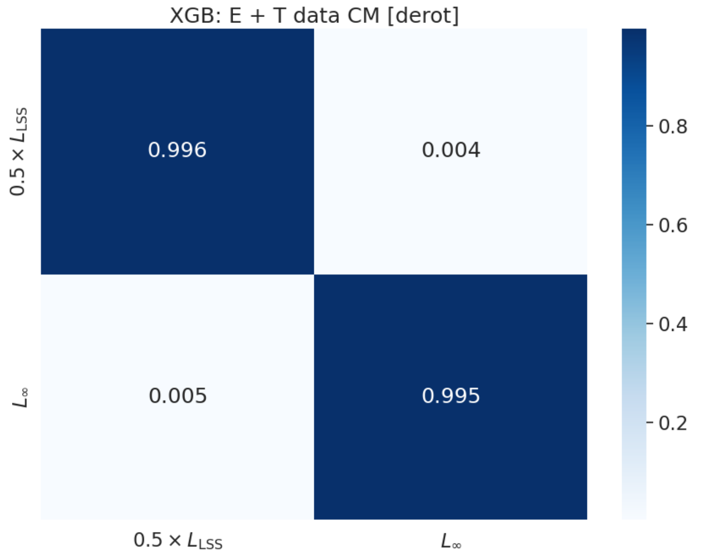


Figure 12: plotting multiple MV realizations for E-mode polarization data ($\ell = 2$). **Top:** randomly rotated realizations. **Bottom:** non-rotated realizations.

Rotation Problem: Current Results

- Cubic E_1 with $L \approx 0.5 \times L_{LSS}$: derotation procedure works well.
- Larger E_1 realizations: the accuracy **drops quickly**.
- The procedure does not work as effectively for non-cubic realizations.
- Need to test the procedure for other topology classes (E_3, E_4, etc.).



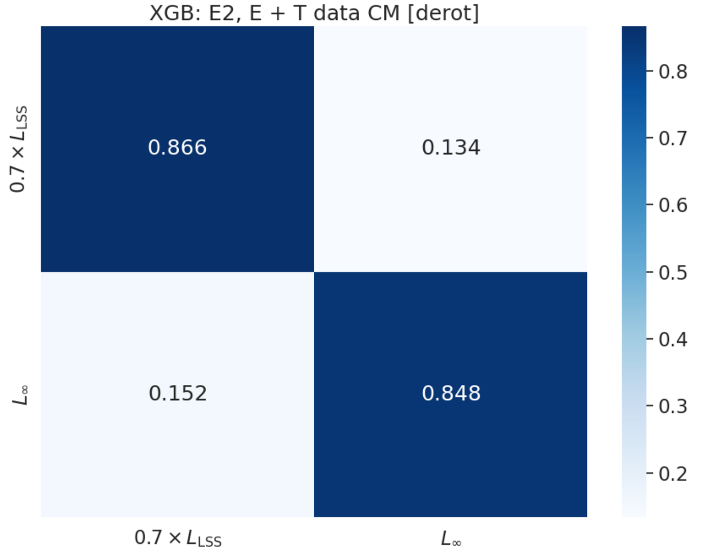


Figure 13: Results for XGBoost trained on the *derotated* dataset. **Top**: $L = 0.5 \times L_{\mathrm{LSS}}$ vs. covering space. **Bottom**: $L = 0.7 \times L_{\mathrm{LSS}}$ vs. covering space.

