

The Quest for Unification

- Intersecting Mathematics, Physics and AI -

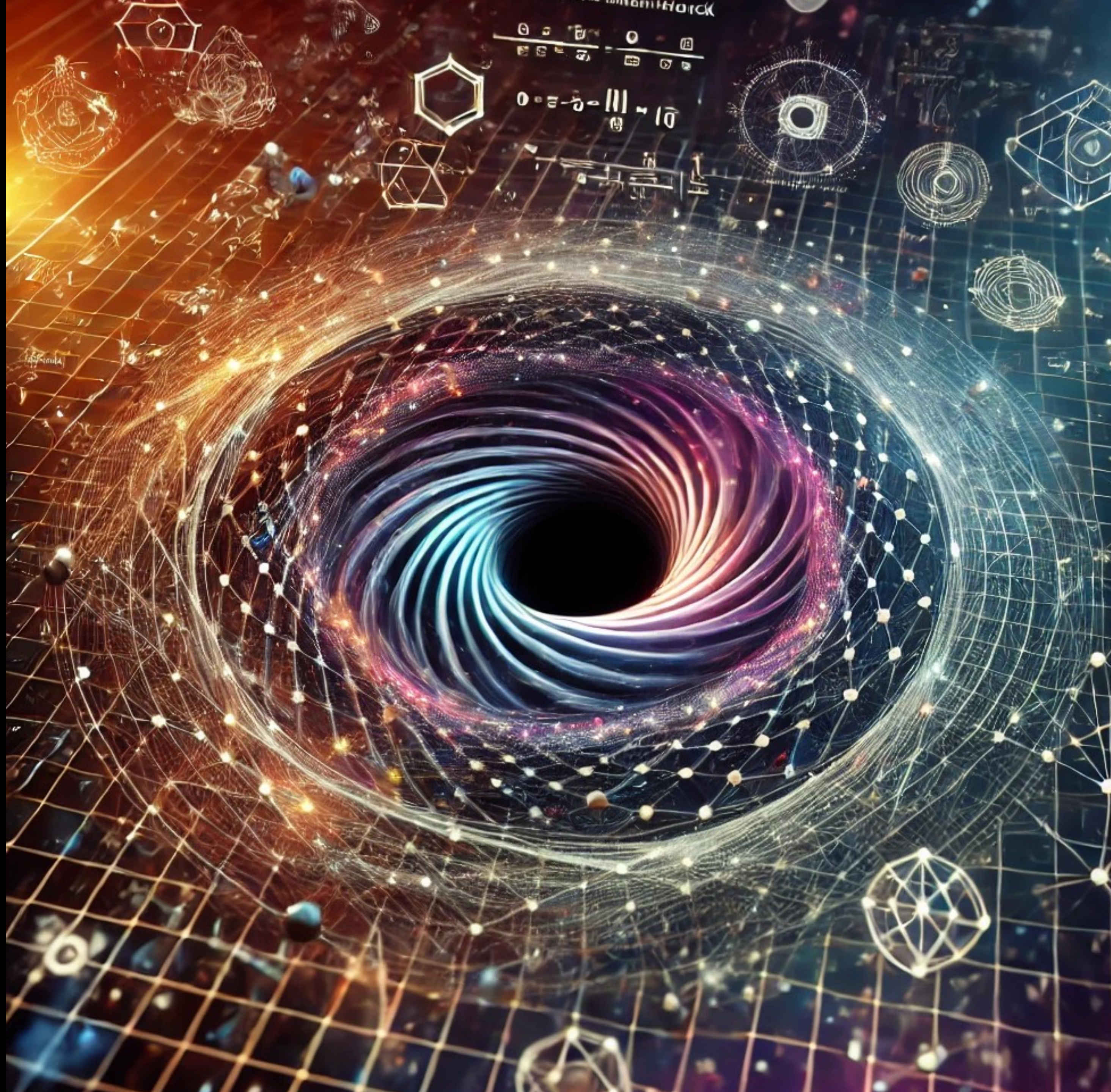
Daniel Persson

Department of Mathematical Sciences
Chalmers University of Technology
University of Gothenburg

WASP | WALLENBERG AI,
AUTONOMOUS SYSTEMS
AND SOFTWARE PROGRAM

Colloquium

Karlstad University
December 17, 2025



The Quest for Unification

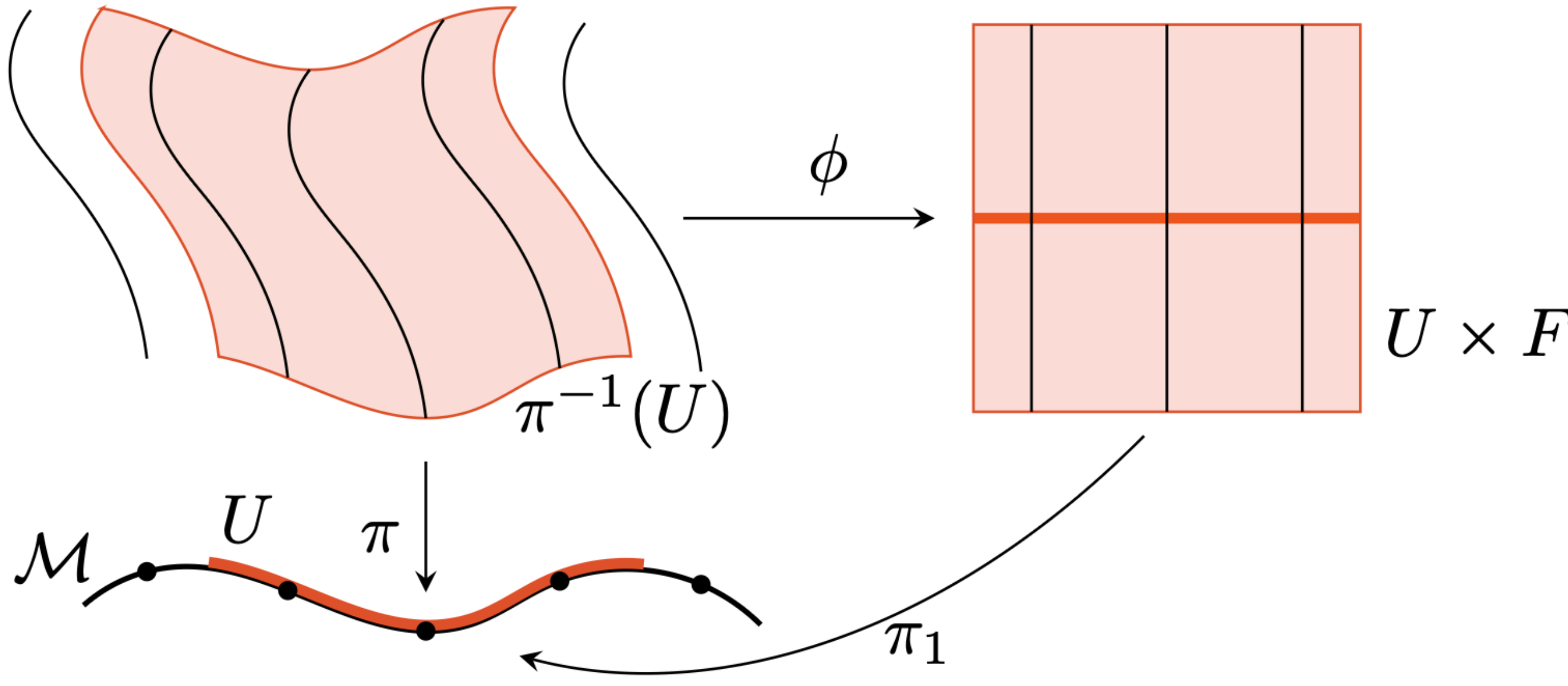
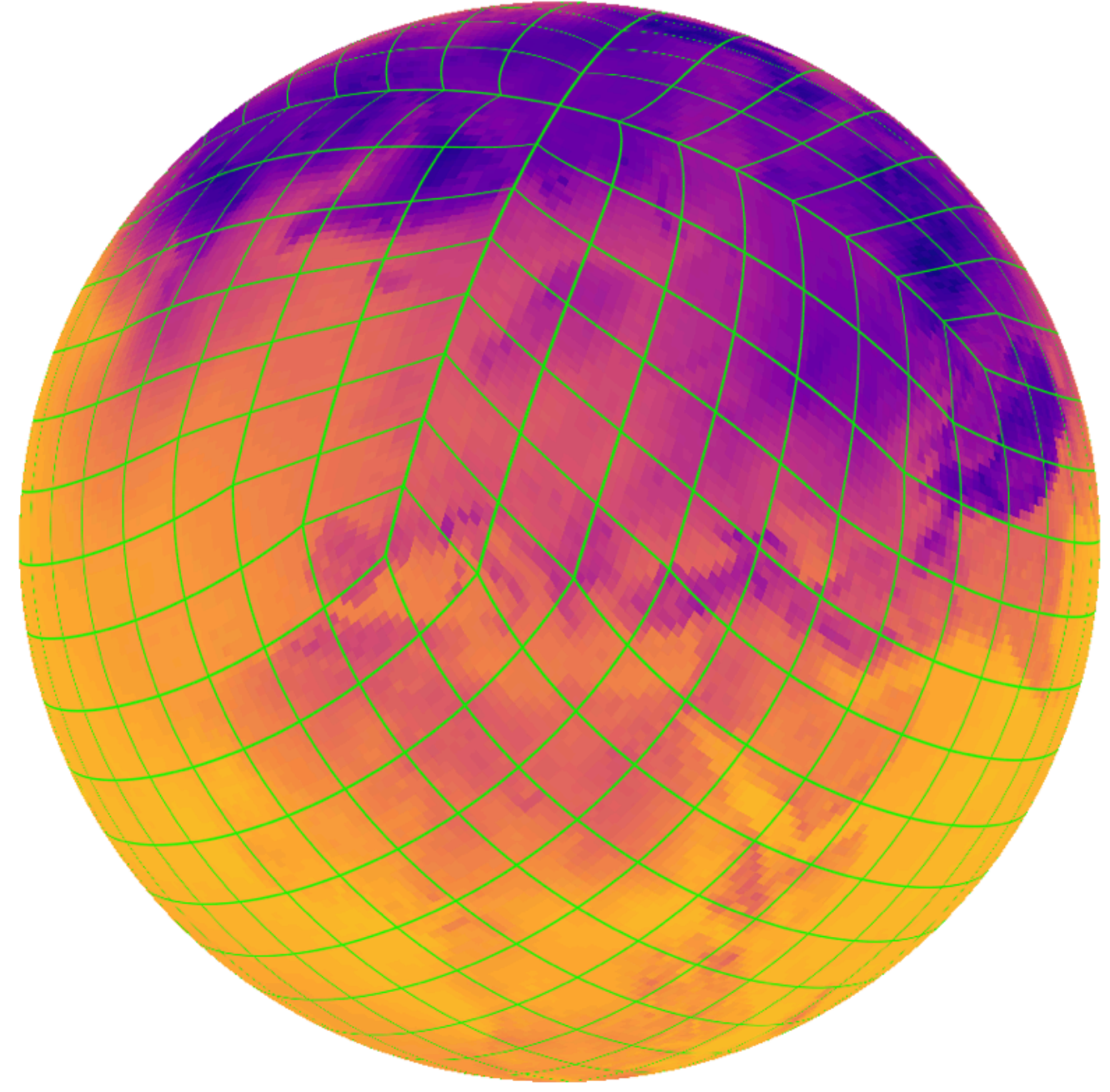
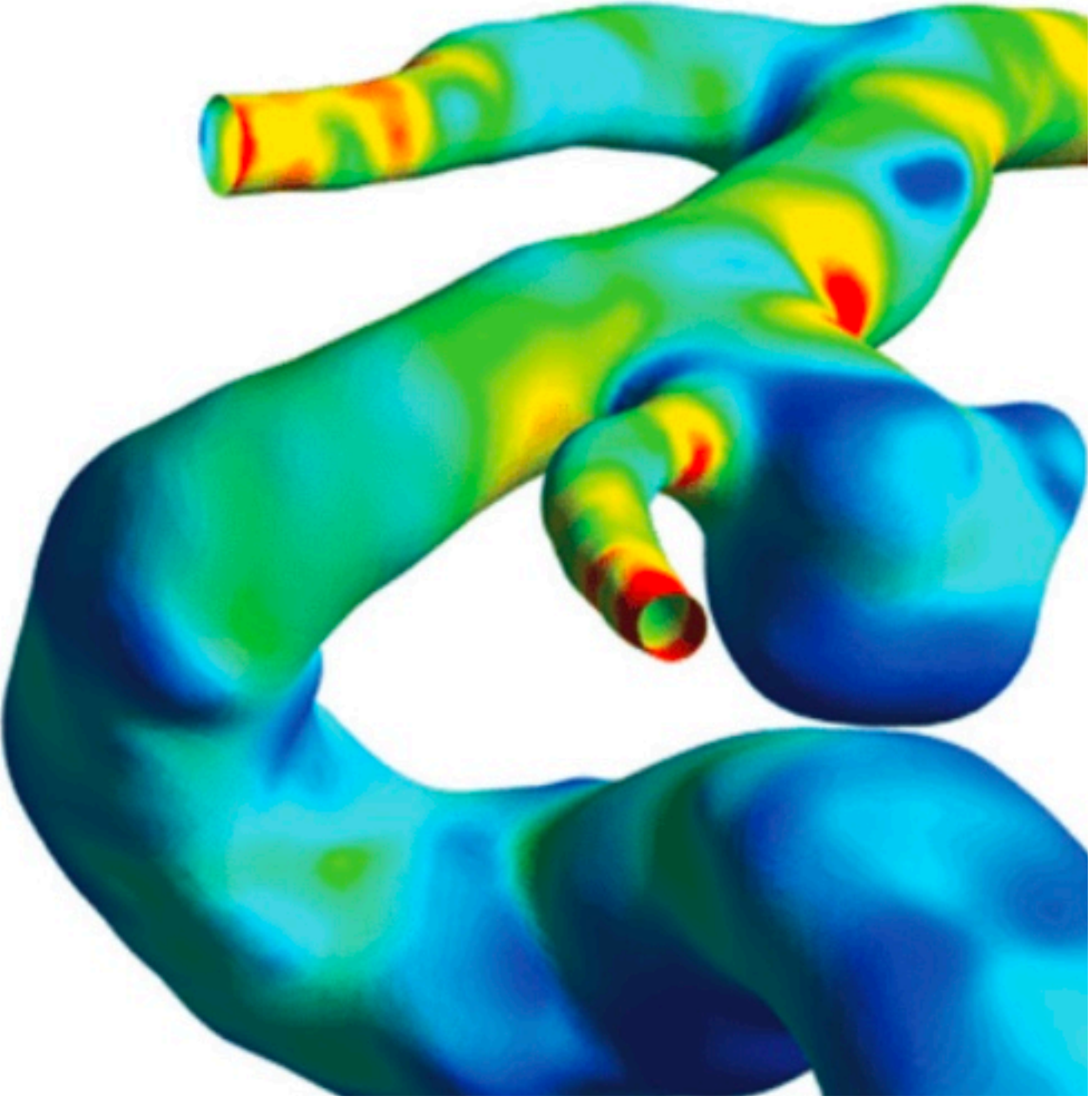
- Intersecting Mathematics, Physics and AI -

Daniel Persson

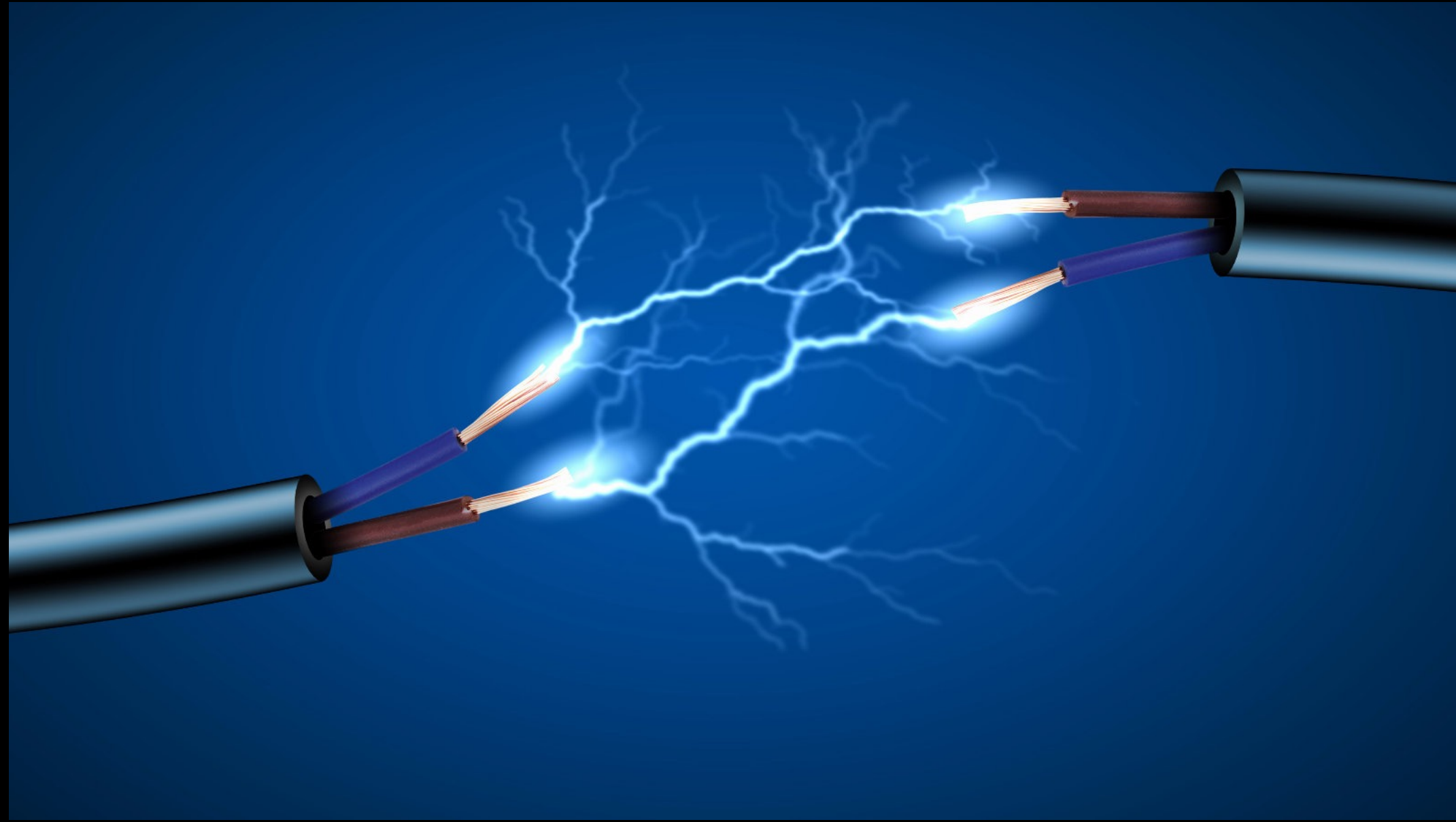
Department of Mathematical Sciences
Chalmers University of Technology
University of Gothenburg



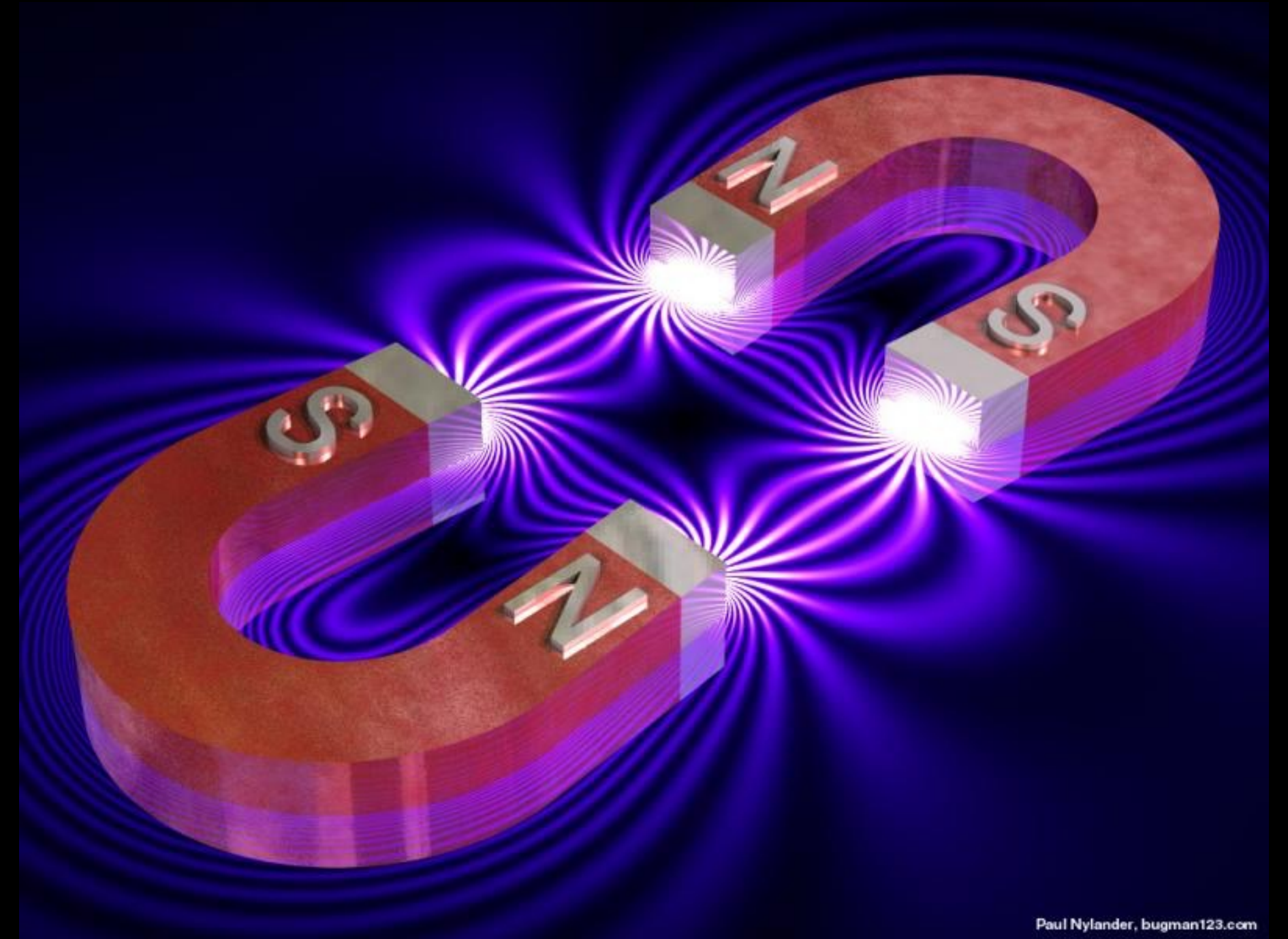
Karlstad University
December 17, 2025



Unification



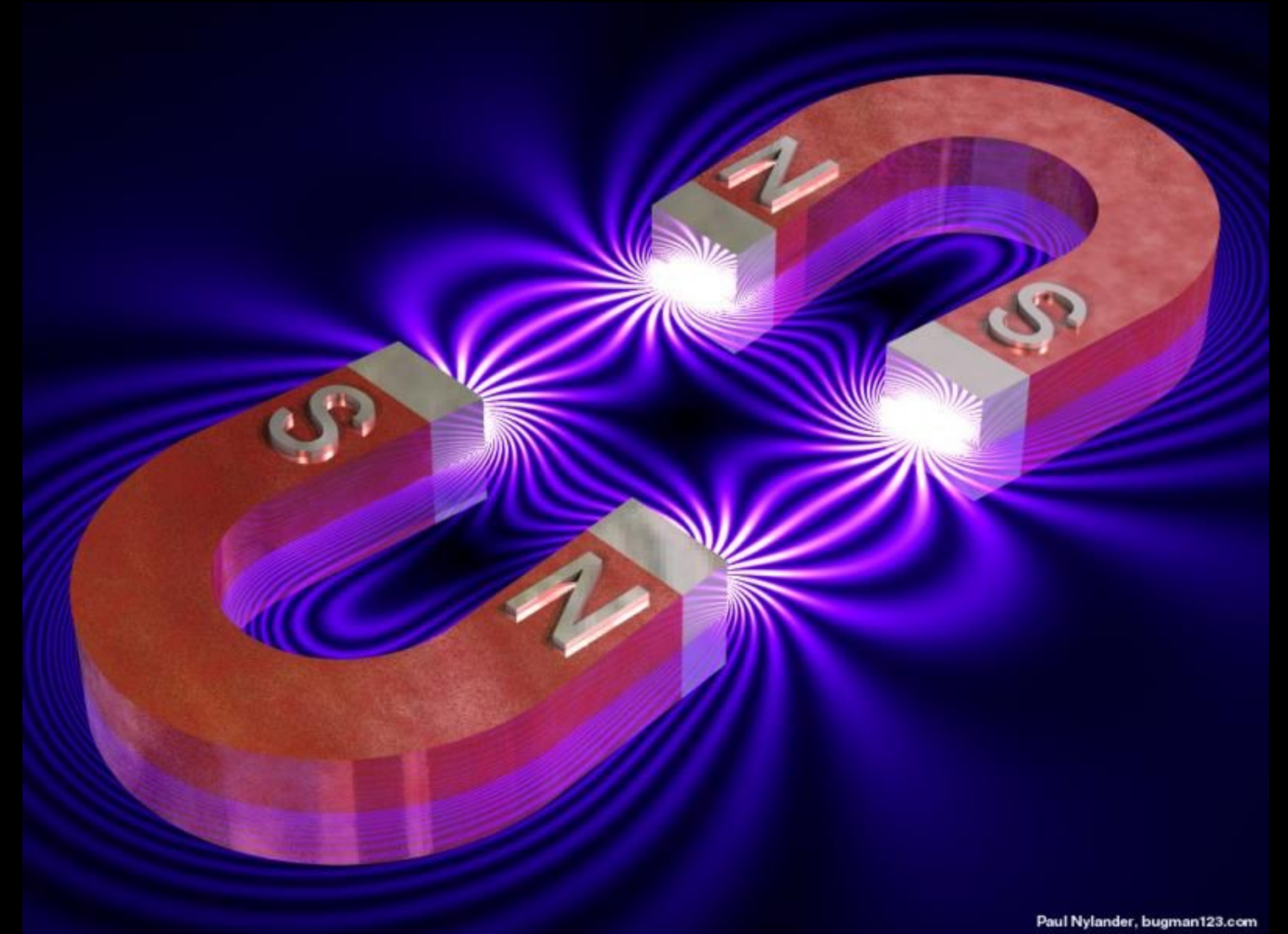
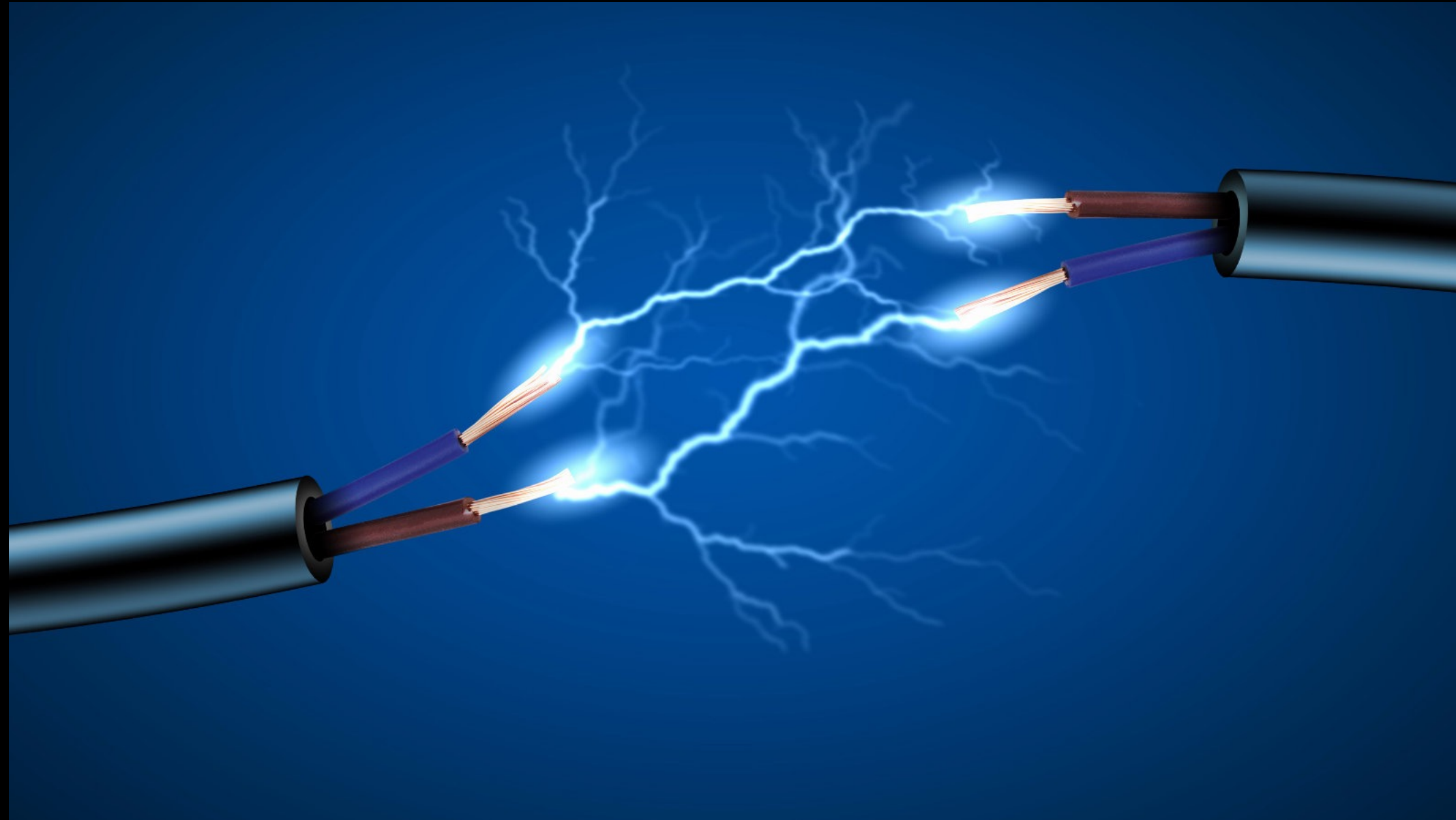
electricity



magnetism

Paul Nylander, bugman123.com

Unification

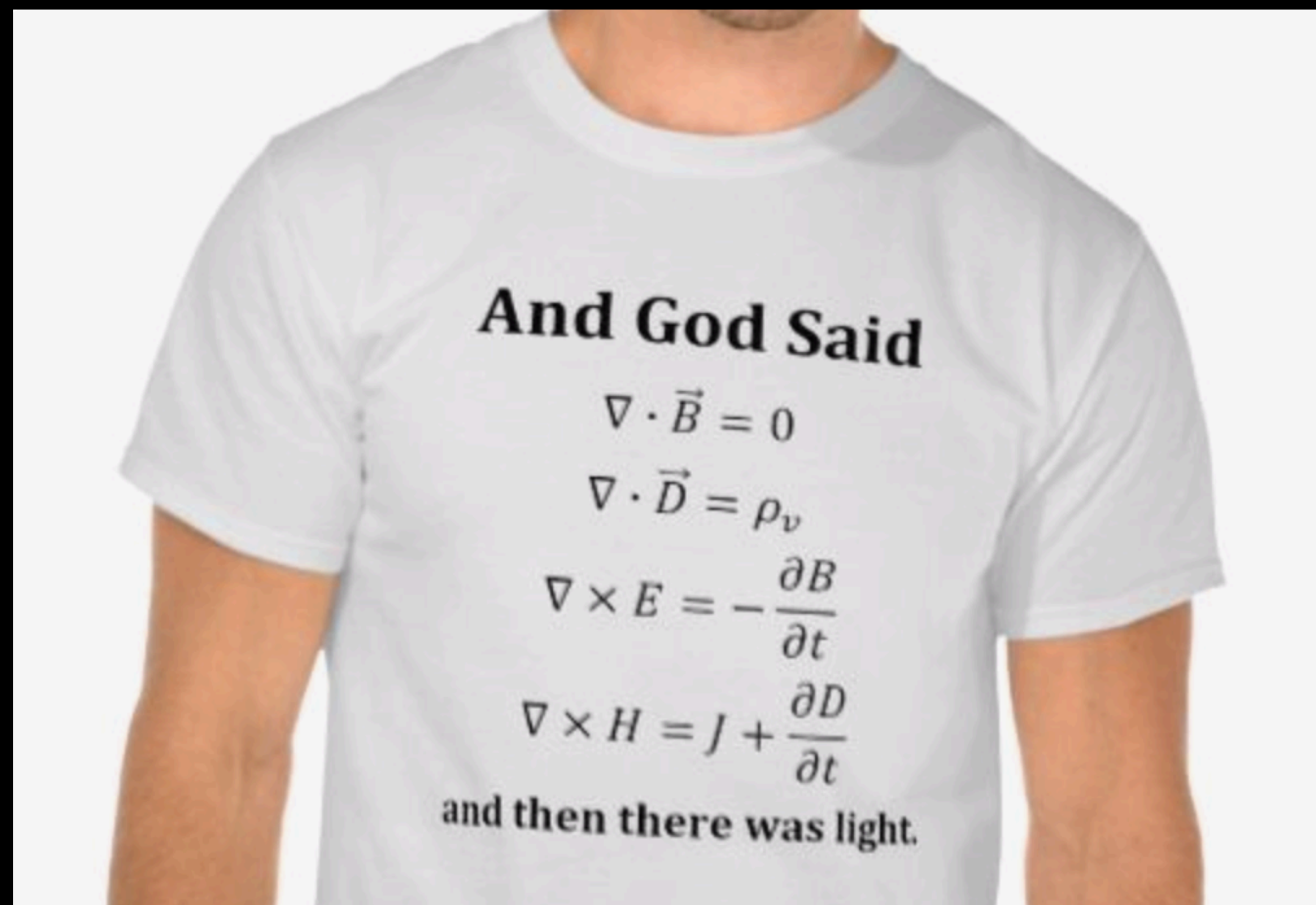
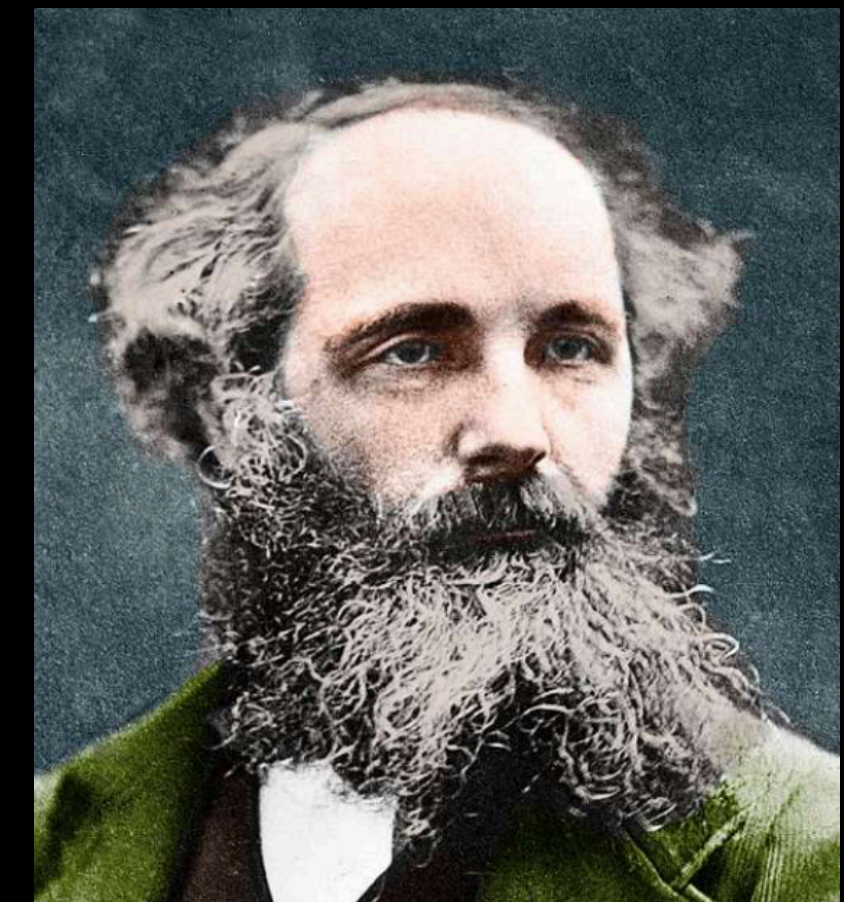


Paul Nylander, bugman123.com

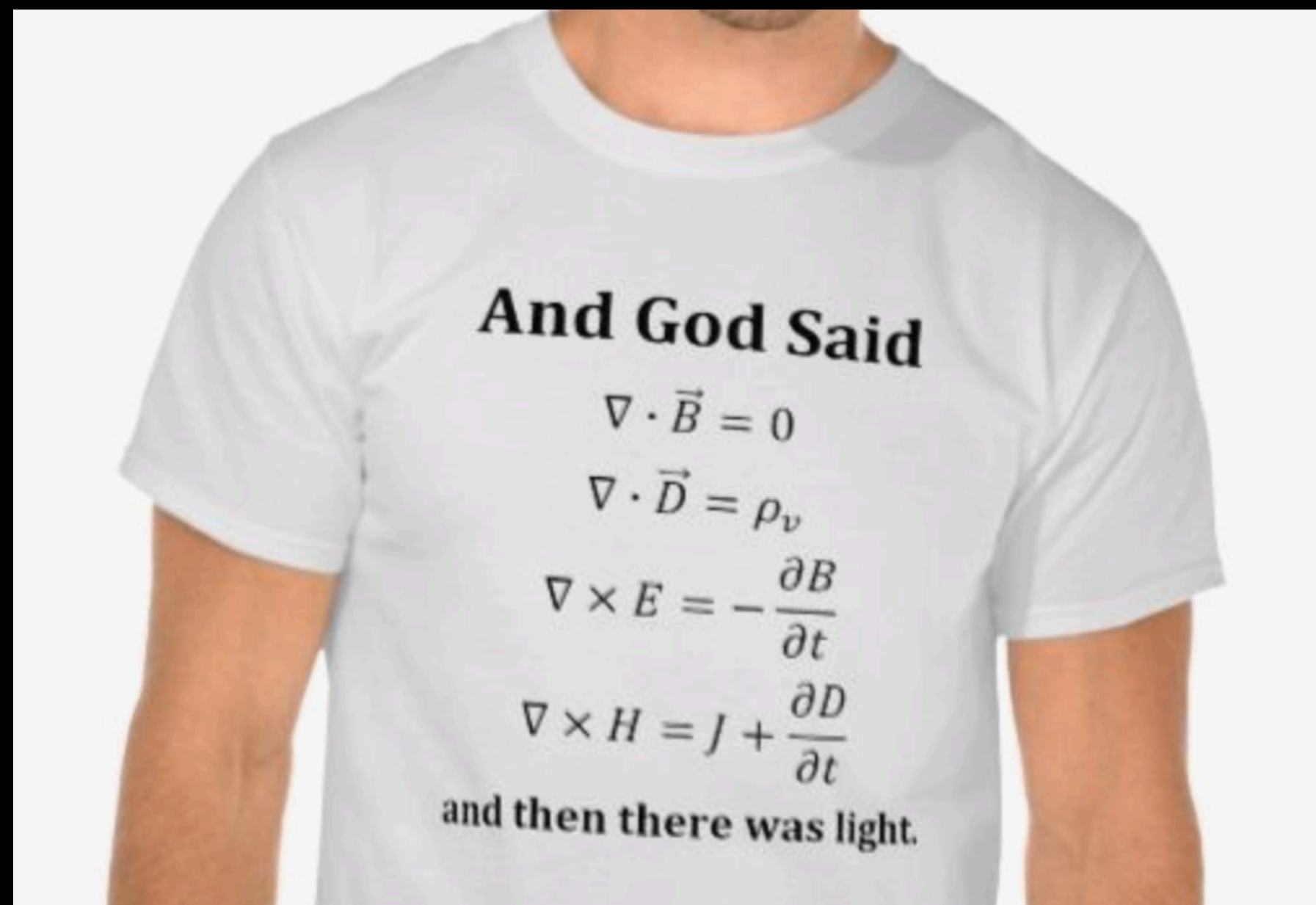
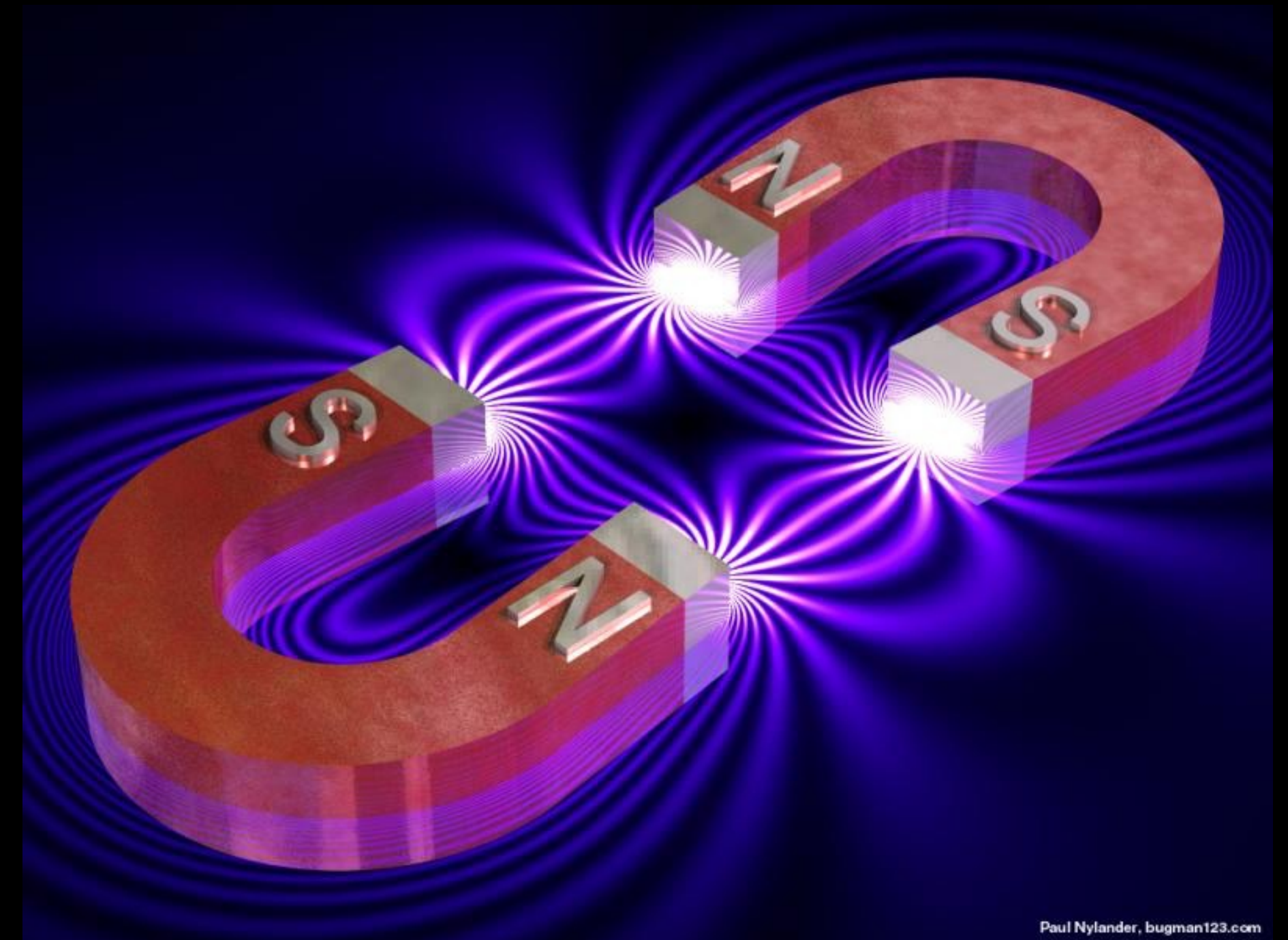
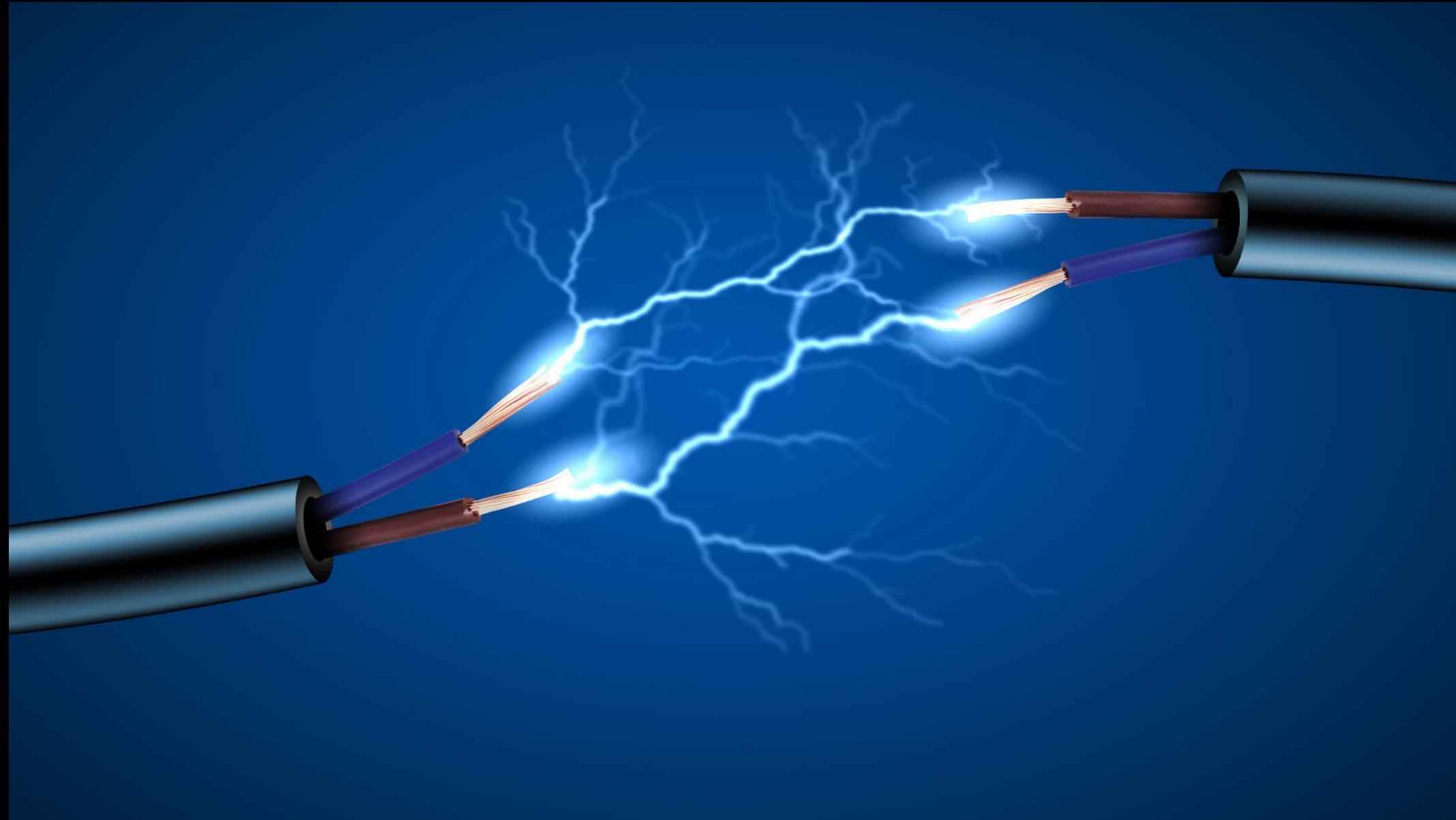
electromagnetism

Maxwell's equations
(1861)

U(1) "gauge theory"
(local symmetry)



Unification



And God Said
 $\nabla \cdot \vec{B} = 0$
 $\nabla \cdot \vec{D} = \rho_v$
 $\nabla \times E = -\frac{\partial B}{\partial t}$
 $\nabla \times H = J + \frac{\partial D}{\partial t}$
and then there was light.

electromagnetism

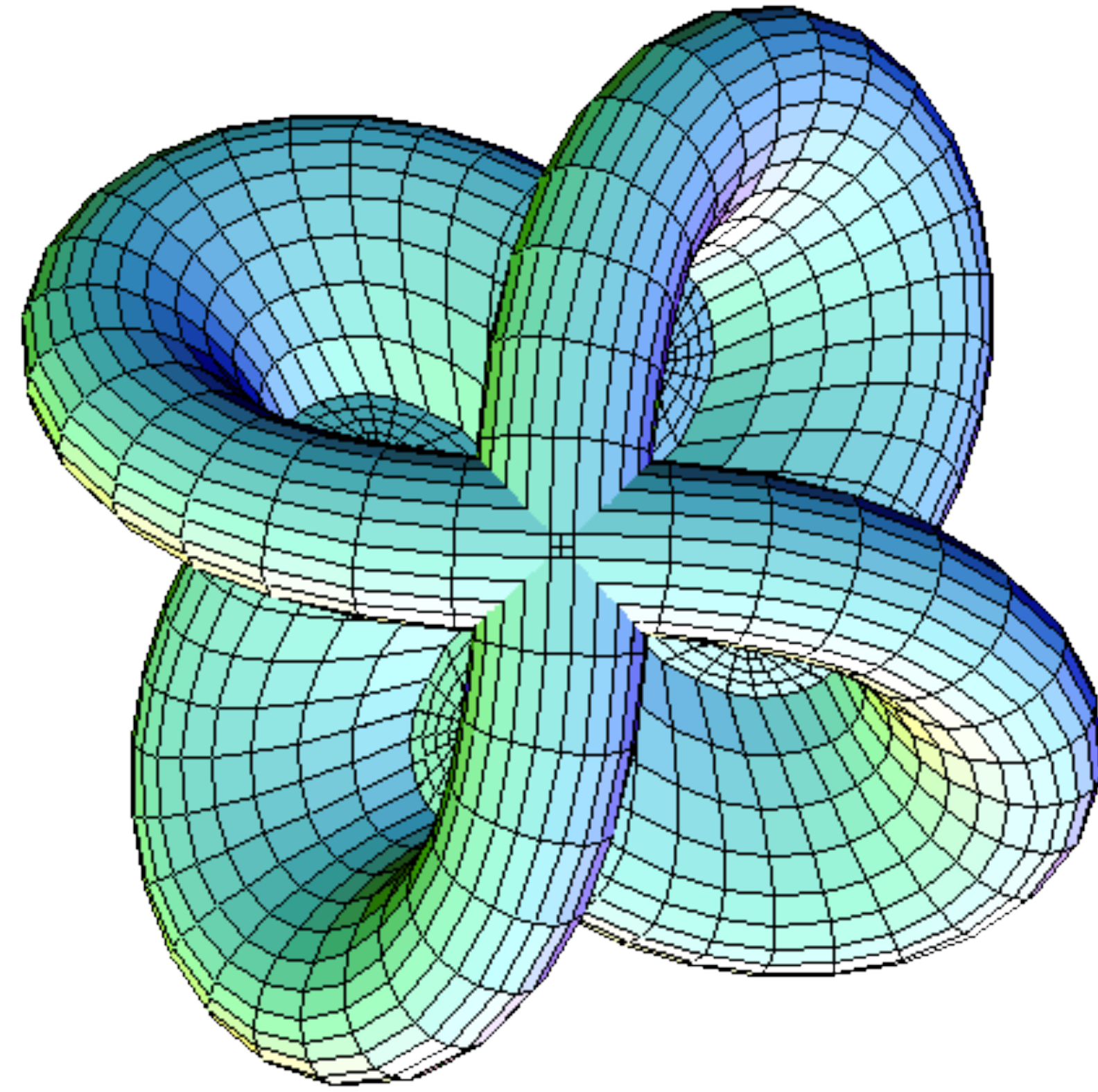
Electric-magnetic duality



(global symmetry)