Using Map Data Directly

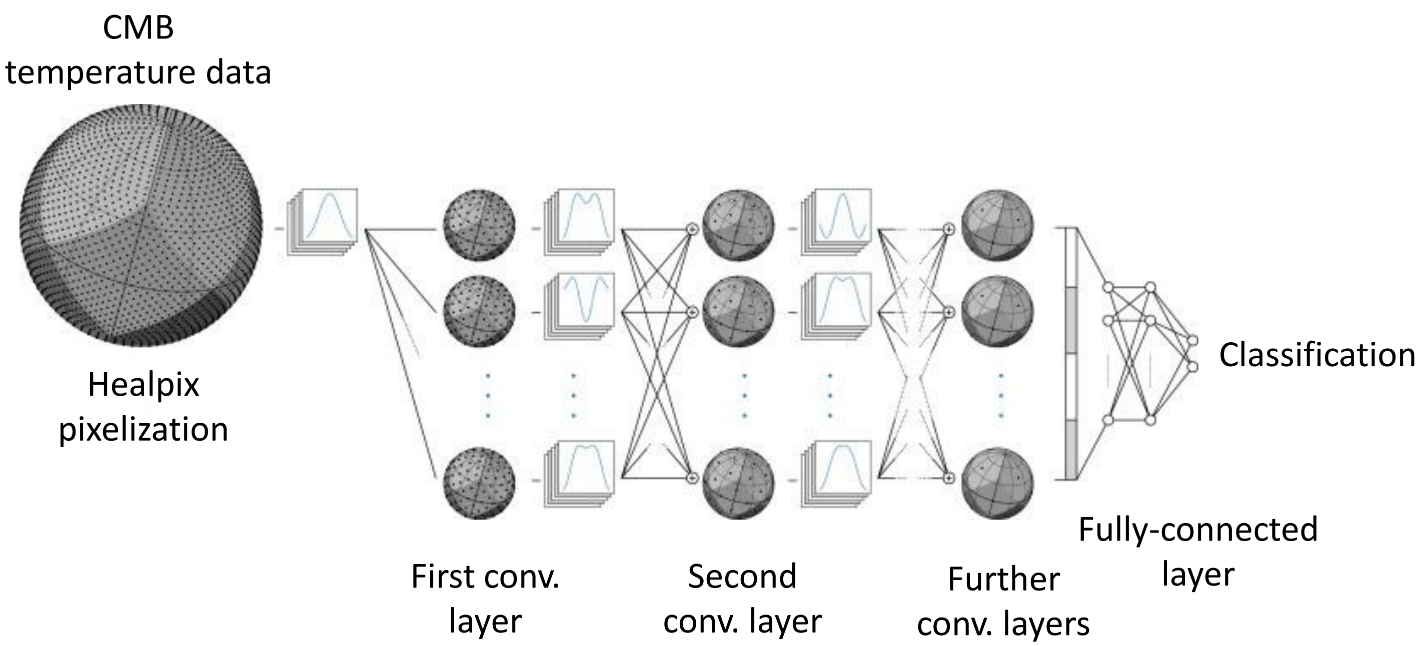


Figure 14: training a convolutional neural network on spherical data directly using DeepSphere (**Perraudin et al. 2019**).

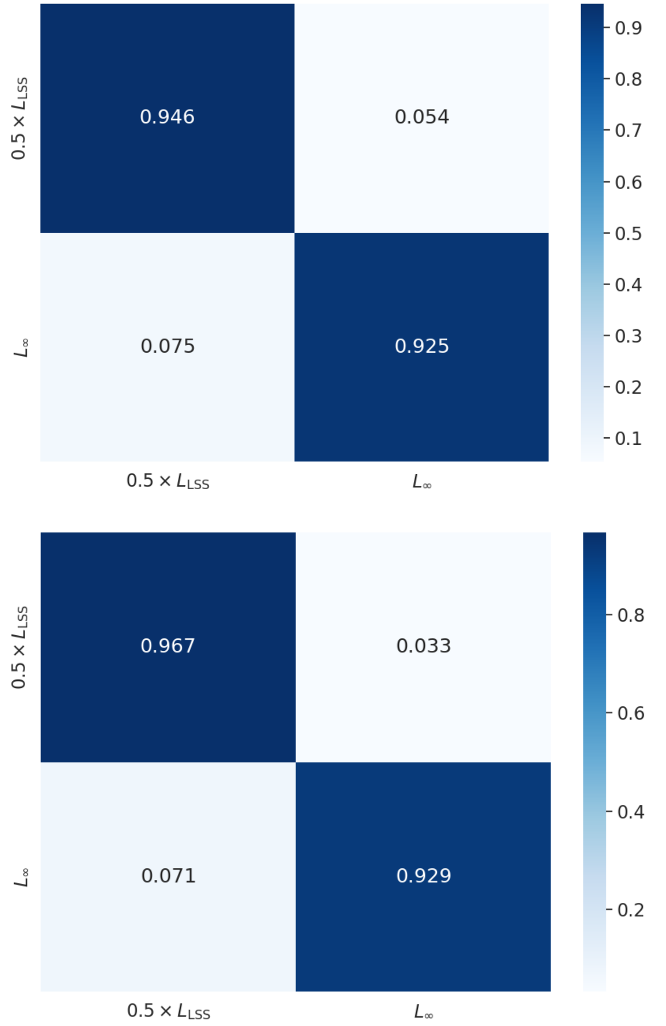


Figure 15:DeepSphere results (E_1 vs. covering space classification) using E-mode polarization data. **Top:** results for the non-rotated dataset. **Bottom:** results for the rotated dataset.

Conclusions and Future Work

- ML approaches are effective for classifying **small** E_1 realizations;
- For larger realizations, results depend on the orientation of the coordinate system;
- Dealing with rotations: **a key challenge**;
- Possible future approaches:
  - ❖ Rotation-equivariant neural networks;
  - ❖ Lorentz-equivariant neural networks + MV data;
  - ❖ Point cloud neural network approaches.

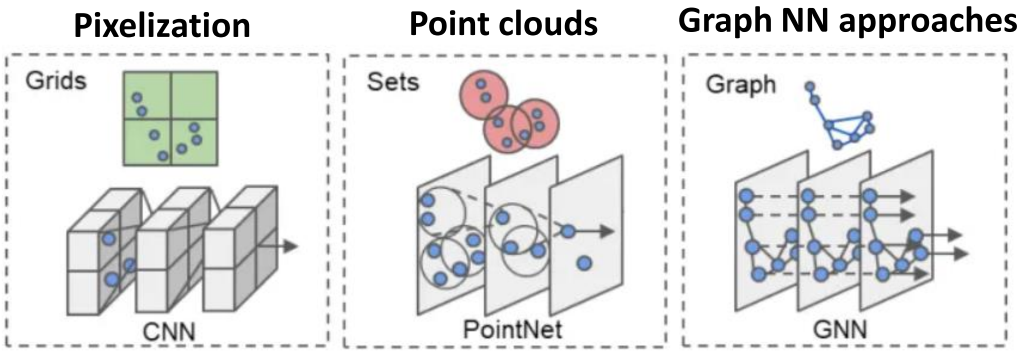


Figure 16: Different ML approaches that work with multipole vector data (**Shi and Rajkumar, 2020**).

andrius-tamosiunas.com

Paper: arXiv:2404.01236