Yang-Mills theory on 4-manifolds

$$S = -\frac{1}{g^2} \int_{\mathcal{M}} \text{Tr}(F_A \wedge \star F_A)$$



\mathcal{M} smooth Riemannian 4-manifold

G compact Lie group (Lie algebra \mathfrak{g})

$A \in \Omega^1(\mathfrak{g})$ connection on $P \rightarrow \mathcal{M}$ (principal G -bundle)

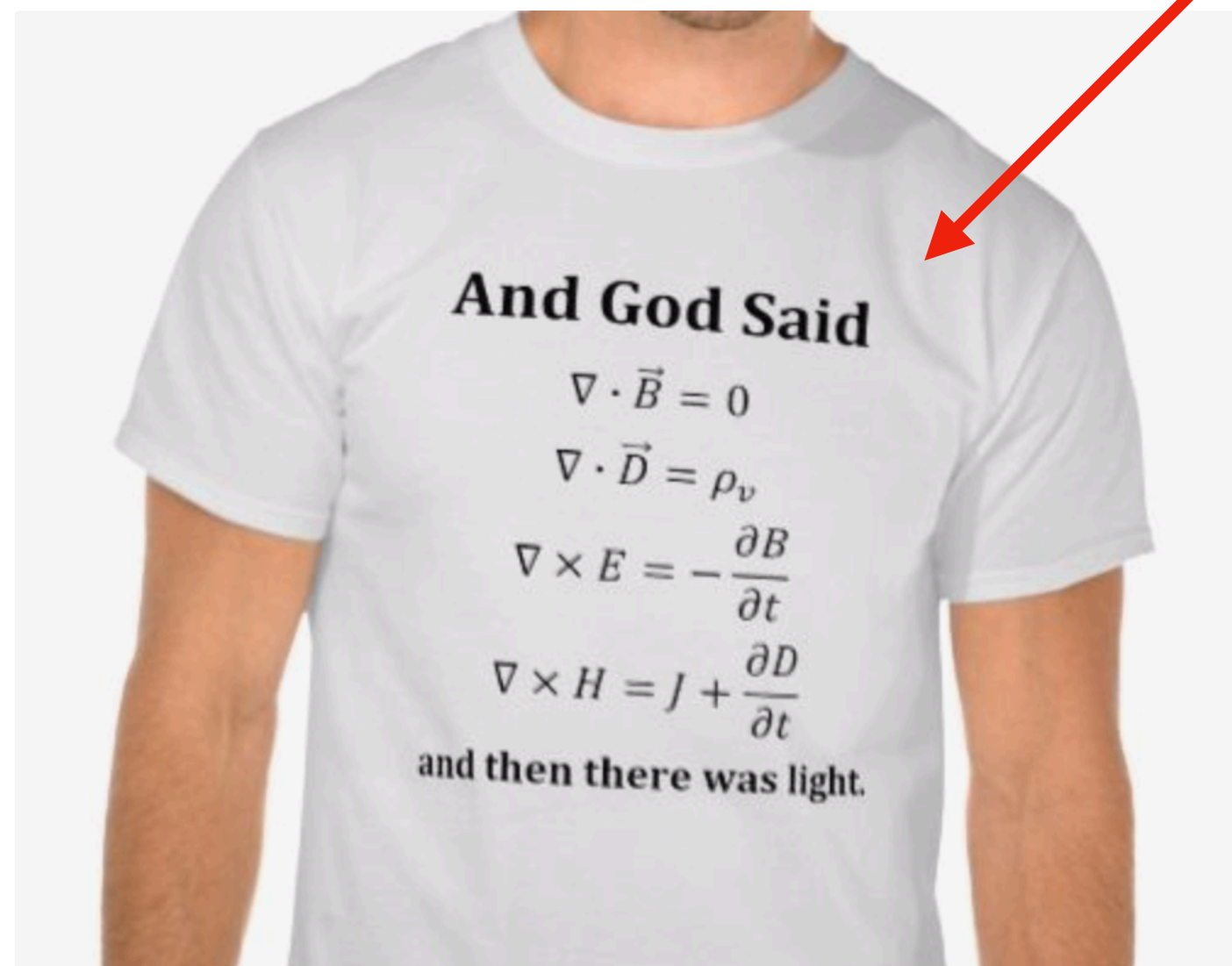
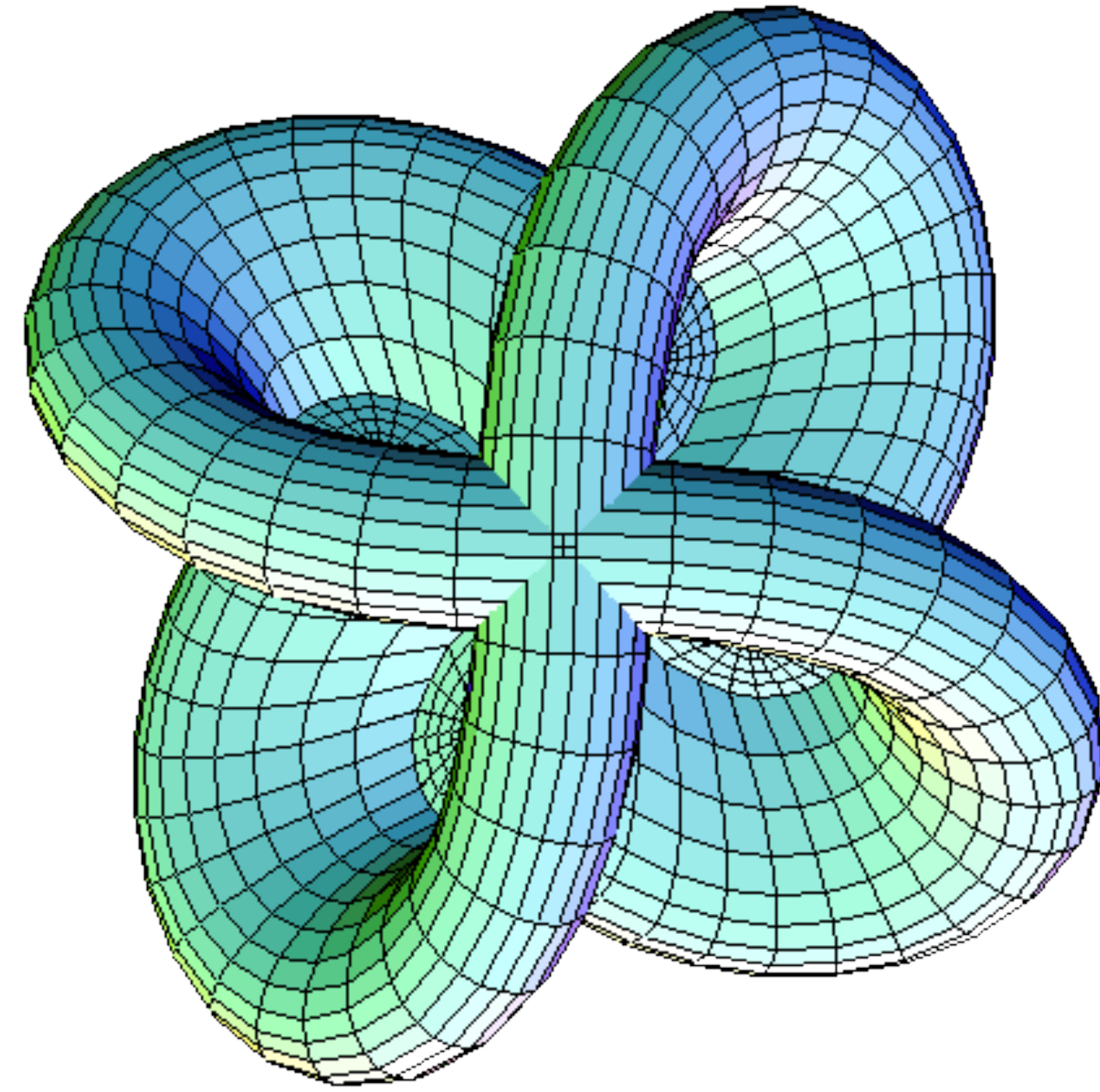
$F_A = dA + A \wedge A \in \Omega^2(\mathfrak{g})$ (curvature/field strength)

Yang-Mills theory on 4-manifolds

$$S = -\frac{1}{g^2} \int_{\mathcal{M}} \text{Tr}(F_A \wedge \star F_A)$$

Critical points give the Yang-Mills equations:

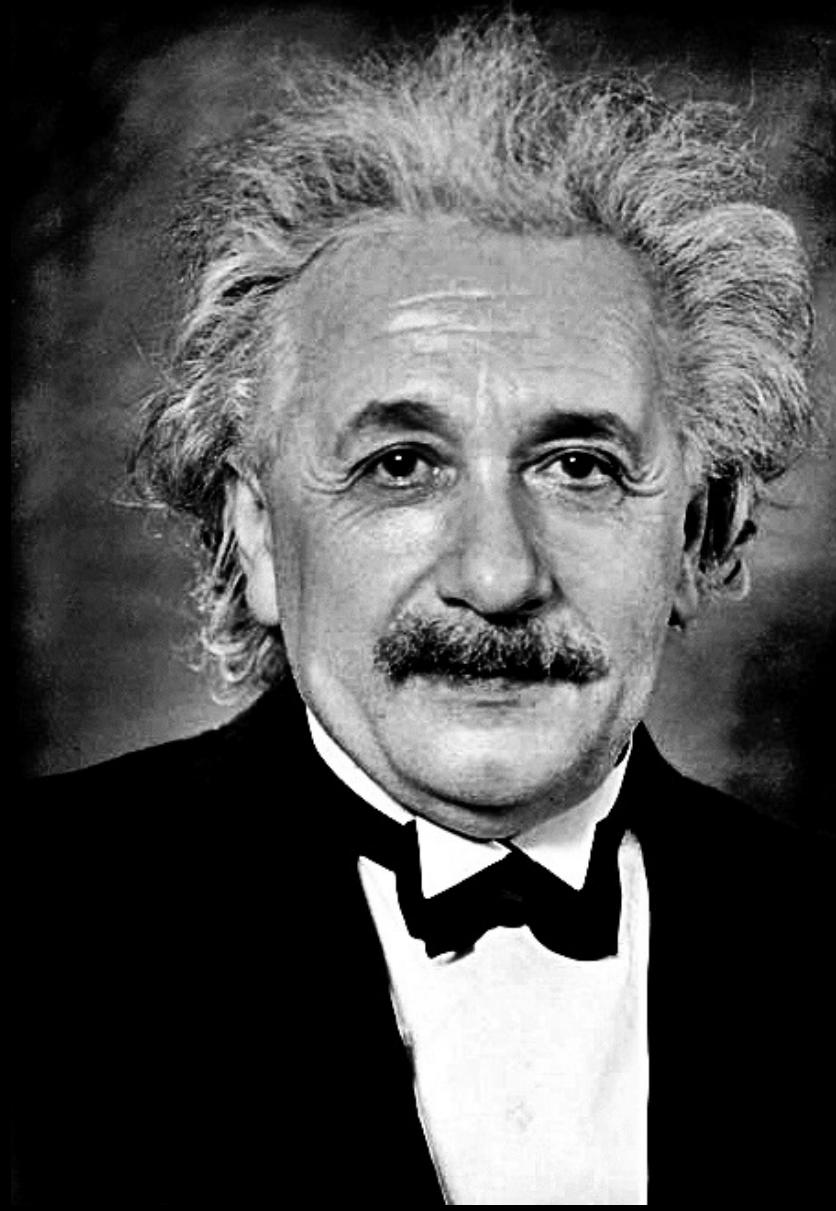
$$\delta S = 0 \quad \longleftrightarrow \quad d \star F_A = 0$$



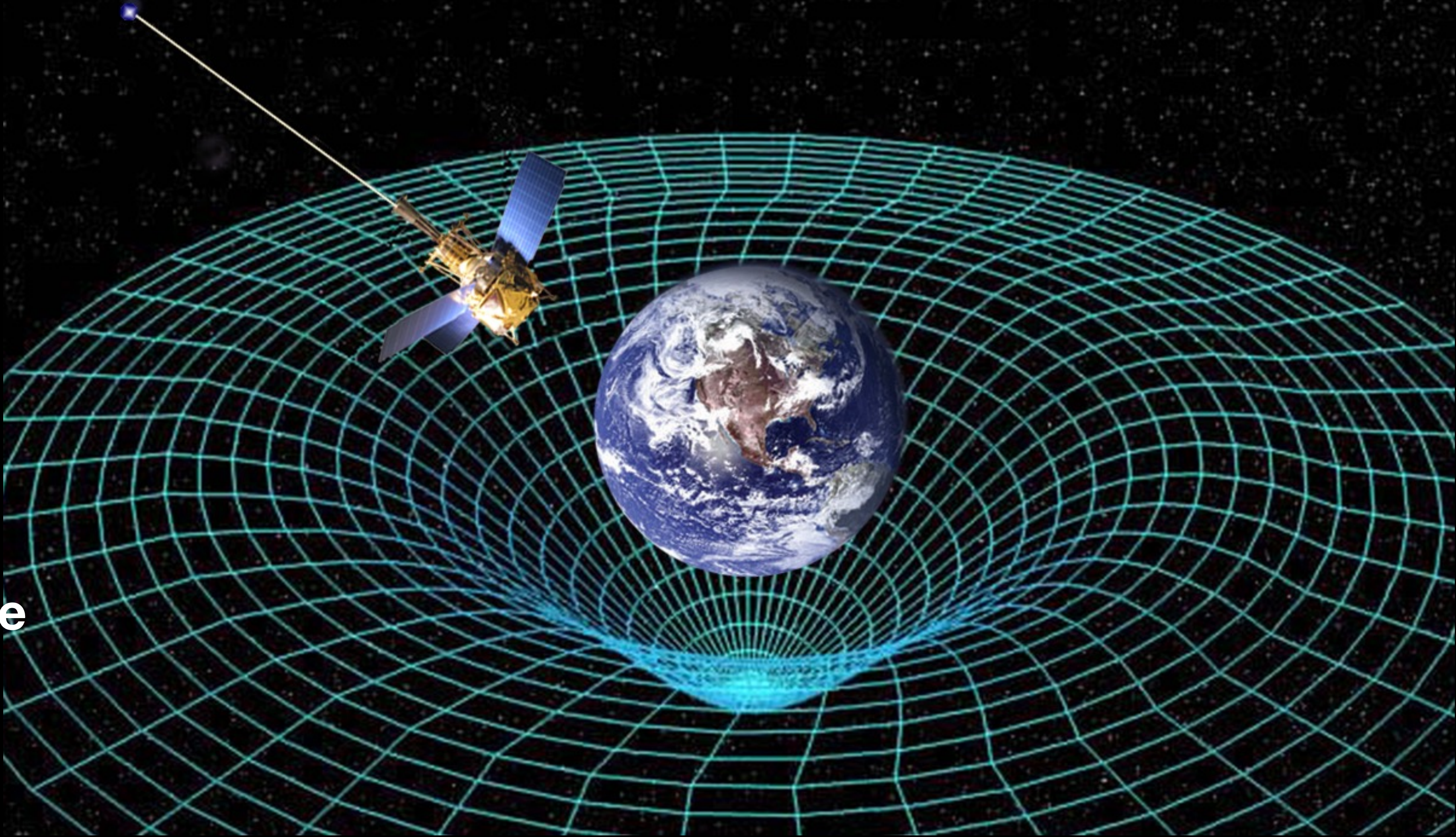
**captures
Maxwell's
equations!**

Local (gauge) symmetry: G

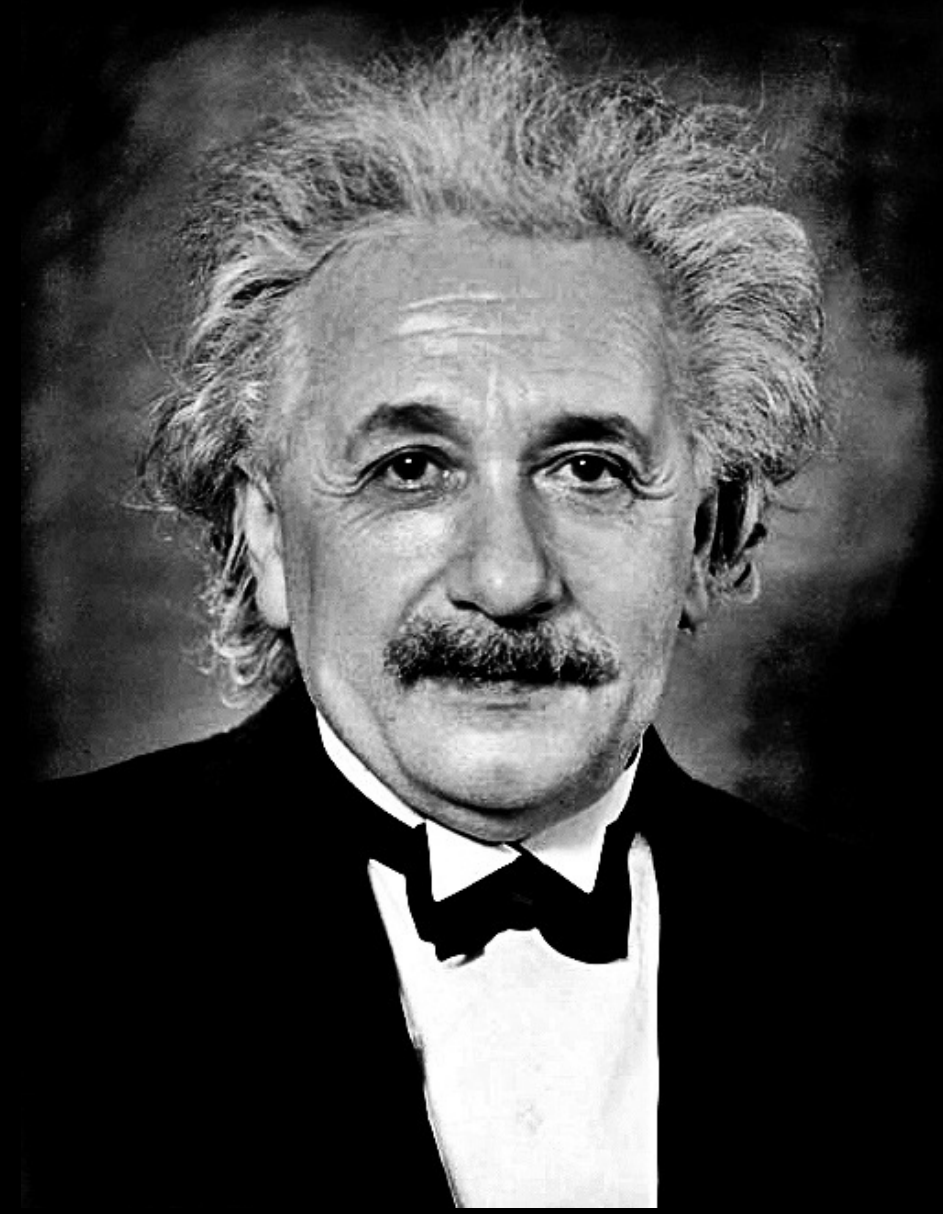
Global symmetry (S-duality): $SL(2, \mathbb{Z})$



space + time = spacetime



**gravity:
mass curves spacetime**

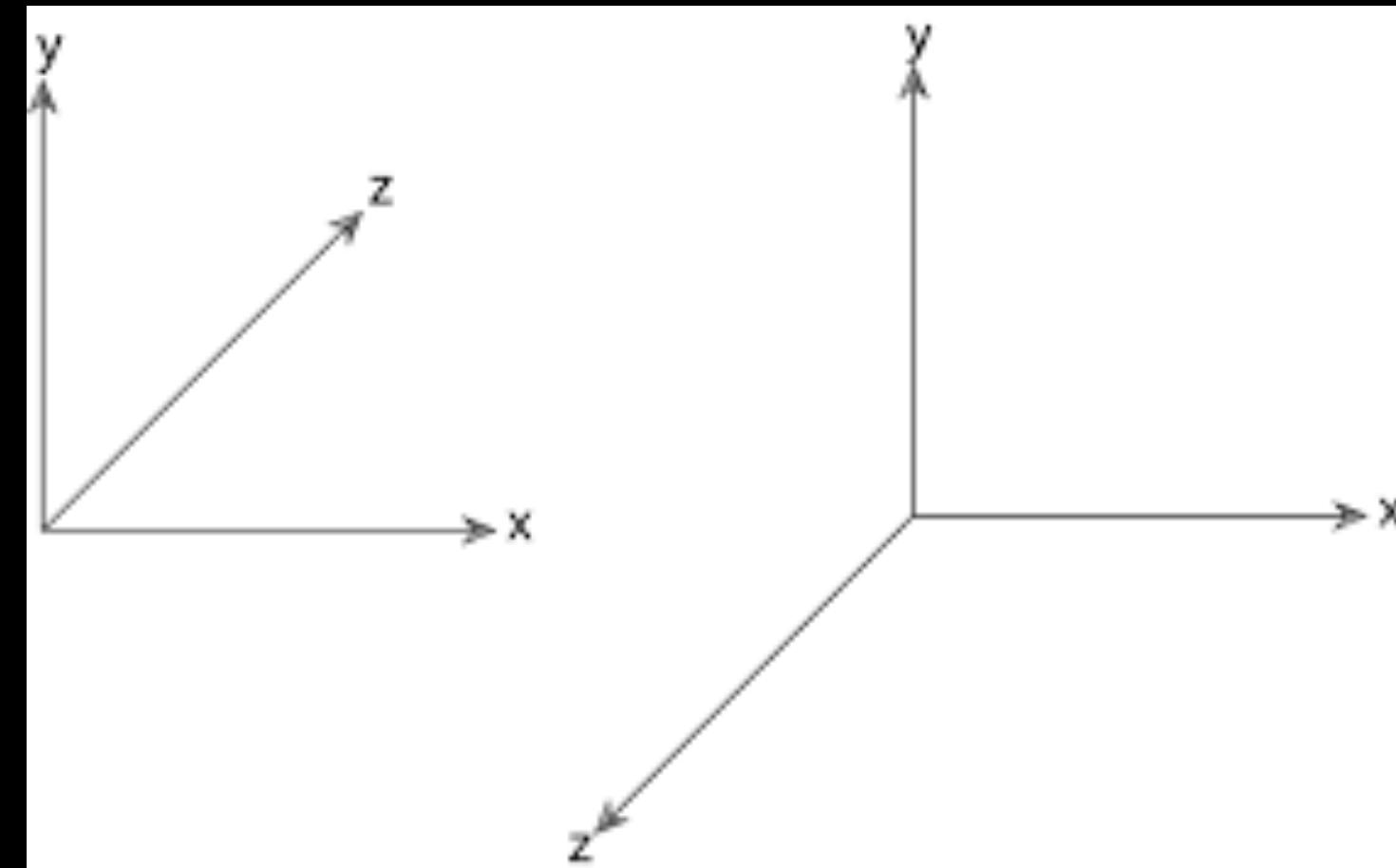


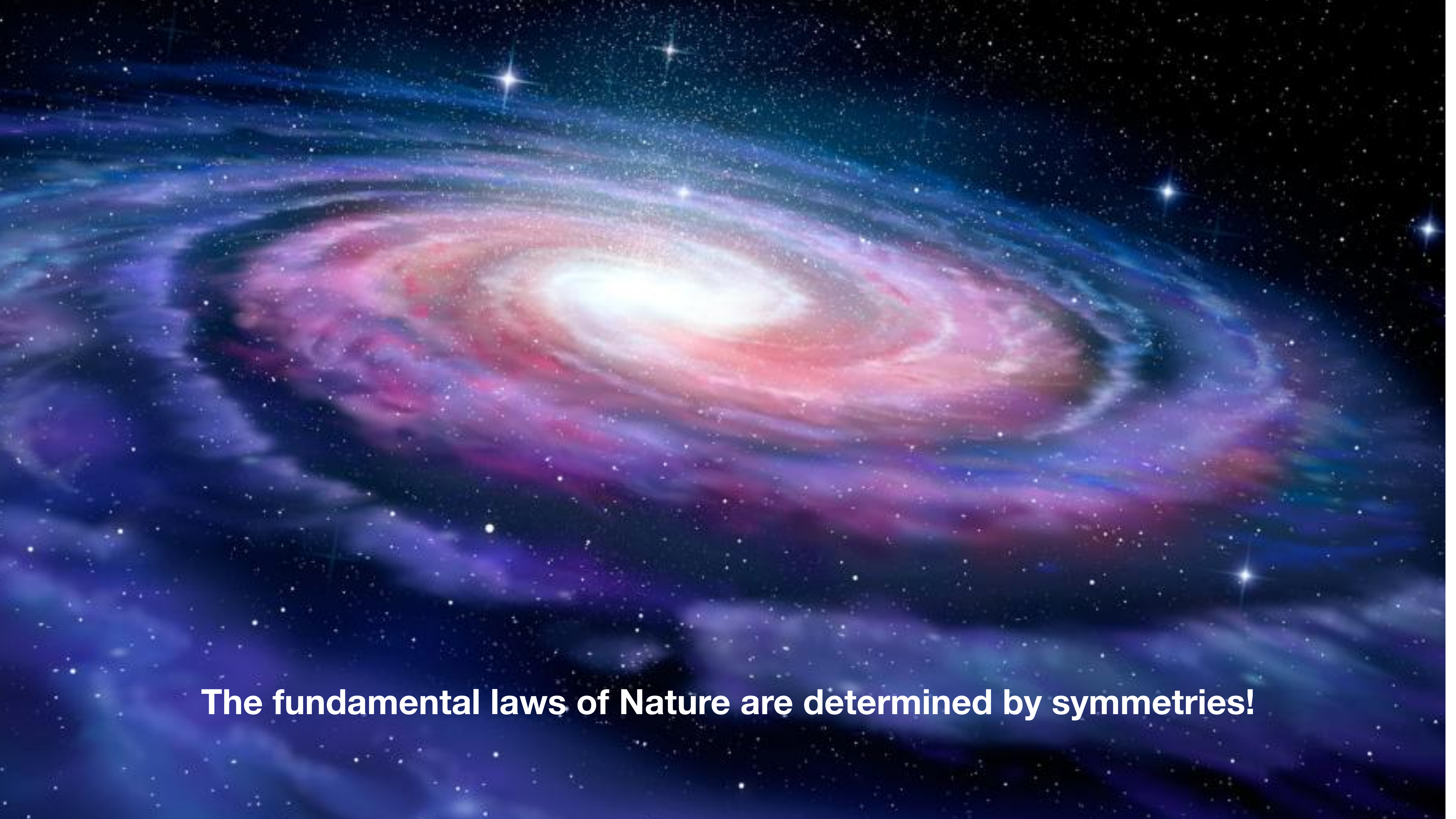
Einstein's general theory of relativity connected differential geometry and physics

Principle of general covariance:

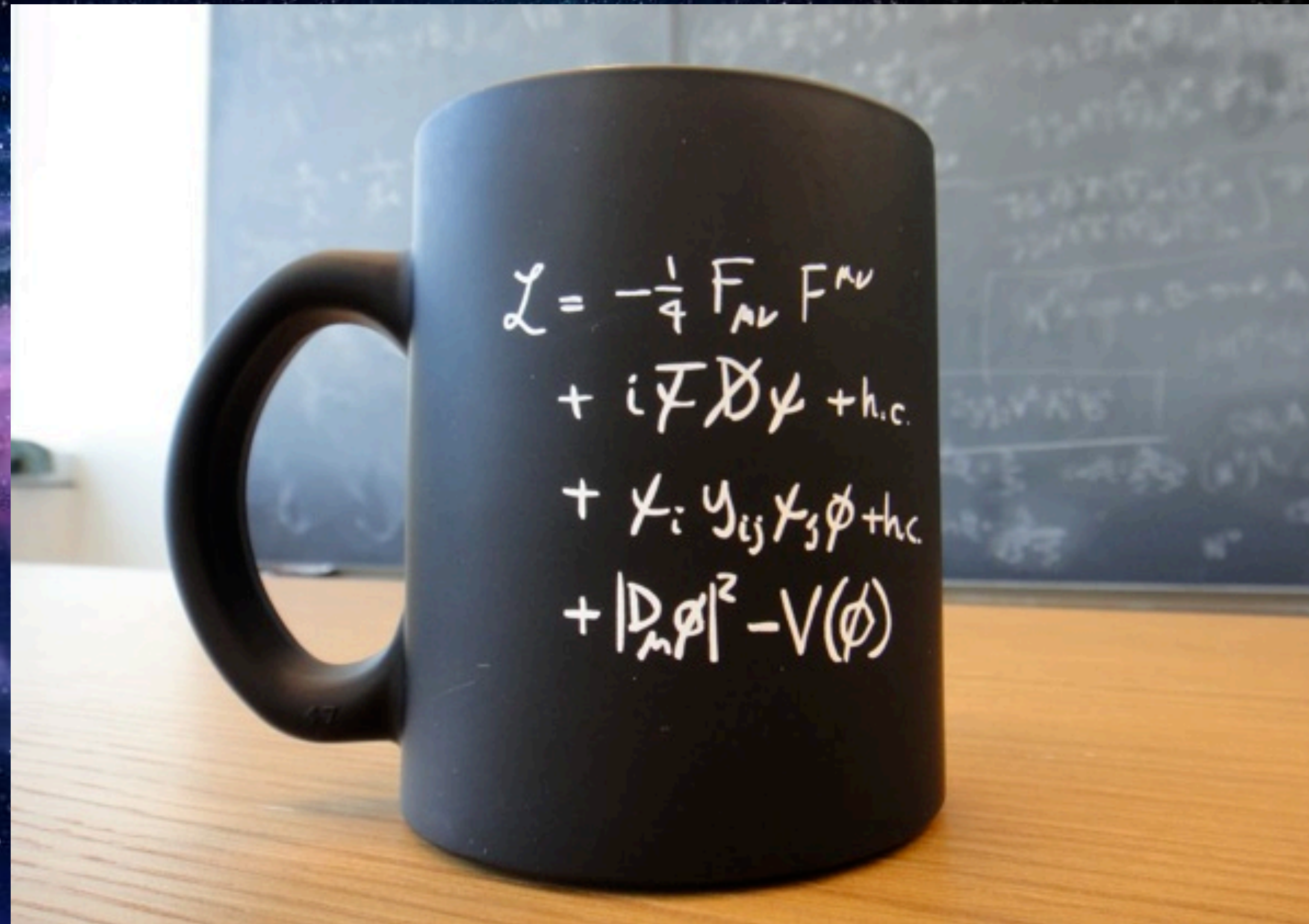
The laws of physics should take the same form independently of which coordinate system we use to represent them

$$R_{\mu\nu}(g) - \frac{1}{2}g_{\mu\nu}R(g) = T_{\mu\nu}(g, F, \dots)$$



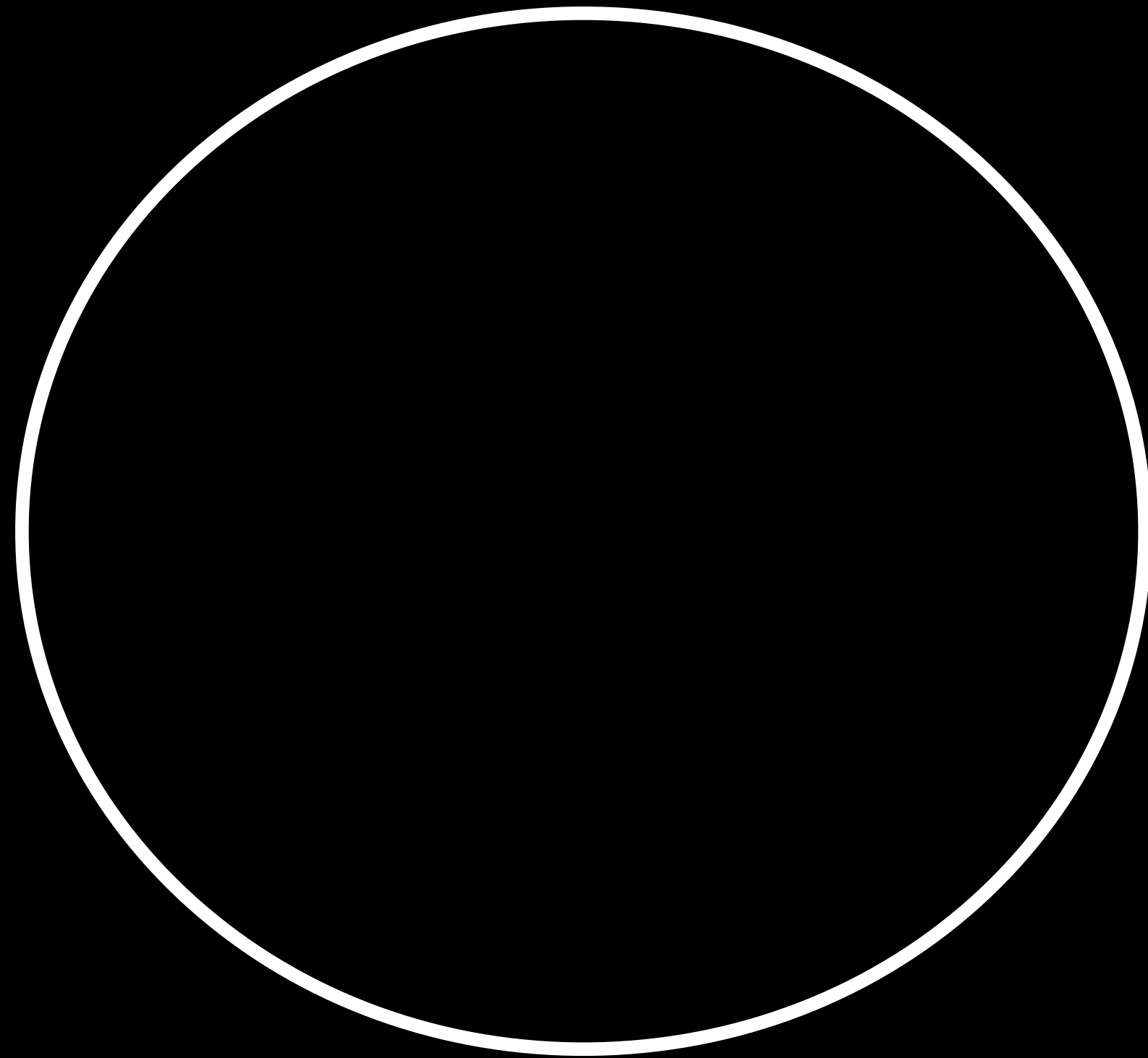


The fundamental laws of Nature are determined by symmetries!

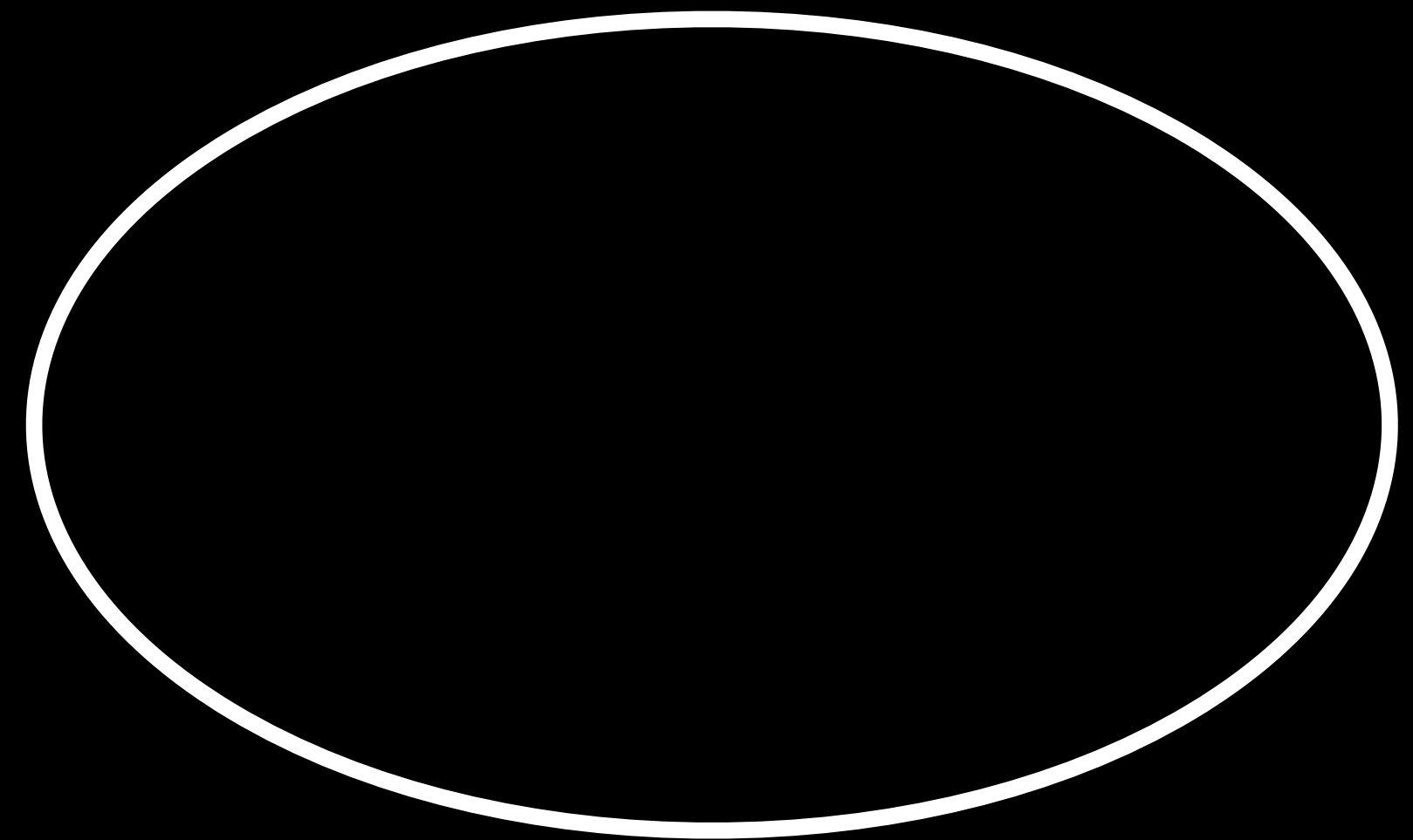


The fundamental laws of Nature are determined by symmetries!

What do we mean by symmetries?

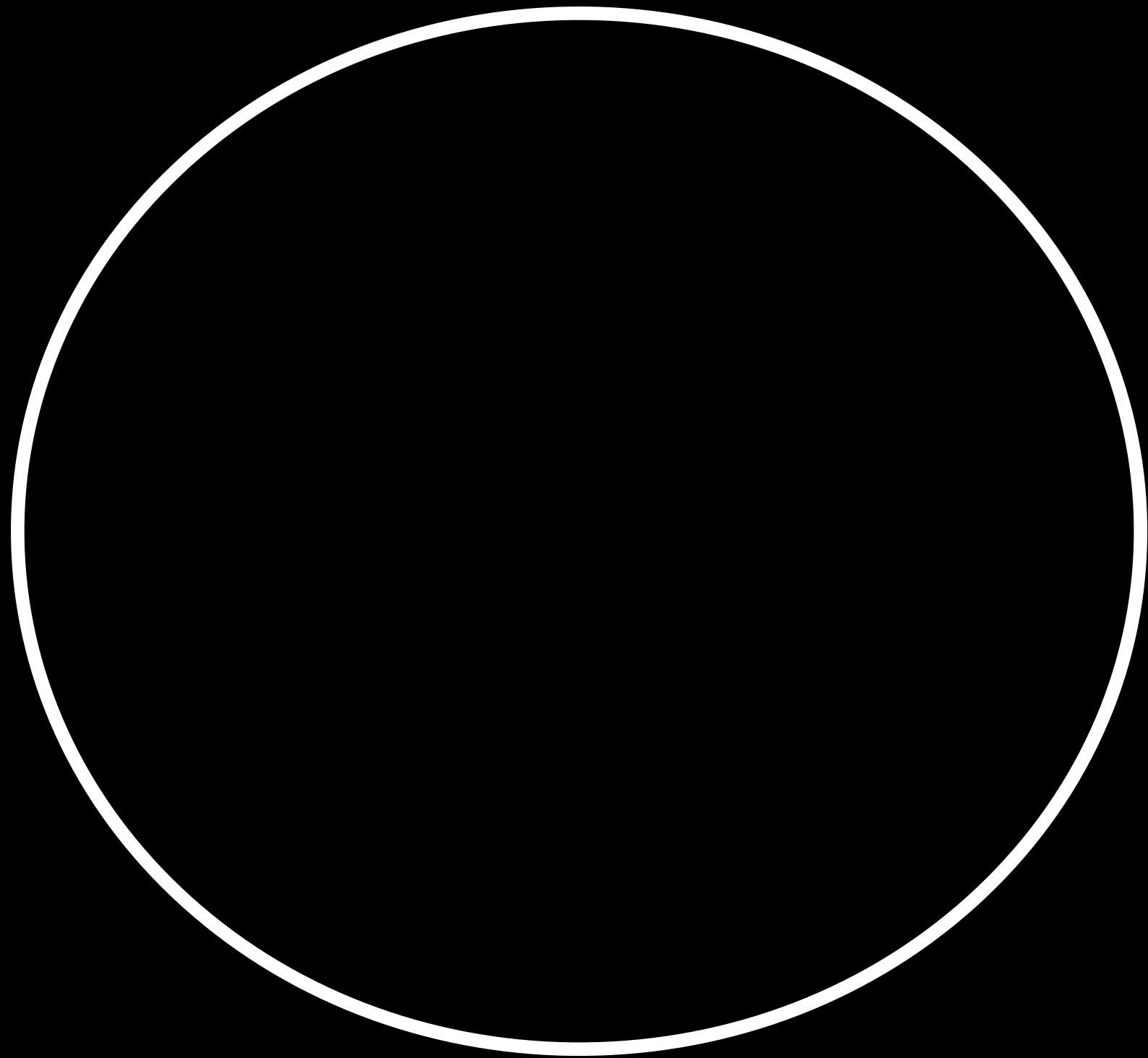


continuous rotational symmetry

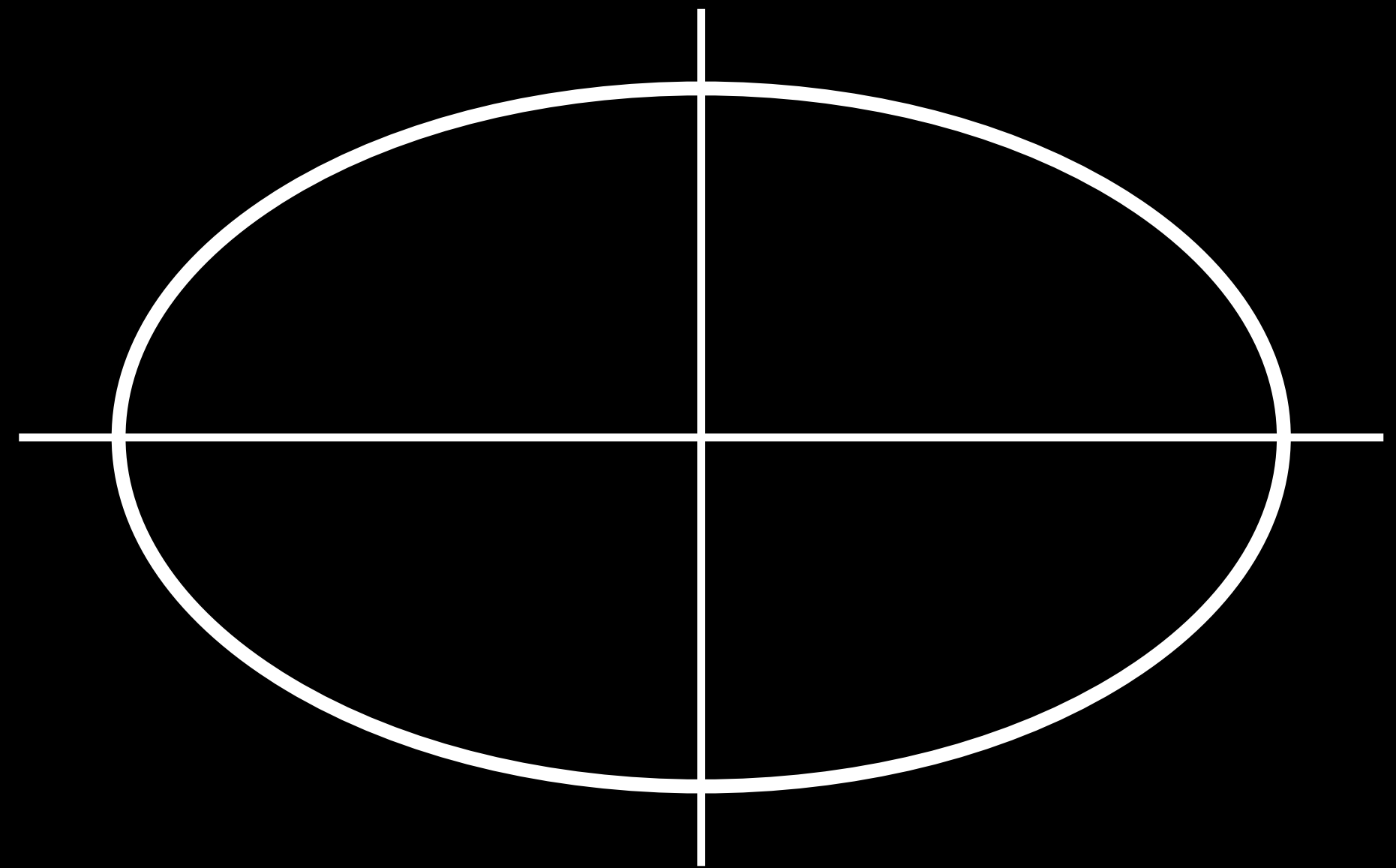


finite reflection symmetry

What do we mean by symmetries?



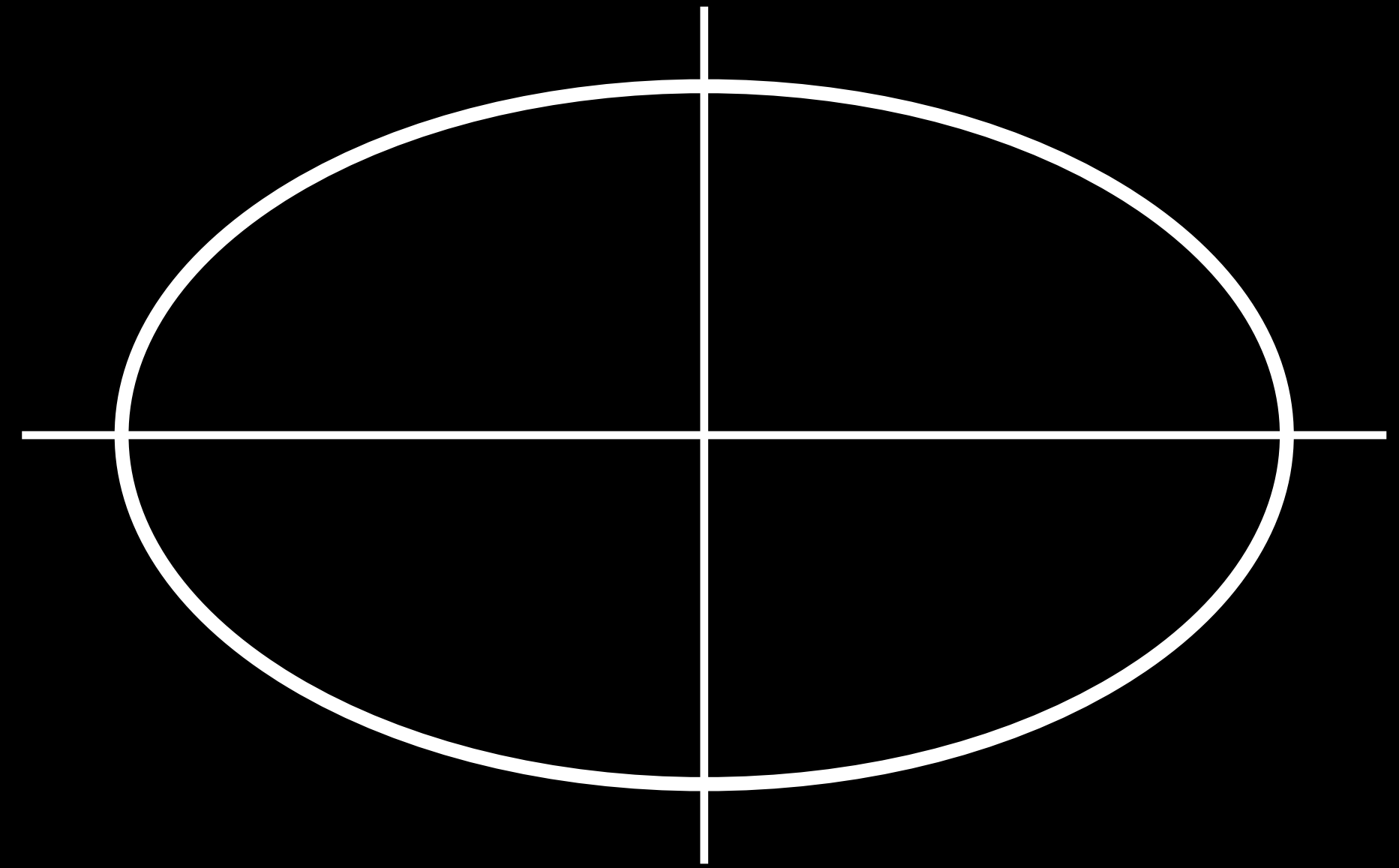
continuous rotational symmetry



finite reflection symmetry

What do we mean by symmetries?

Symmetry transformations
form a ***group***



finite reflection symmetry

Finite simple groups

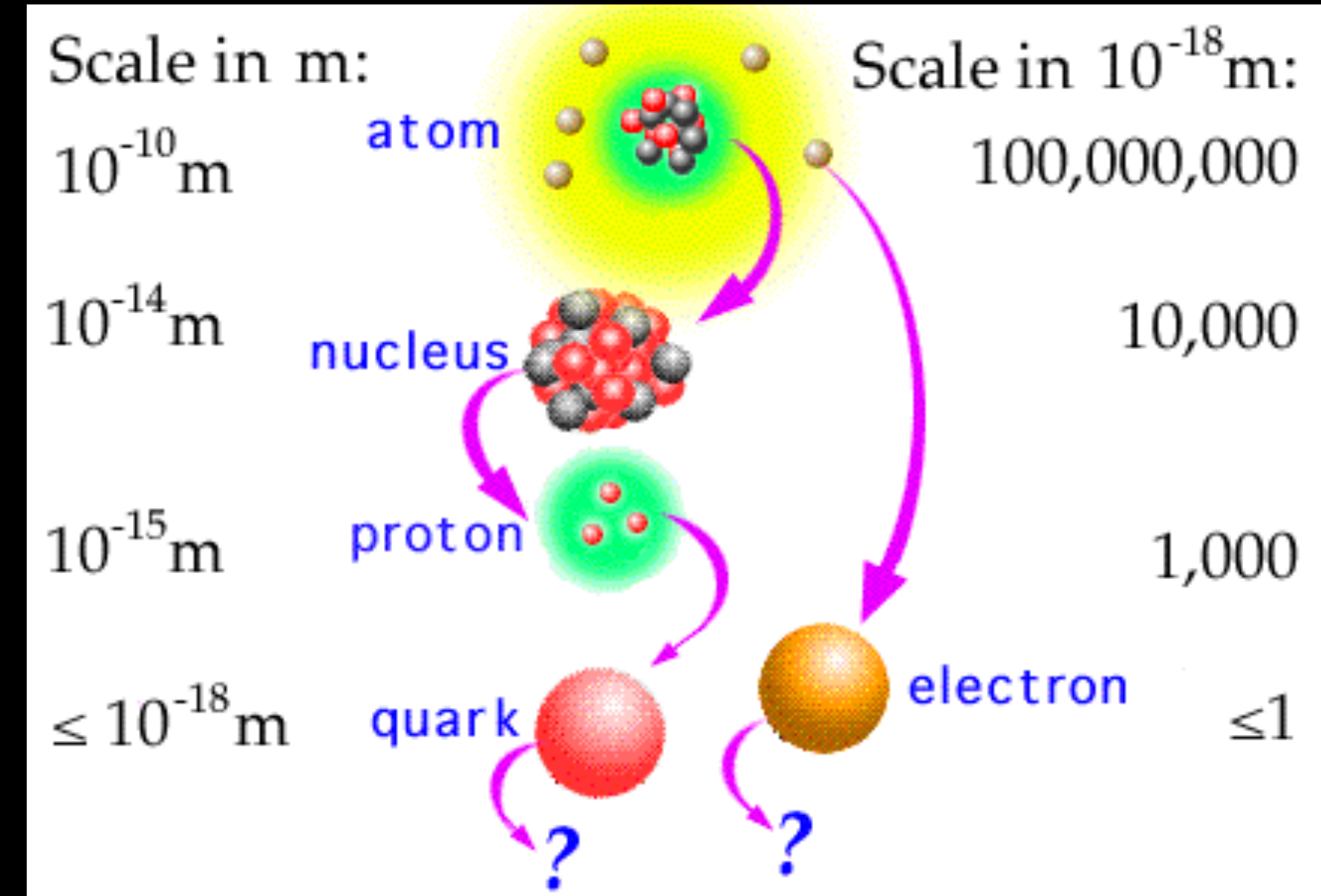
Finite groups that cannot be divided into smaller pieces are called **simple**

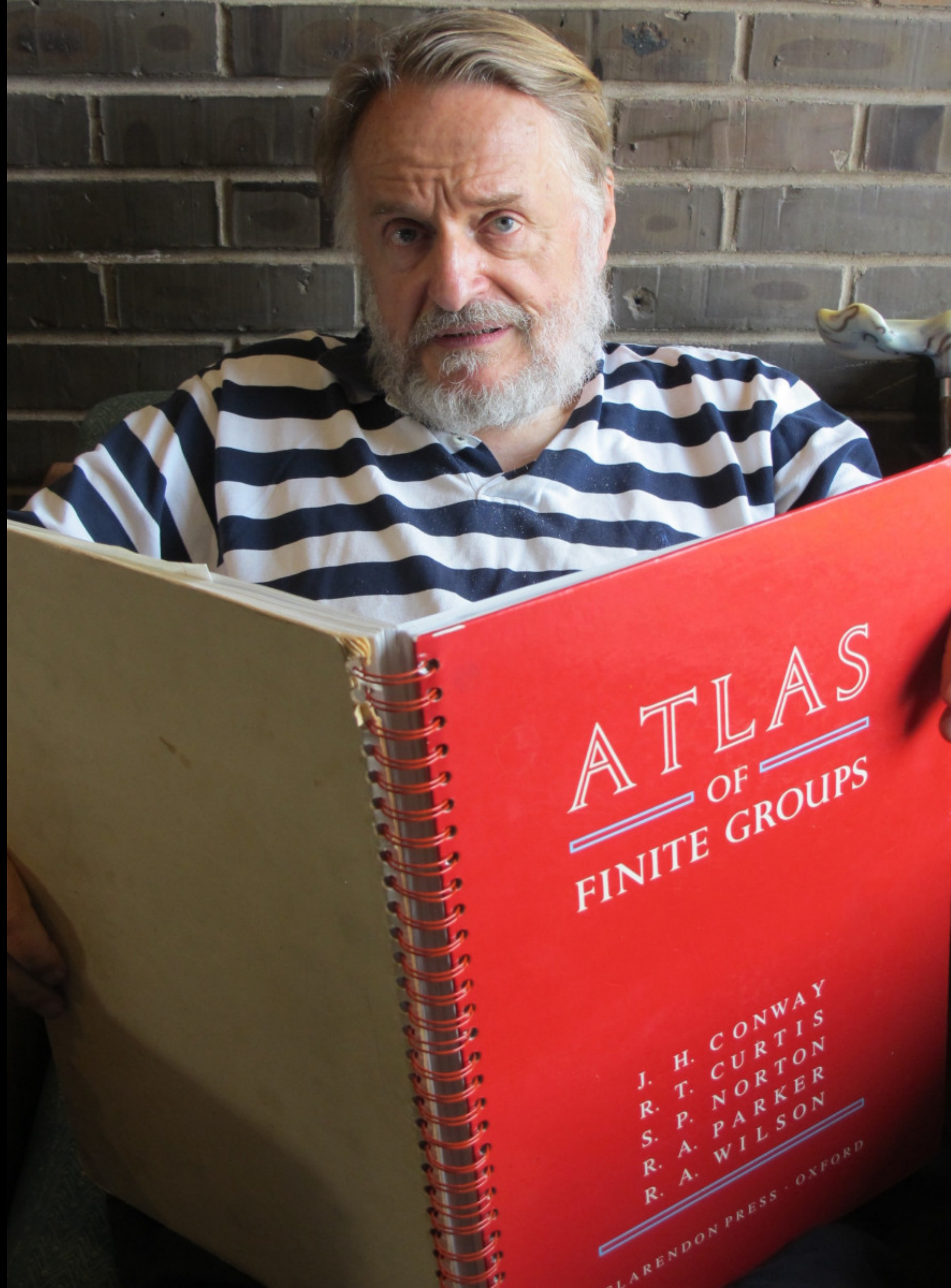


They are like **building blocks of symmetries**

Prime numbers

2, 3, 5, 7, 11, 13,
17, 19, 23, 29, 31,
37, 41, 43, 47, 53,
59, 61, 67, 71, 73,
79, 83, 89, 97





The **classification** of finite simple groups is one of the most extensive projects in mathematics

Complete proof consists of:

10 000+ pages

100+ journals

100+ mathematicians

Monster group

M



Borcherds

Fields medal in 1998

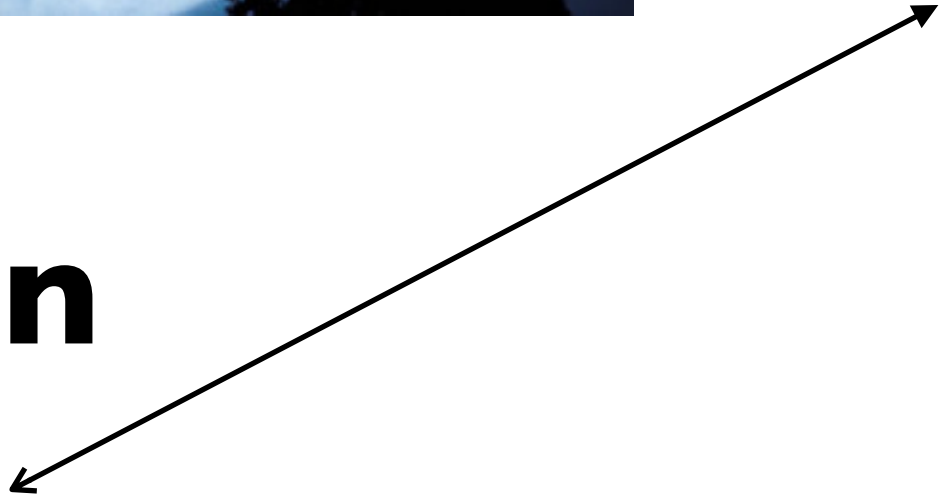


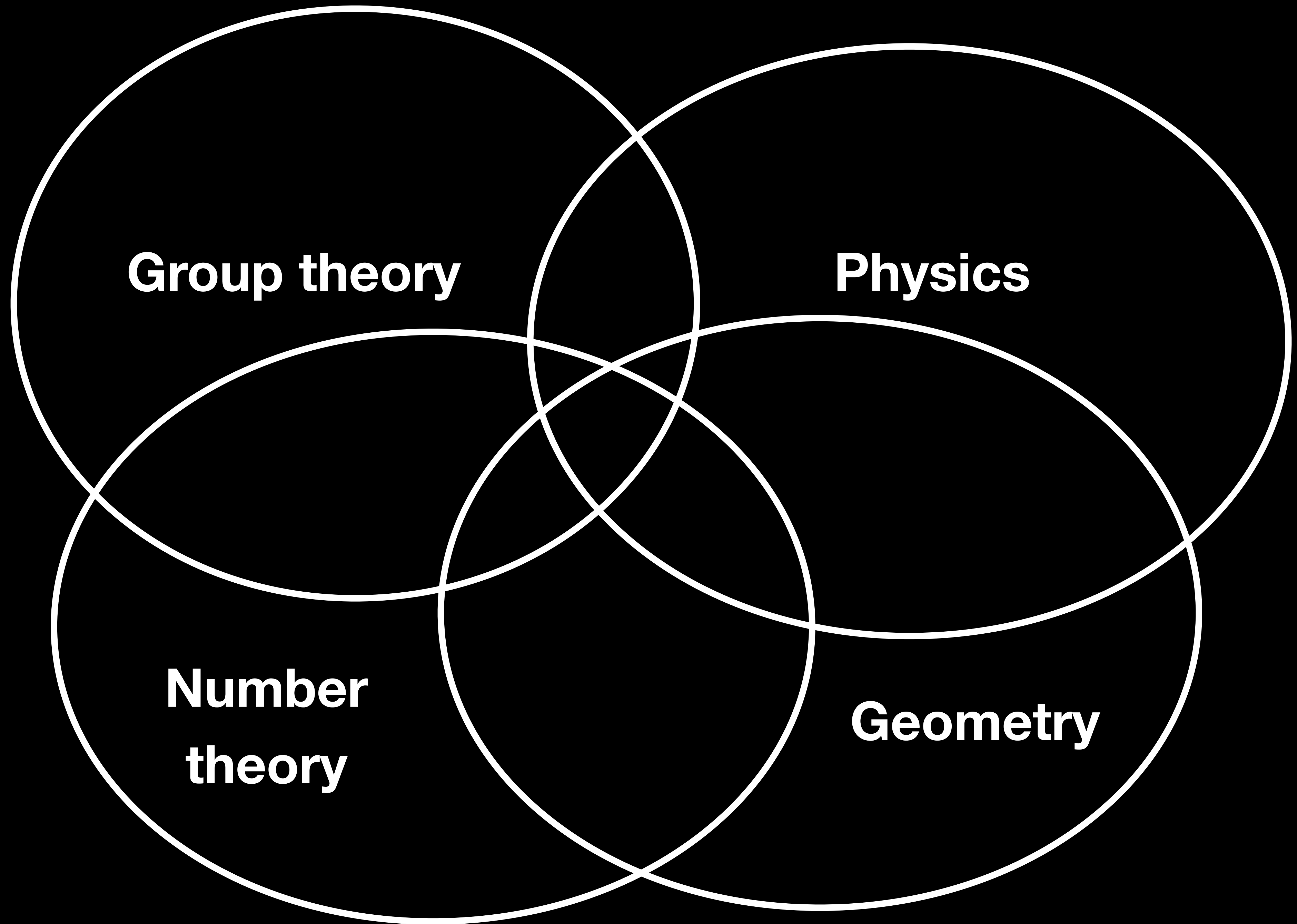
2d conformal field theory
(vertex operator algebra)

[Frenkel, Lepowsky,
Meurman]

Modular function

$J(\tau)$





Group theory

Physics

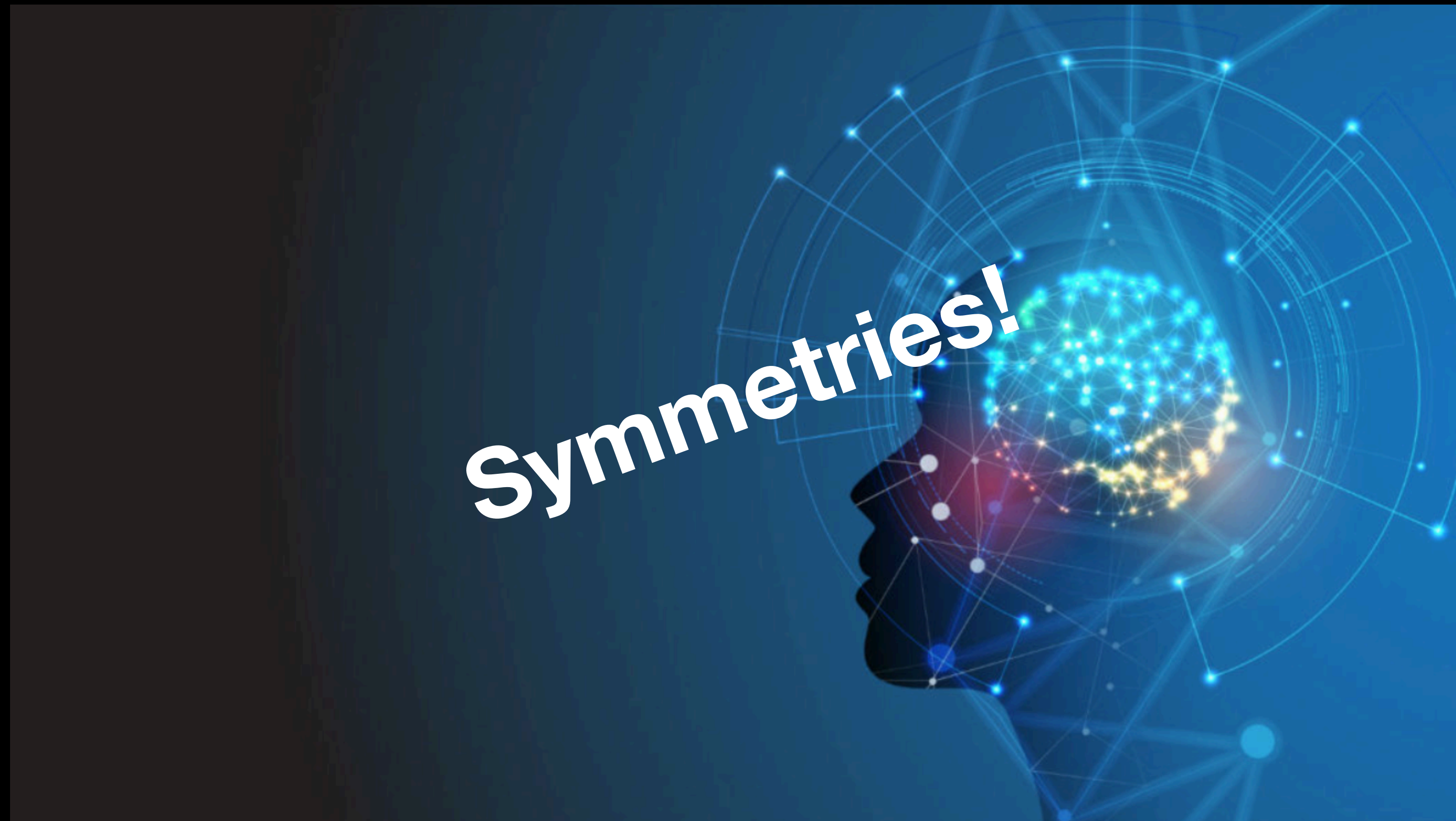
**Number
theory**

Geometry

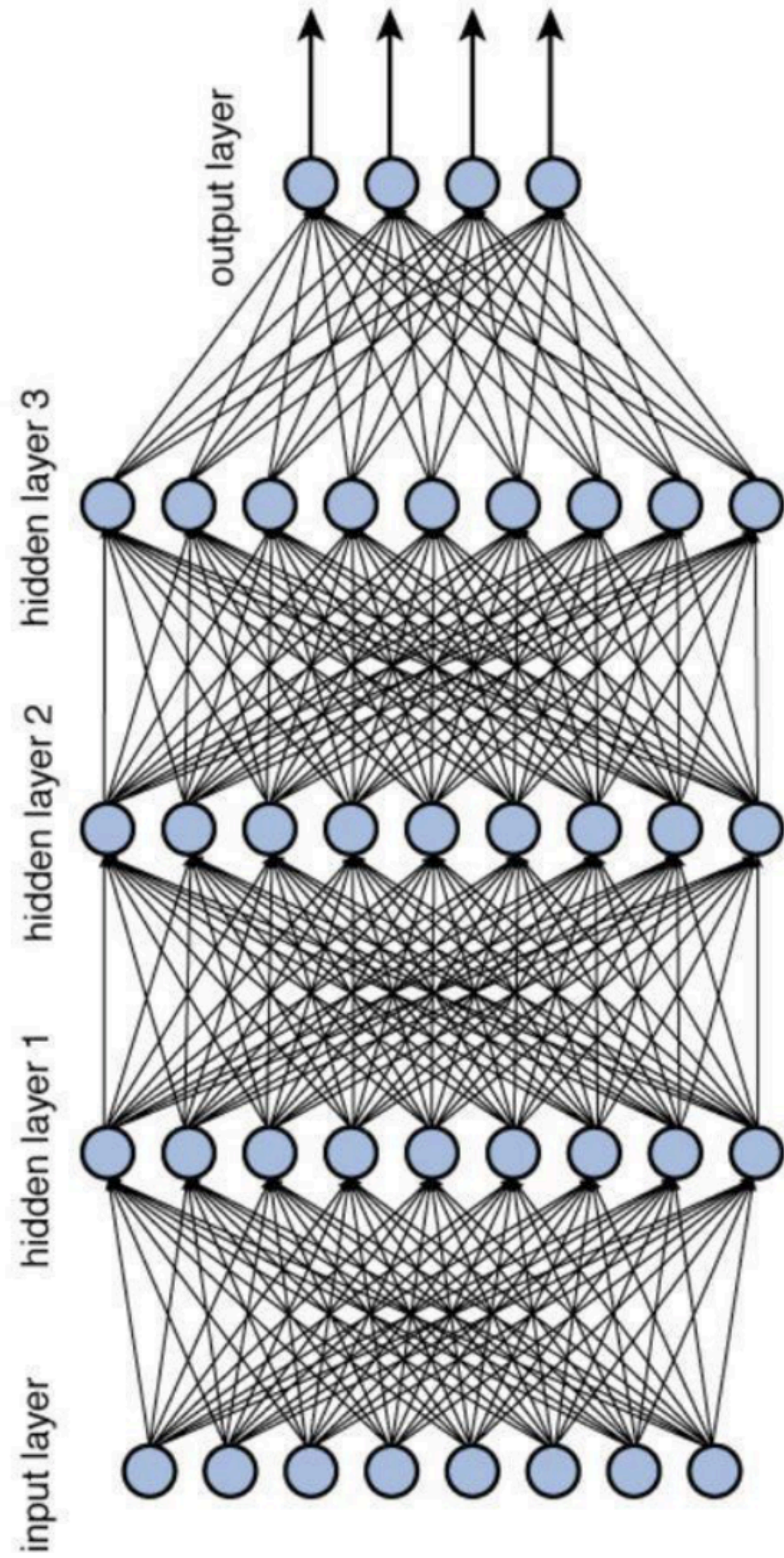
What does all this have to do with AI?



What does all this have to do with AI?



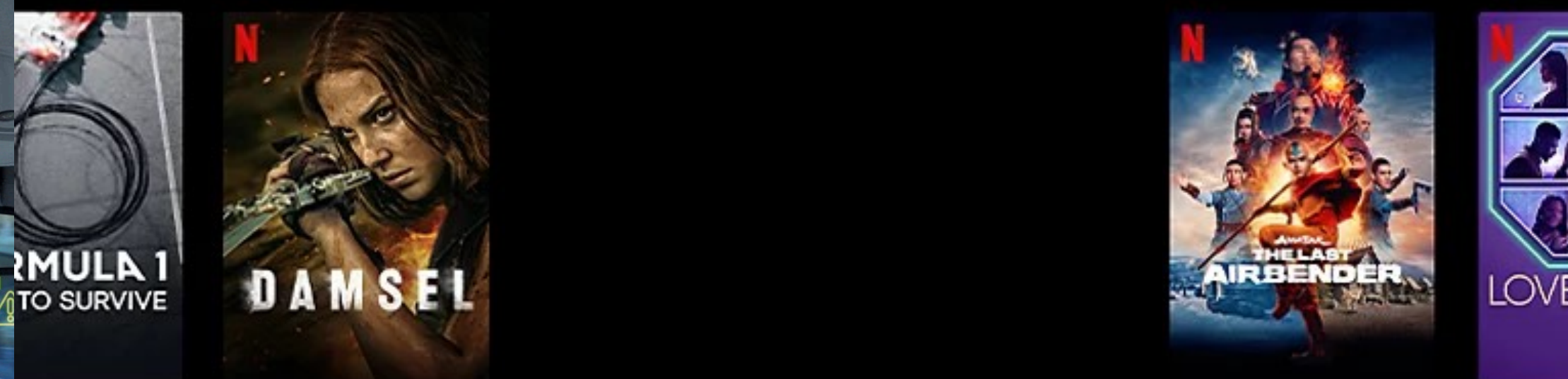
Deep Neural Network



Artificial Intelligence:
Mimicking the intelligence or behavioural pattern of humans or any other living entity.

Machine Learning:
A technique by which a computer can "learn" from data, without using a complex set of different rules. This approach is mainly based on training a model from datasets.

Deep Learning:
A technique to perform machine learning inspired by our brain's own network of neurons.





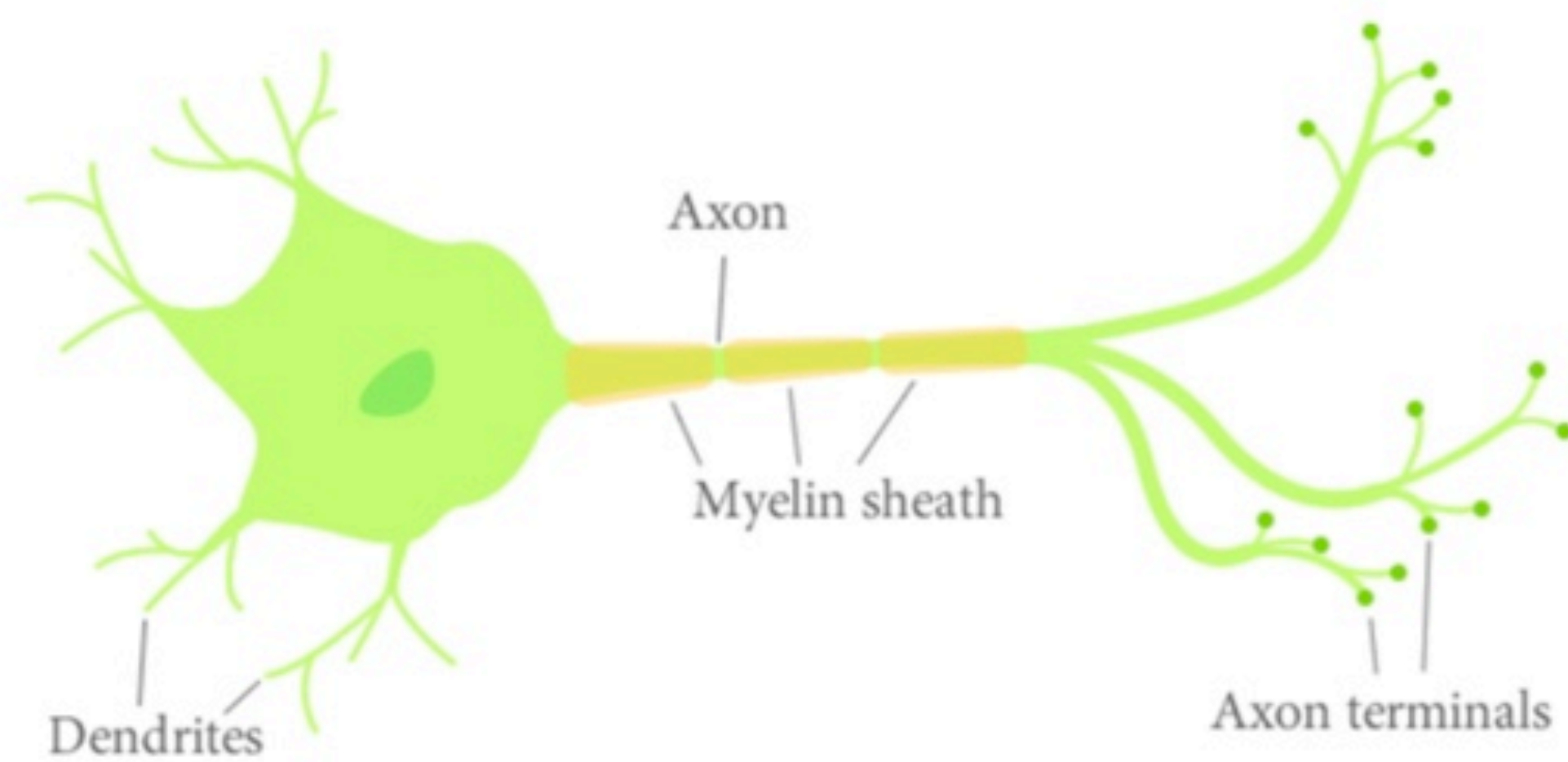




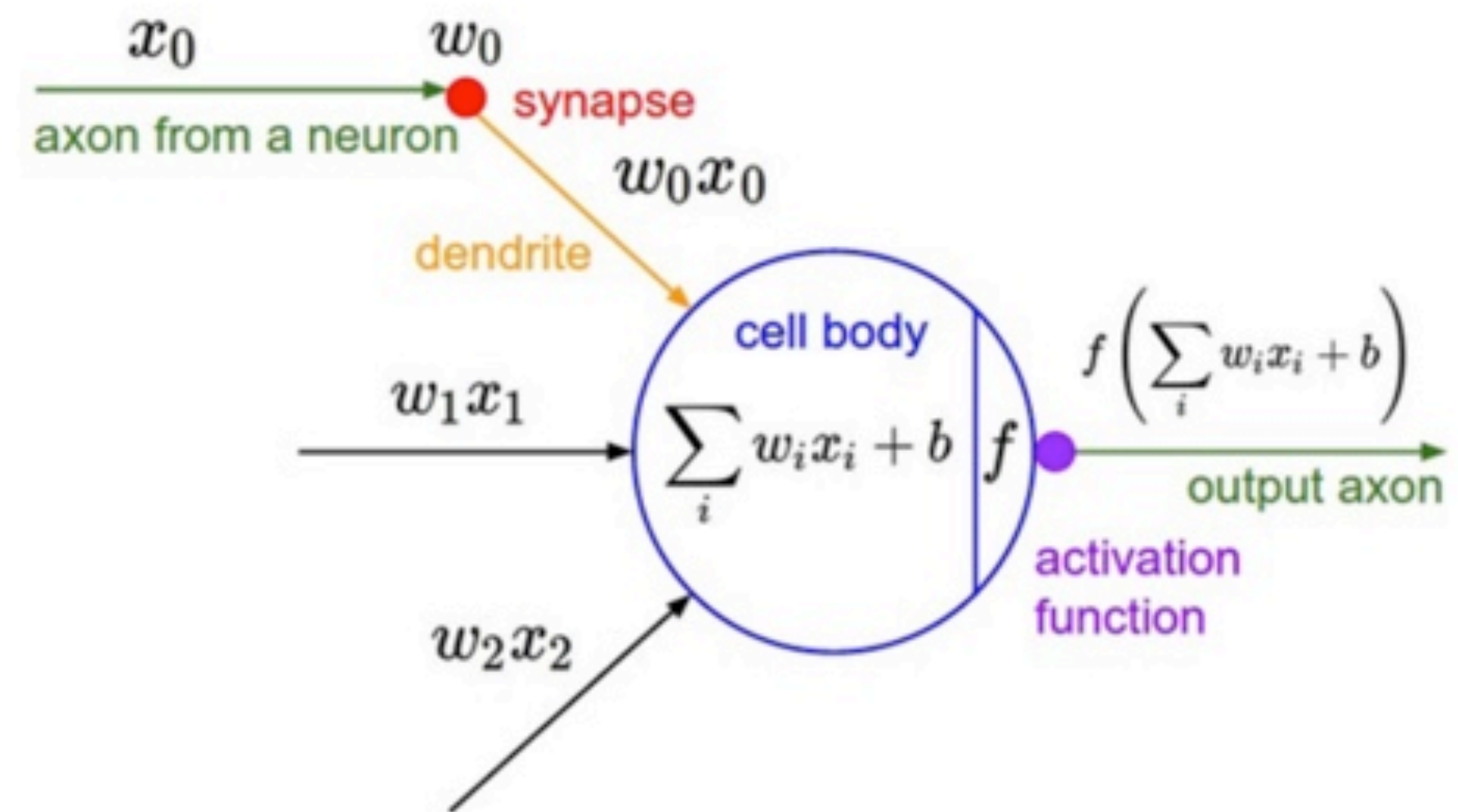
[Pic from thispersondoesnotexist.com]



Real Neuron



Artificial Neuron

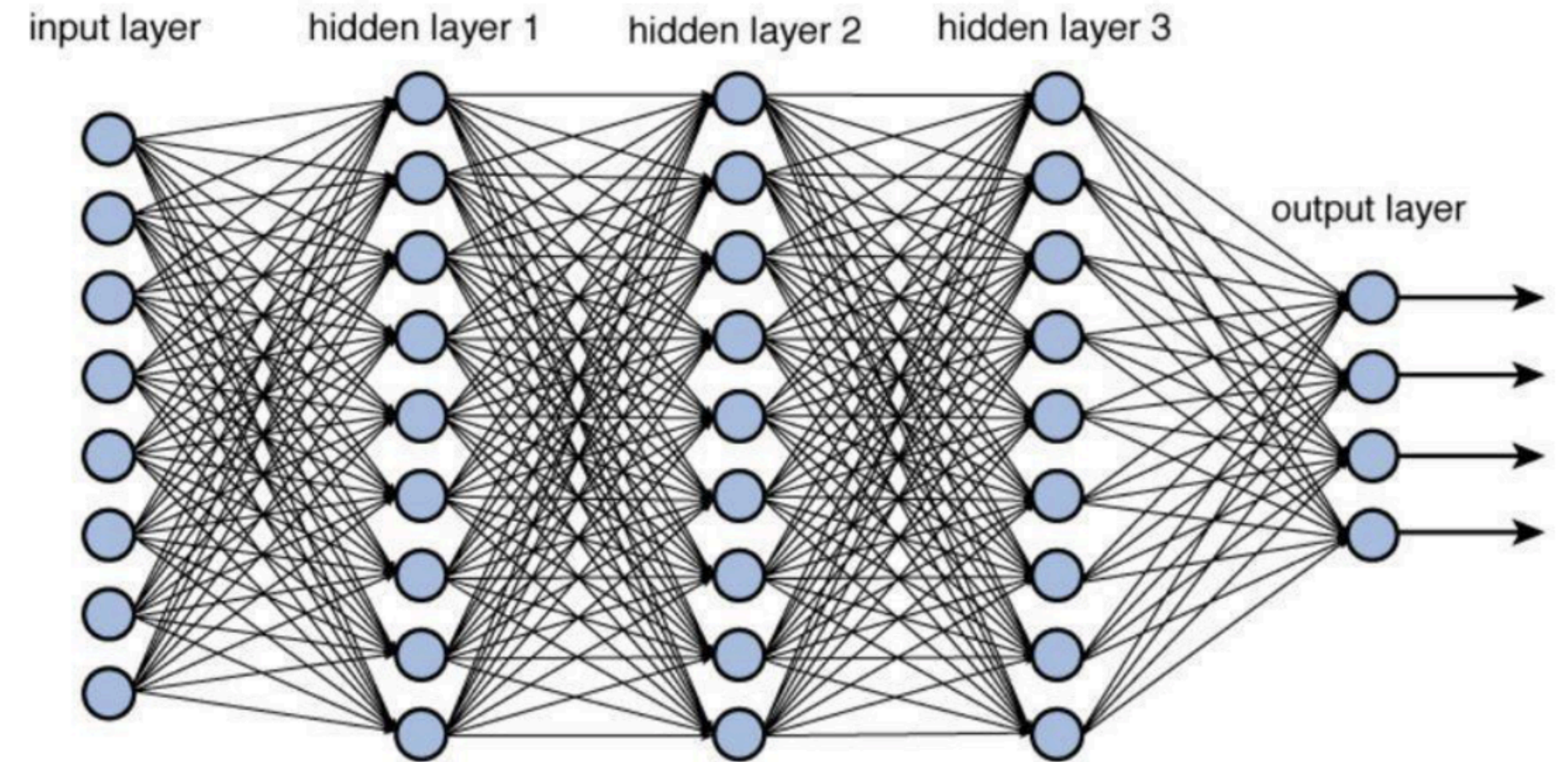


Artificial neuron:

1. Input layer
2. Hidden layers
3. Output layer



Deep Neural Network



Artificial neuron:

1. Input layer
2. Hidden layers
3. Output layer

Inductive bias

Incorporate known constraints into the architecture/training

symmetries

differential equations

geometry

Image classification

Image classification

“skateboard”

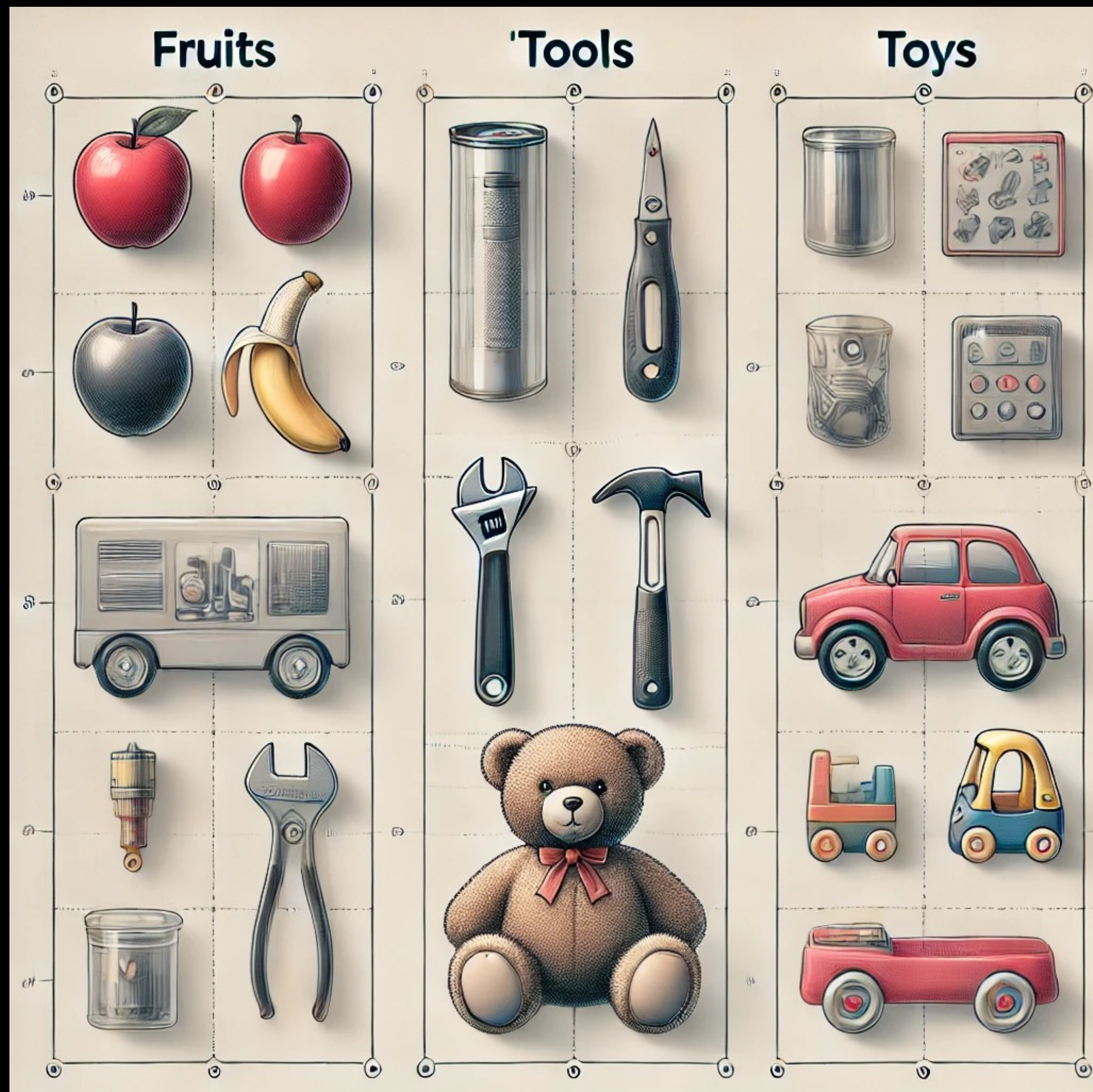


“ribs”

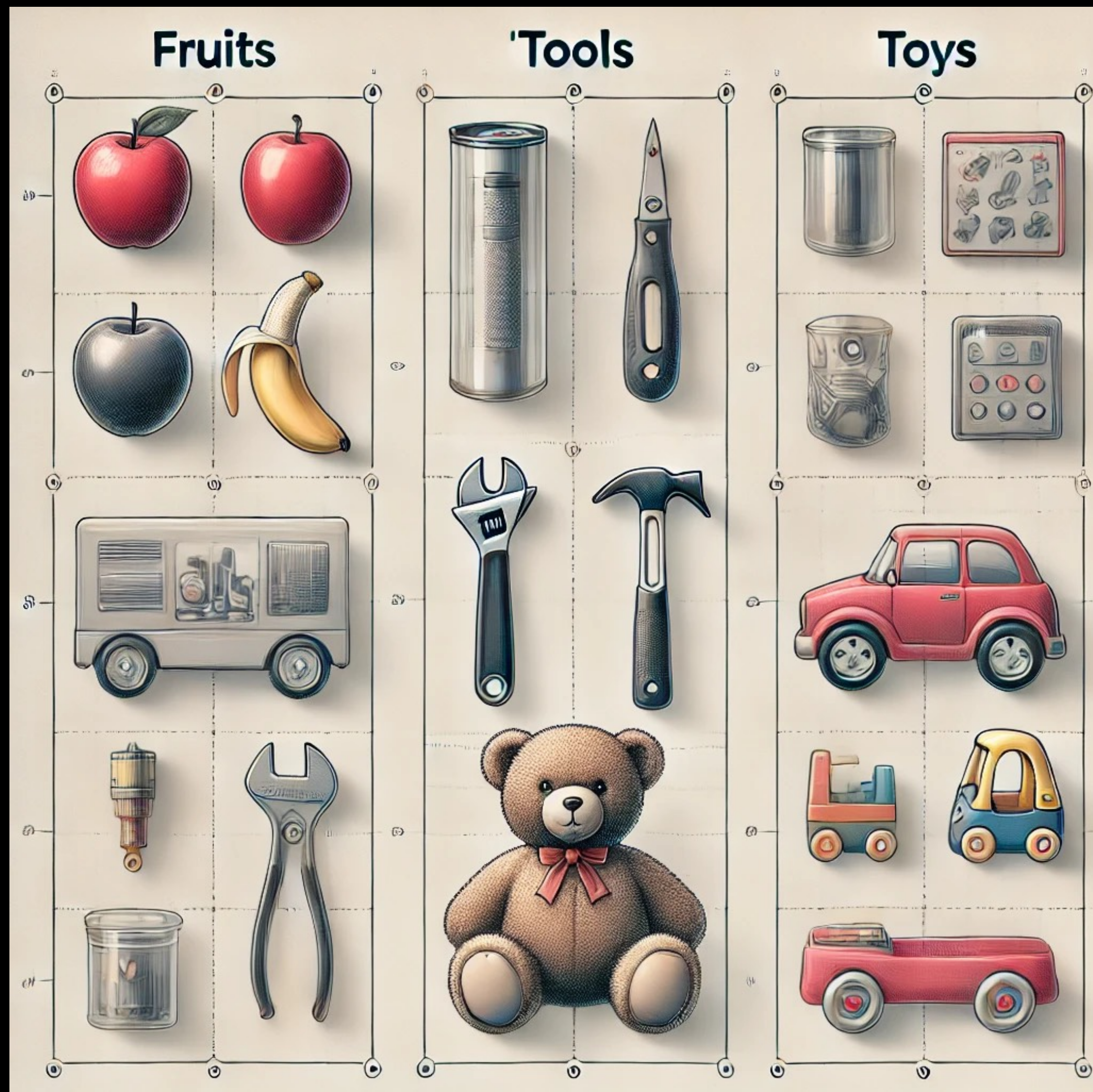


“boxing gloves”

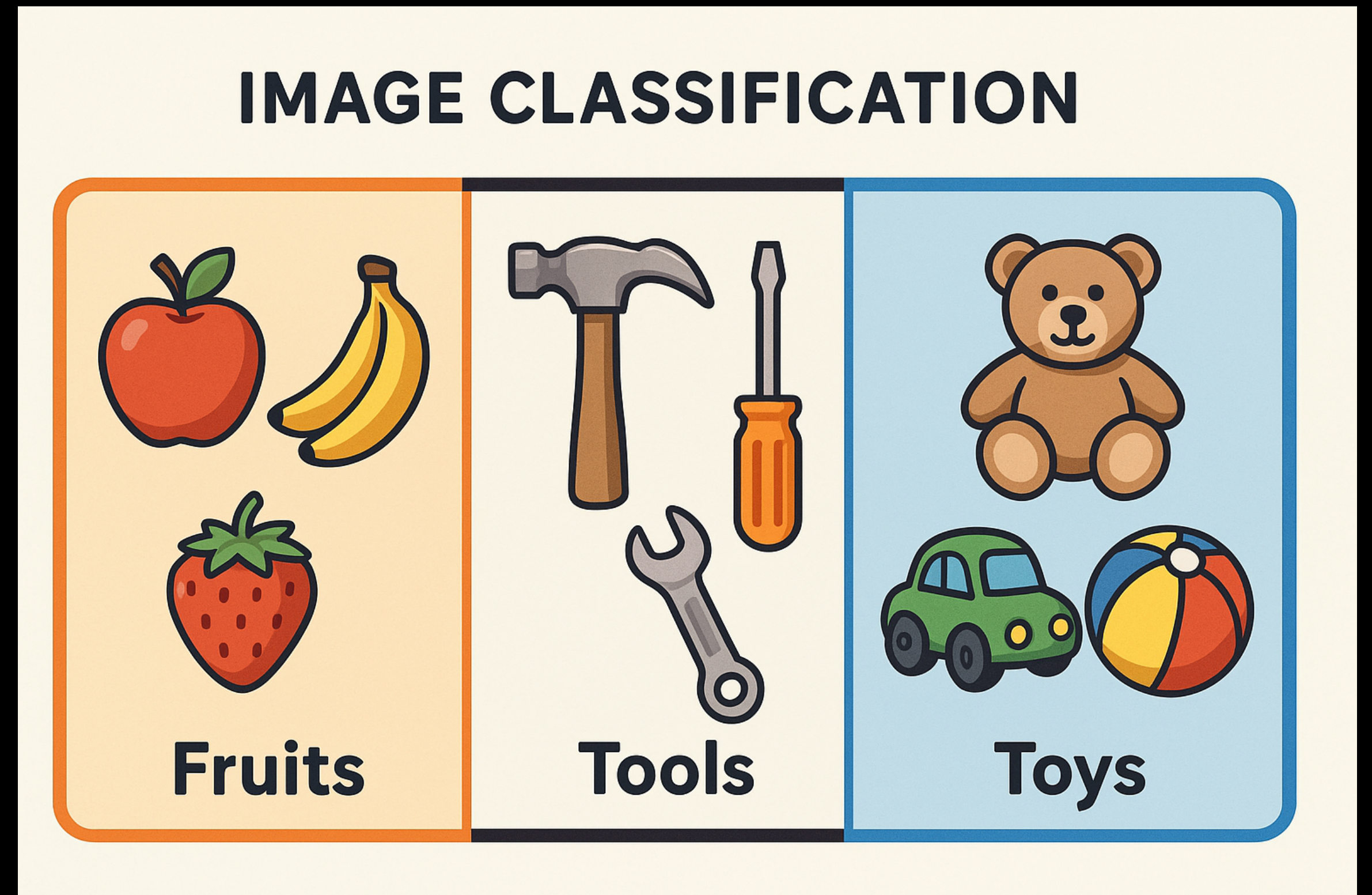




chatGPT's
proposal
1 year ago



chatGPT's
proposal
1 year ago



chatGPT's
proposal
today

Reflection symmetry



reflect



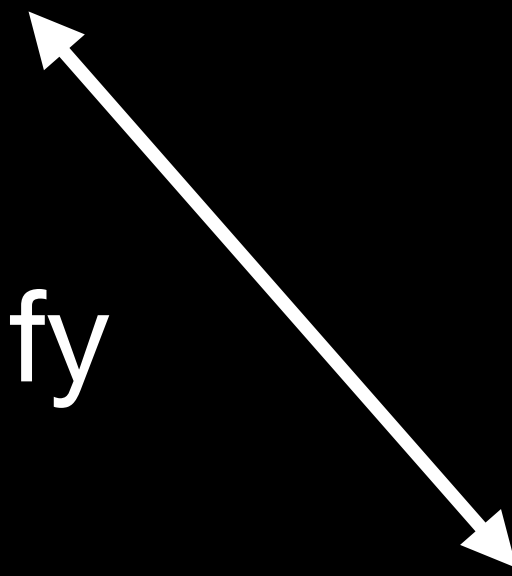
Reflection symmetry



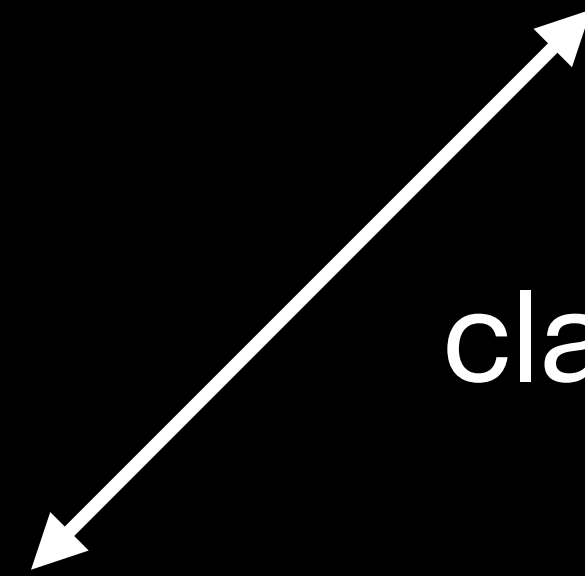
reflect



classify



classify



“boxing glove”

Reflection symmetry



reflect



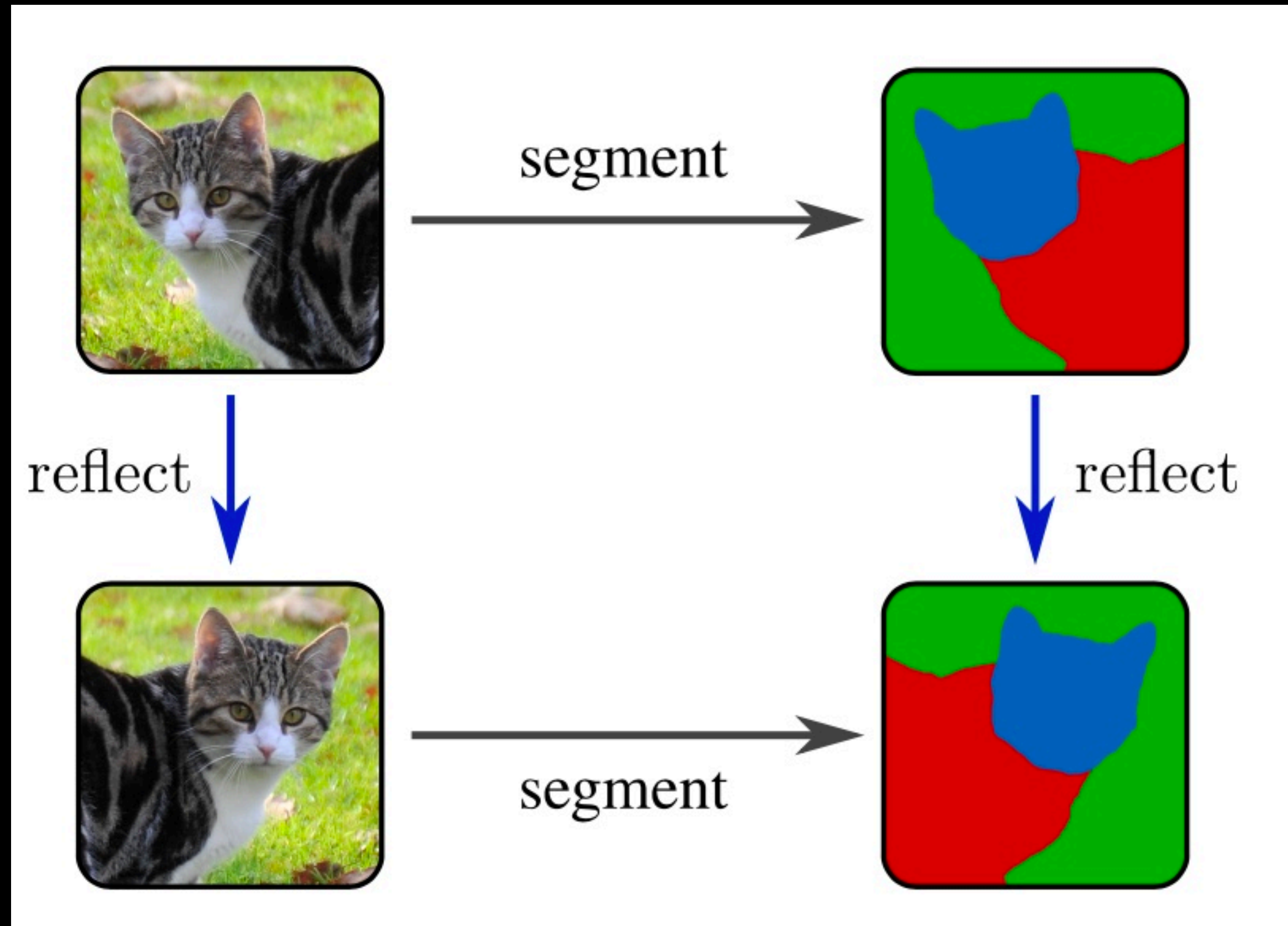
classify

classify

“boxing glove”

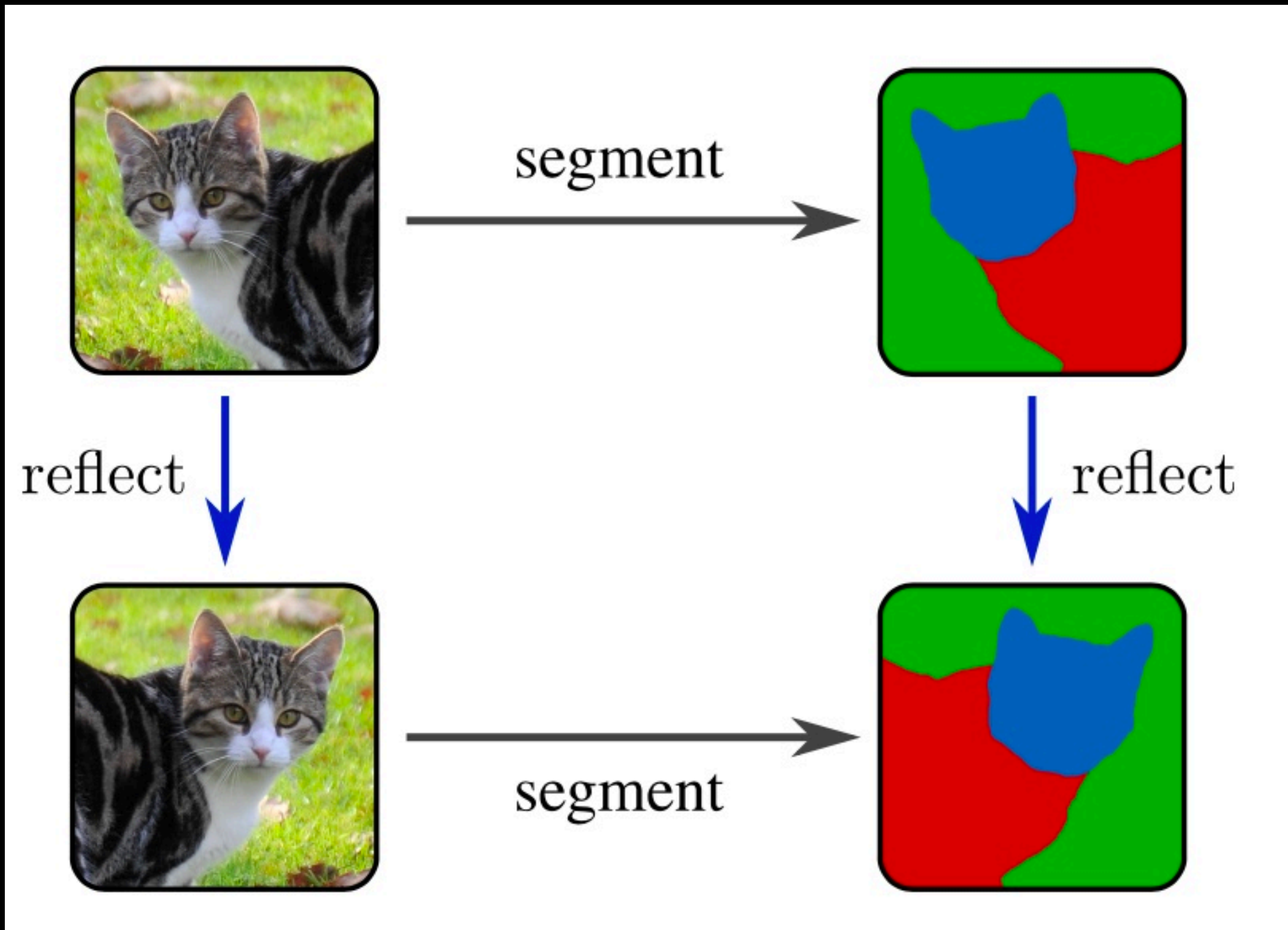
Invariance:
The output is *unchanged*
when we change the input

Segmentation



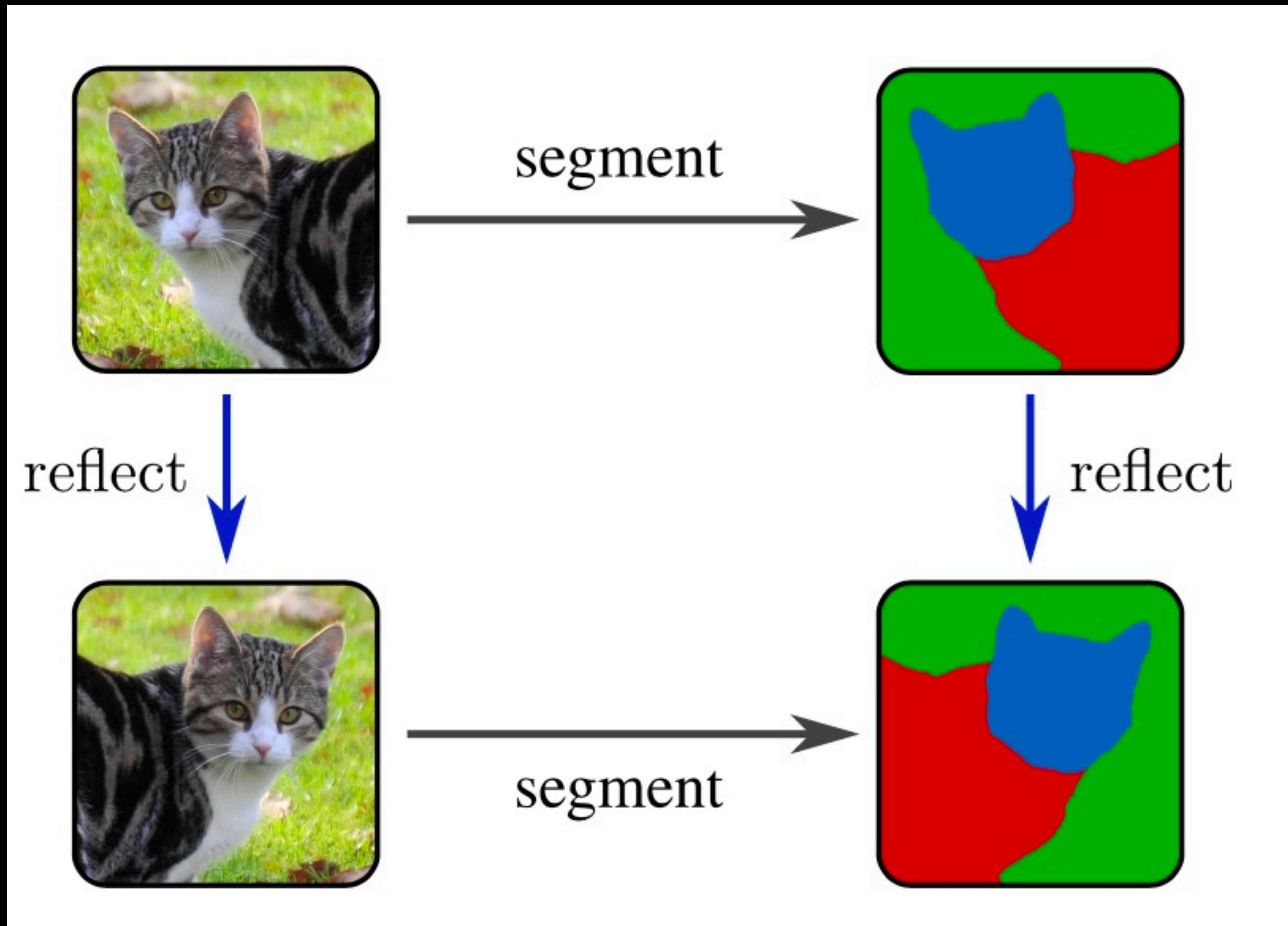
[Image from: Weiler, Forré, Verlinde, Welling (2023)]

Segmentation

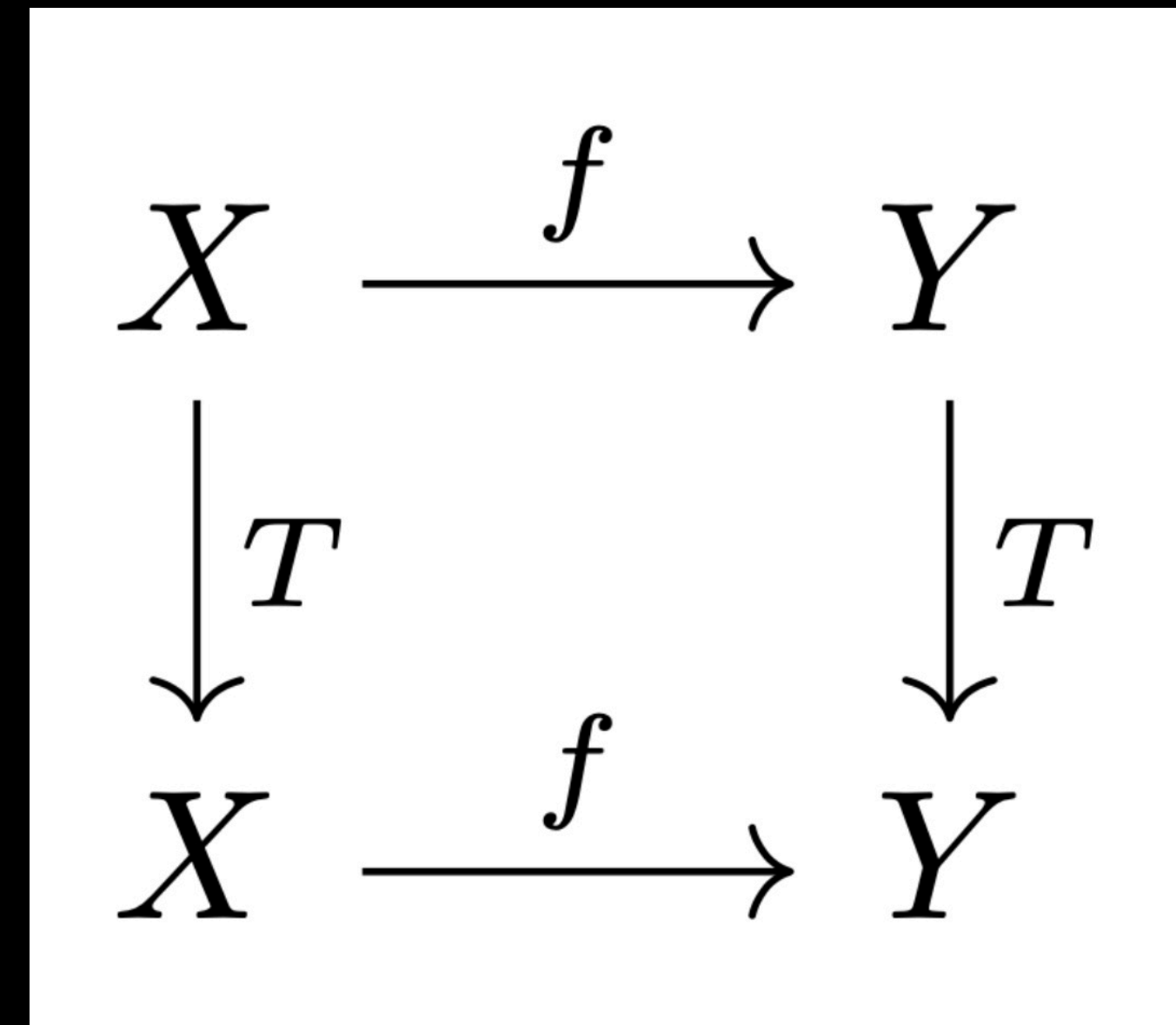


Equivariance:
The output *transforms*
according to the
transformation of the input

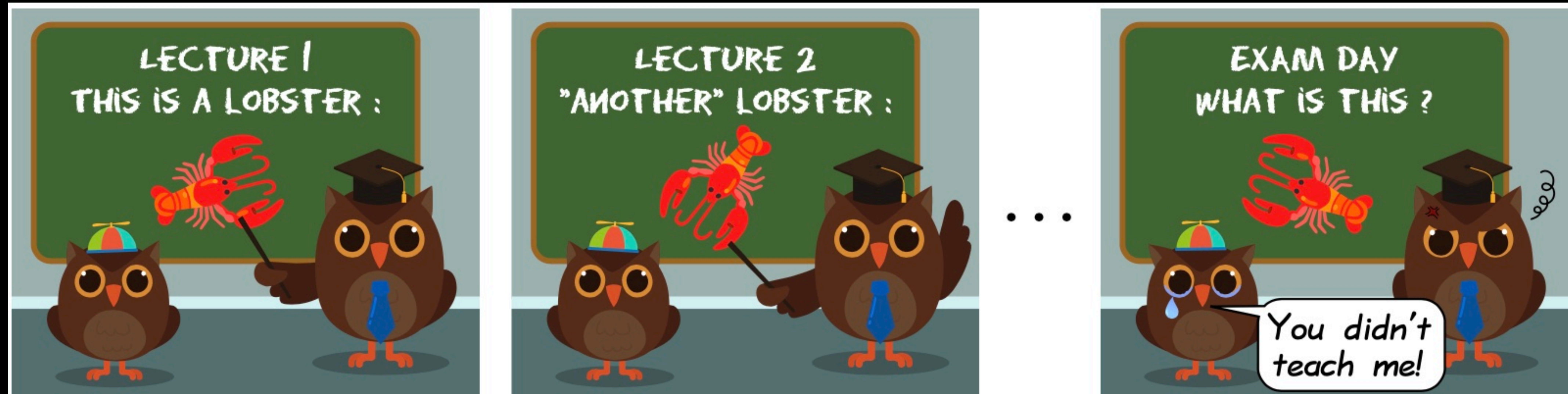
Segmentation



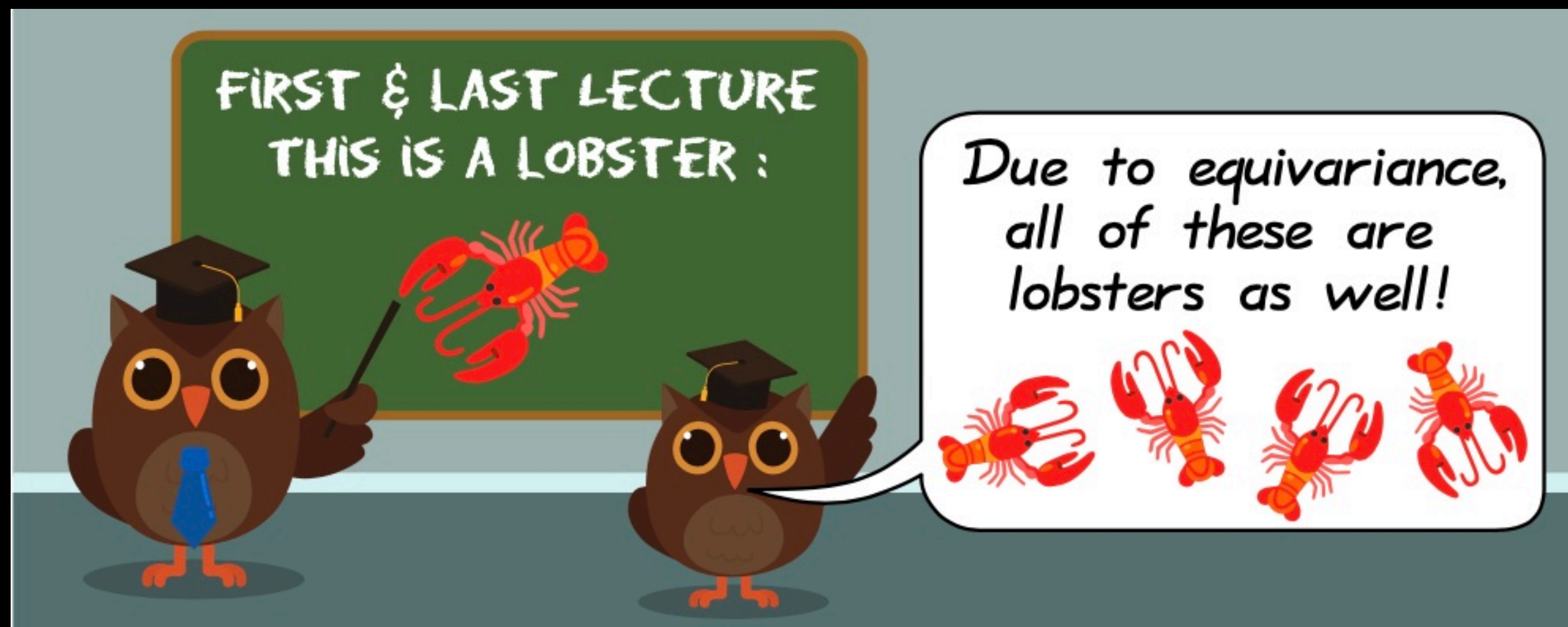
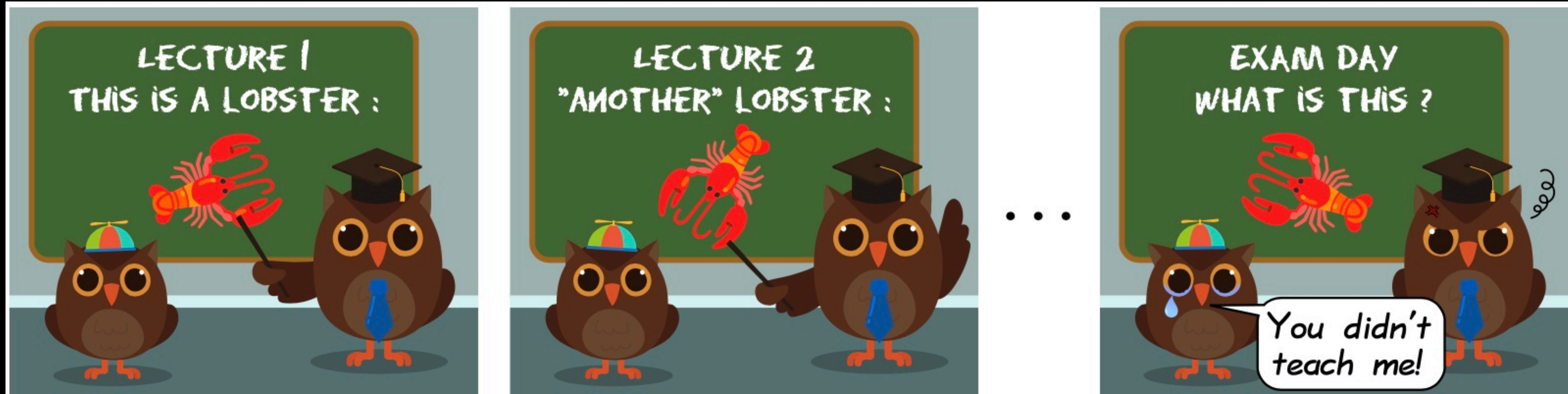
commutative
diagram

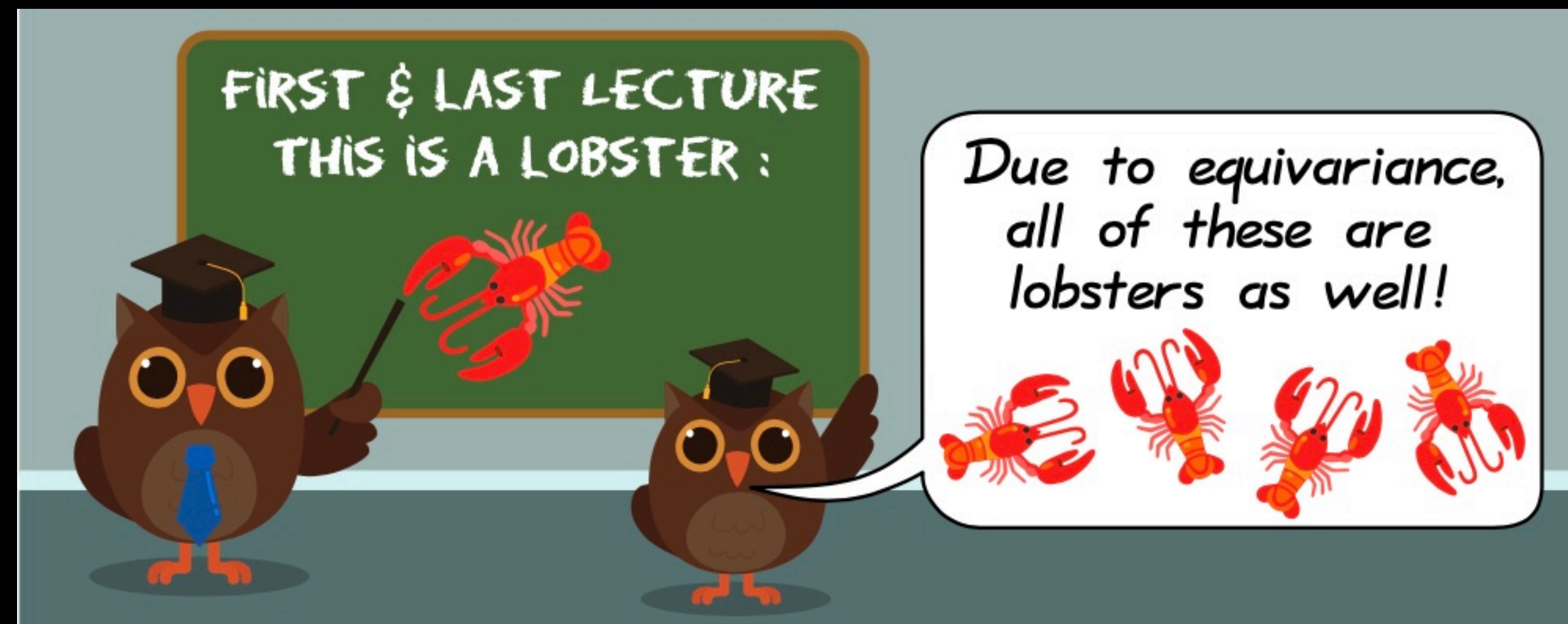
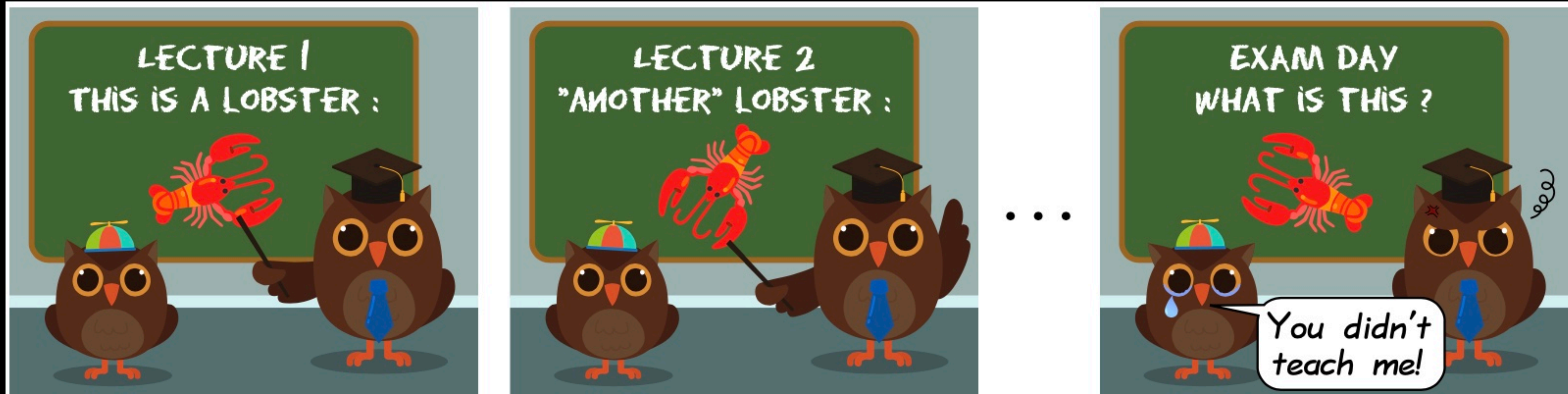


$$f(Tx) = Tf(x)$$



[Images from: Weiler, Forré, Verlinde, Welling (2023)]



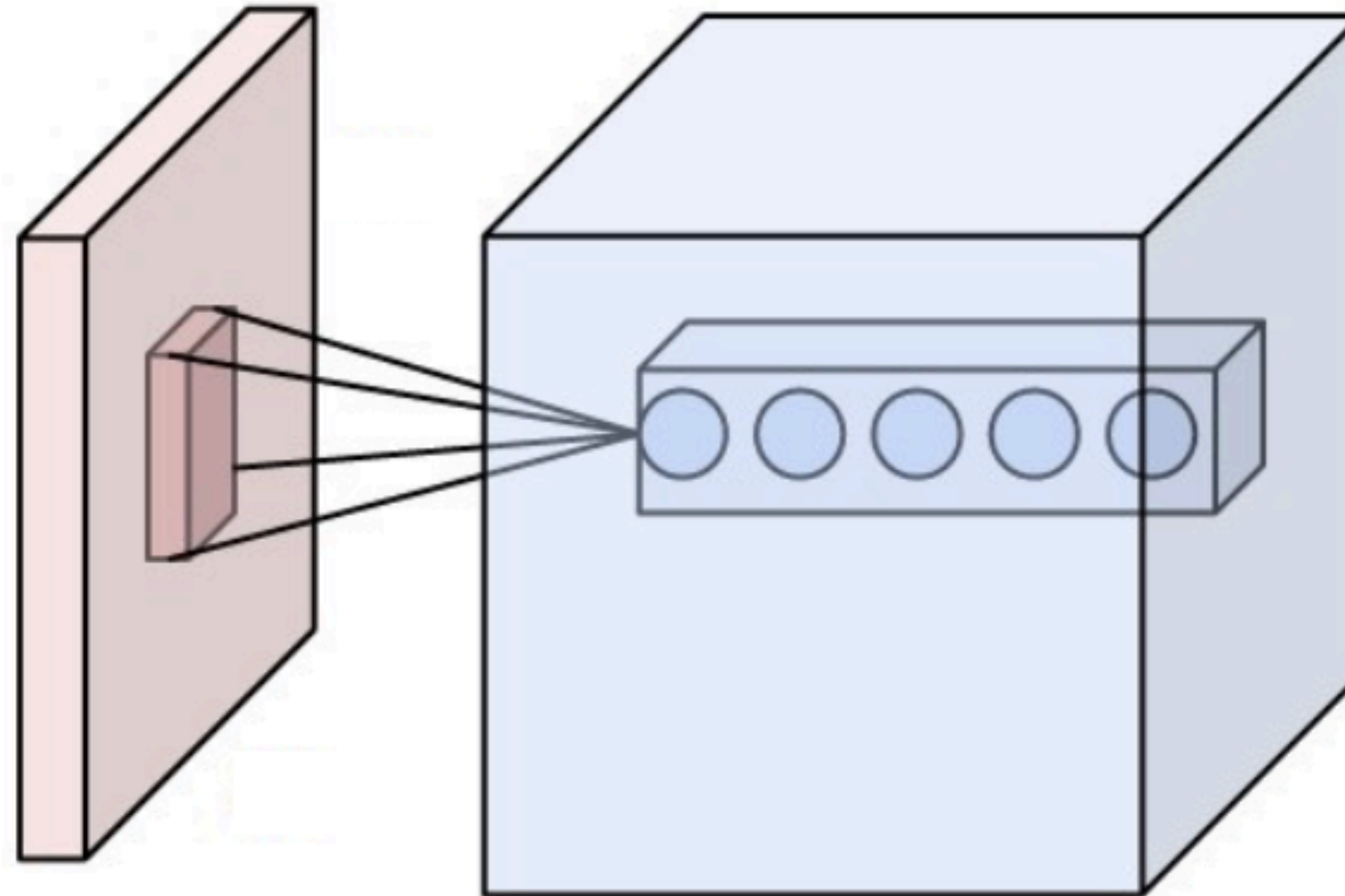


Design neural networks
that have **intrinsic symmetries**
(equivariance)

Convolutional Neural Networks

“Convolutional networks are simply neural networks that use convolution in place of general matrix multiplication in at least one of their layers.”

[Goodfellow, Bengio, Courville]



Mathematical structure

For each layer we have a **feature map**:

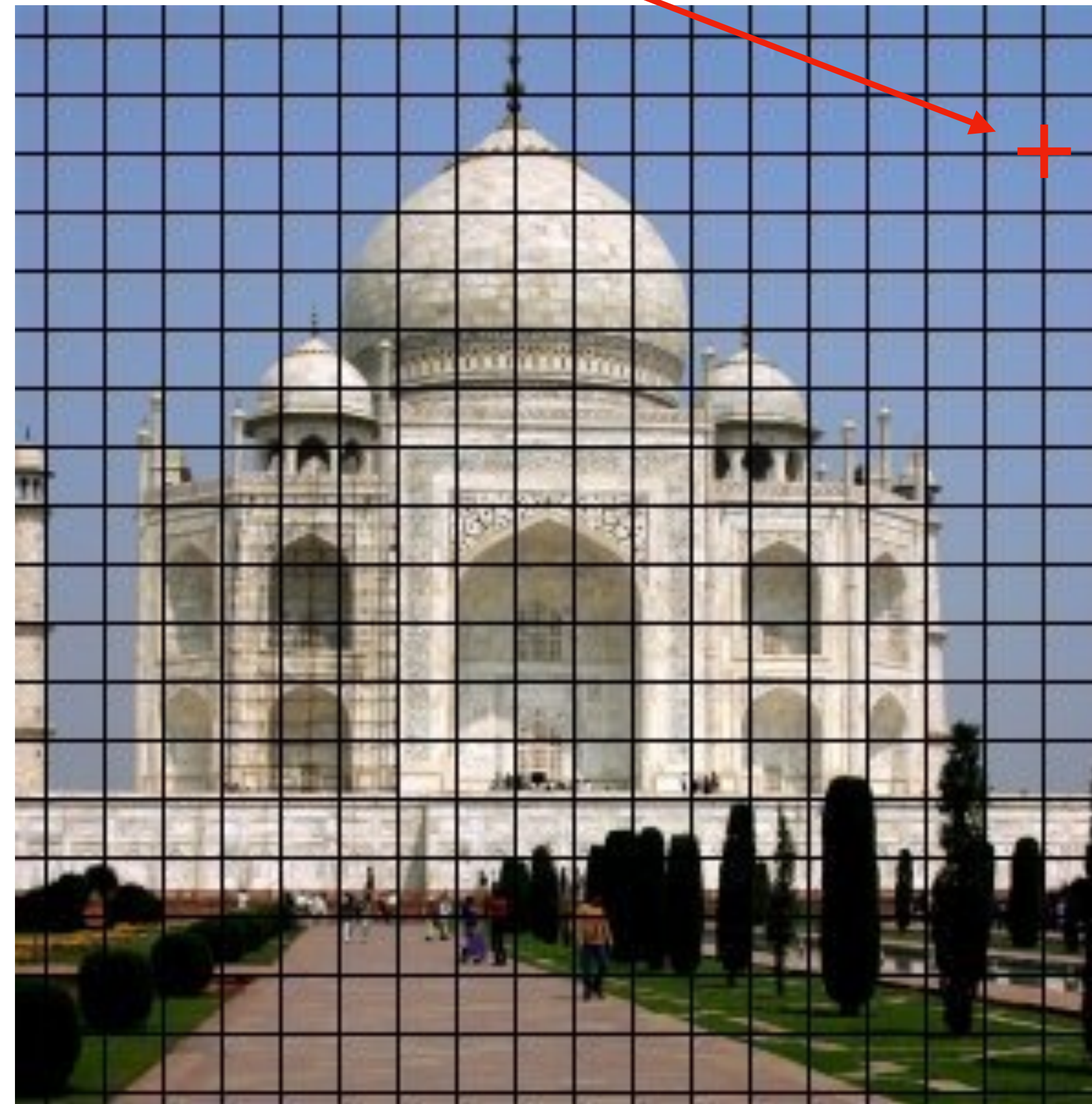
$$f : \mathbb{Z}^2 \rightarrow \mathbb{R}^K$$

Mathematical structure

For each layer we have a **feature map**:

no. of channels

$$f : \mathbb{Z}^2 \rightarrow \mathbb{R}^K$$



(p, q)

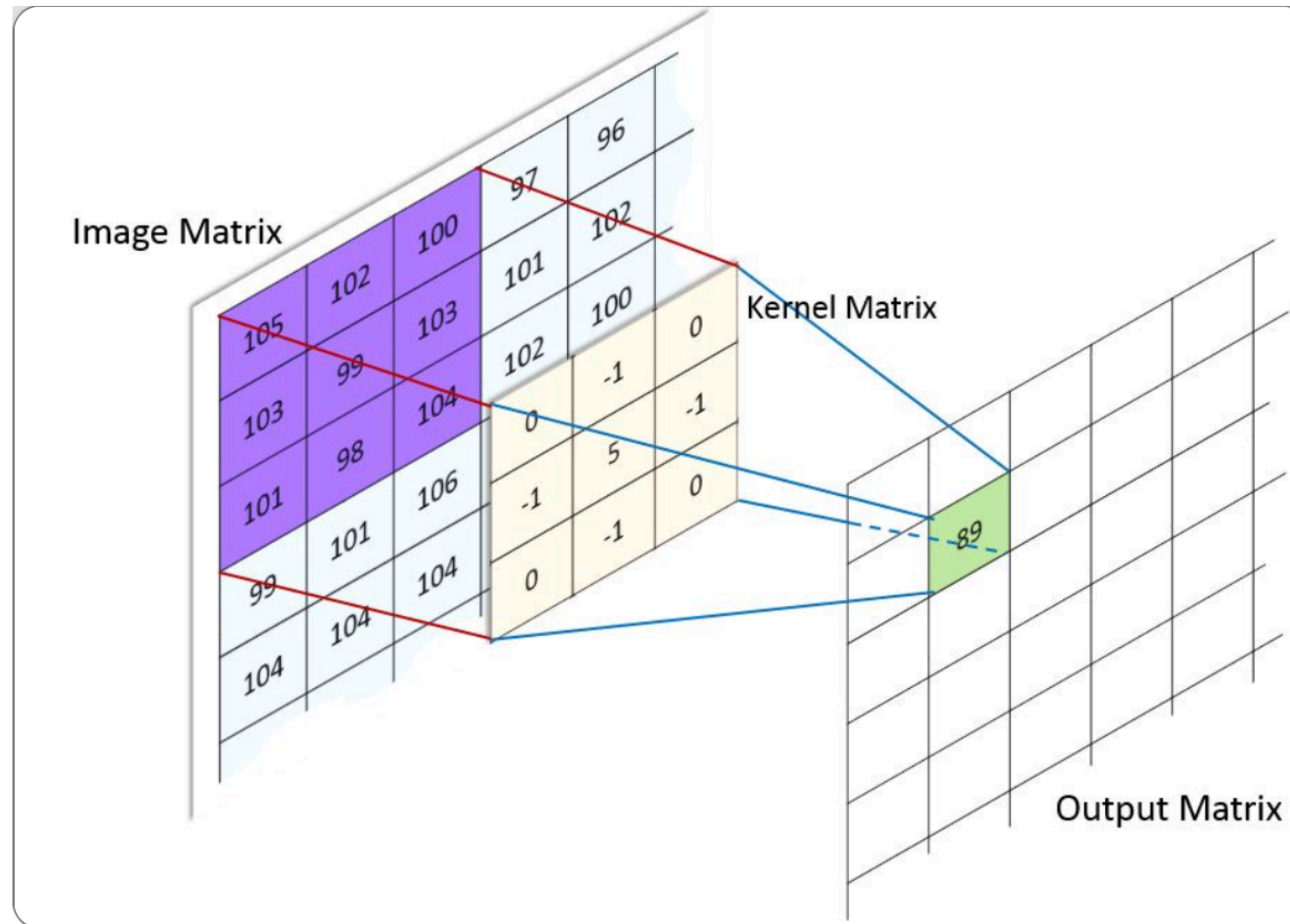
pixel coordinate

Kernel (filter): $\psi : \mathbb{Z}^2 \rightarrow \mathbb{R}^K$

Convolution: $[f * \psi](x) = \sum_{y \in \mathbb{Z}^2} \sum_{k=1}^K f_k(y) \psi_k(x - y)$

Kernel (filter): $\psi : \mathbb{Z}^2 \rightarrow \mathbb{R}^K$

Convolution: $[f * \psi](x) = \sum_{y \in \mathbb{Z}^2} \sum_{k=1}^K f_k(y) \psi_k(x - y)$



[Figure from machinelearninguru.com]

Kernel (filter): $\psi : \mathbb{Z}^2 \rightarrow \mathbb{R}^K$

Convolution: $[f * \psi](x) = \sum_{y \in \mathbb{Z}^2} \sum_{k=1}^K f_k(y) \psi_k(x - y)$

Translation map: $[T(t)f](x) = f(x + t)$

Kernel (filter): $\psi : \mathbb{Z}^2 \rightarrow \mathbb{R}^K$

Convolution: $[f * \psi](x) = \sum_{y \in \mathbb{Z}^2} \sum_{k=1}^K f_k(y) \psi_k(x - y)$

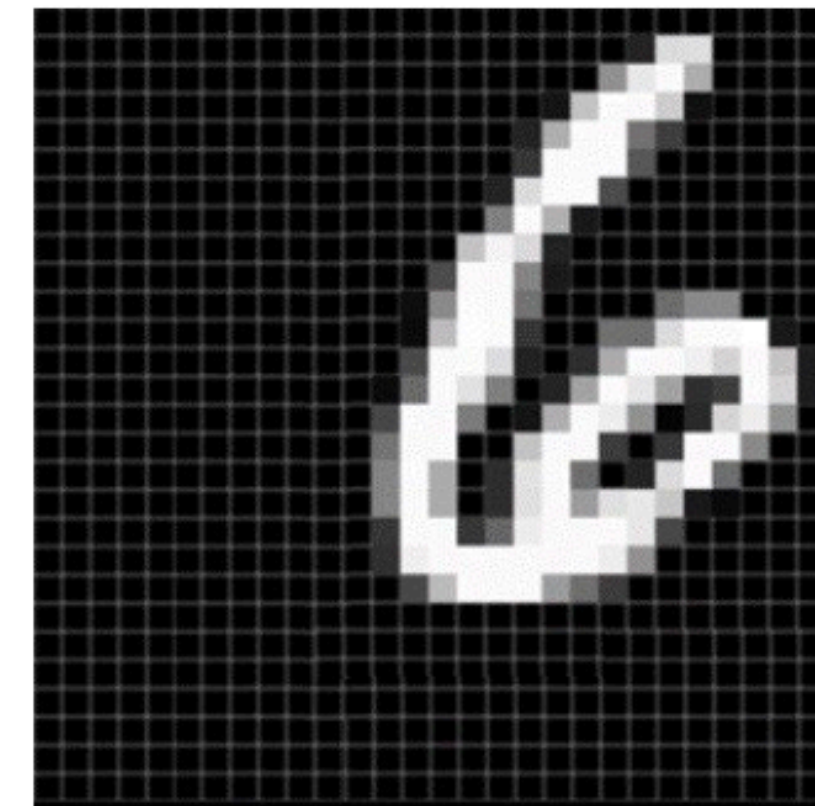
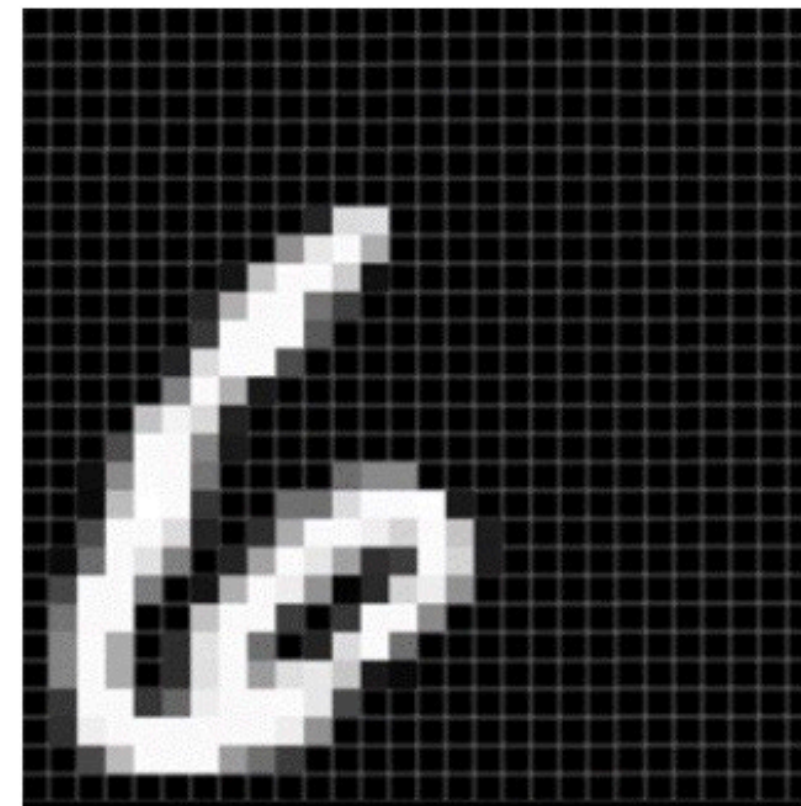
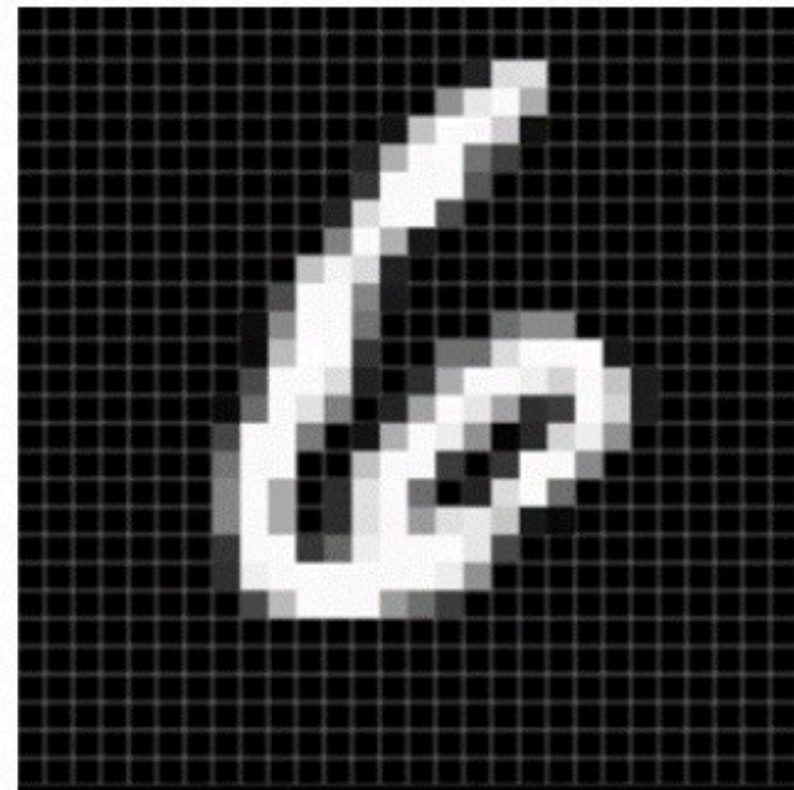
Translation map: $[T(t)f](x) = f(x + t)$

Convolution is equivariant w.r.t. translations

$$[T(t)f] * \psi = T(t)[f * \psi]$$

Convolution is **equivariant w.r.t. translations**

$$[T(t)f] * \psi = T(t)[f * \psi]$$



But what about more general symmetries?

Locally, we can think of feature maps as functions $f : G/H \rightarrow V$

Example: $G = \mathbb{Z}^2$ $H = \{1\}$ $V = \mathbb{R}^K$

Feature maps are sections! $f : \mathbb{Z}^2 \rightarrow \mathbb{R}^K$

Locally, we can think of feature maps as functions $f : G/H \rightarrow V$

Example: $G = \mathbb{Z}^2$ $H = \{1\}$ $V = \mathbb{R}^K$

Feature maps are sections! $f : \mathbb{Z}^2 \rightarrow \mathbb{R}^K$

This is a **vector bundle**:

$$\begin{array}{ccc} V & \longrightarrow & P \\ & & \downarrow p \\ & & G/H \end{array}$$

Locally, we can think of feature maps as functions $f : G/H \rightarrow V$

Example: $G = \mathbb{Z}^2$ $H = \{1\}$ $V = \mathbb{R}^K$

Feature maps are sections! $f : \mathbb{Z}^2 \rightarrow \mathbb{R}^K$

General structure of group equivariant CNNs:

$$\left\{ \begin{array}{l} \text{Feature space of a} \\ G - \text{equivariant CNN} \end{array} \right\} \cong \left\{ \begin{array}{l} \text{Sections of vector} \\ \text{bundles } P \rightarrow G/H \end{array} \right\}$$

Layers defined with group-equivariant convolutions:

$$[f * \psi](g) = \int_G \sum_{k=1}^K f_k(h) \psi_k(gh) dh$$

$$\left\{ \begin{array}{l} \text{Feature space of a} \\ G\text{-equivariant CNN} \end{array} \right\} \cong \left\{ \begin{array}{l} \text{Sections of vector} \\ \text{bundles } P \rightarrow G/H \end{array} \right\}$$

Sections of $P \rightarrow G/H$ belong to the **induced representation**:

$$\mathcal{F} = \text{Ind}_H^G \rho = \{f : G \rightarrow V \mid f(gh) = \rho(h^{-1})f(g)\}$$

$$\left\{ \begin{array}{l} \text{Feature space of a} \\ G - \text{equivariant CNN} \end{array} \right\} \cong \left\{ \begin{array}{l} \text{Sections of vector} \\ \text{bundles } P \rightarrow G/H \end{array} \right\}$$

Sections of $P \rightarrow G/H$ belong to the **induced representation**:

$$\mathcal{F} = \text{Ind}_H^G \rho = \{f : G \rightarrow V \mid f(gh) = \rho(h^{-1})f(g)\}$$

$$\left\{ \begin{array}{l} \text{Maps between layers} \\ \text{in the CNN} \end{array} \right\} \cong \left\{ \begin{array}{l} G - \text{equivariant linear maps} \\ \text{between feature spaces } \mathcal{F} \rightarrow \mathcal{F}' \end{array} \right\}$$

$$\left\{ \begin{array}{l} \text{Feature space of a} \\ G\text{-equivariant CNN} \end{array} \right\} \cong \left\{ \begin{array}{l} \text{Sections of vector} \\ \text{bundles } P \rightarrow G/H \end{array} \right\}$$

Sections of $P \rightarrow G/H$ belong to the **induced representation**:

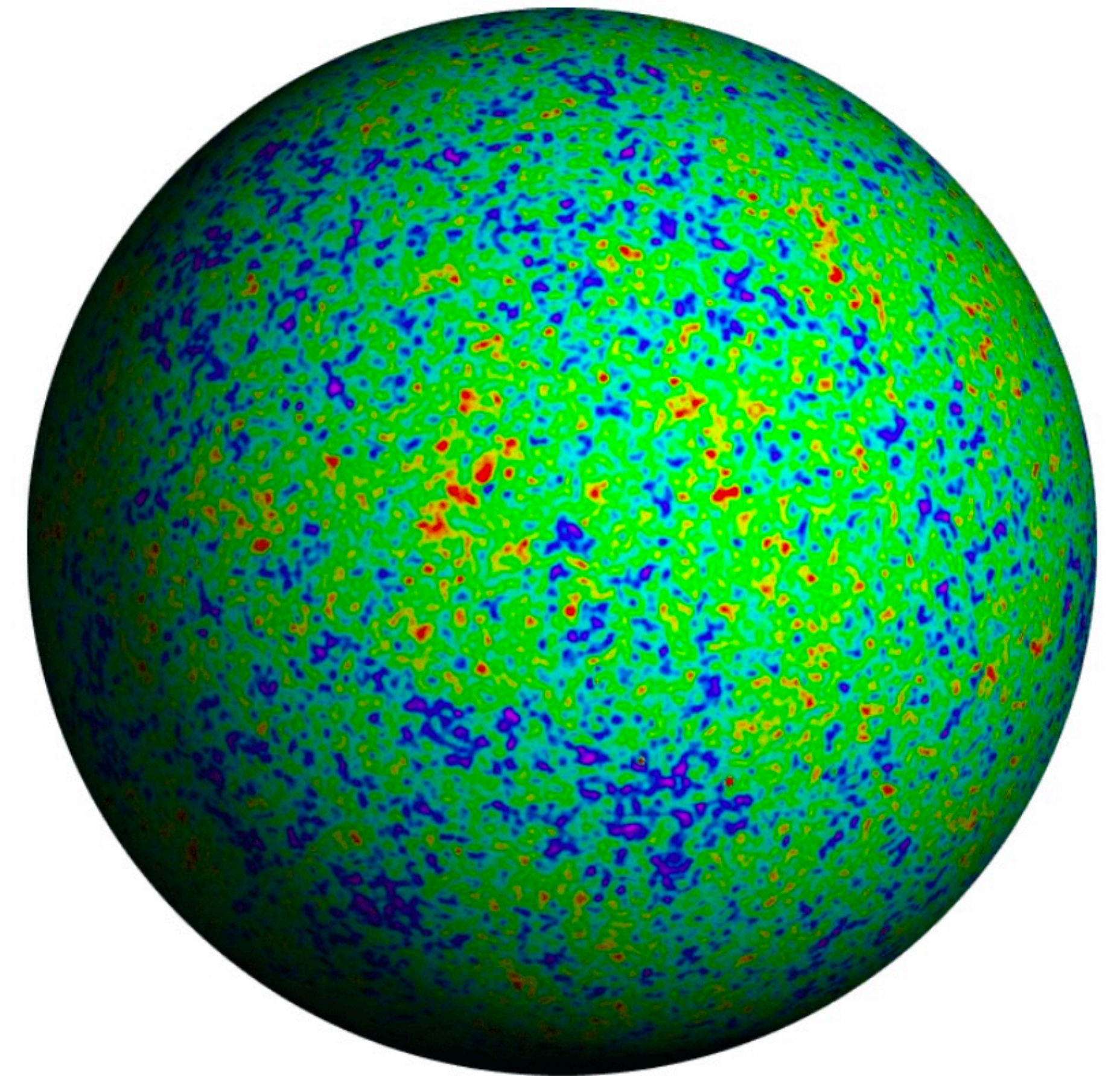
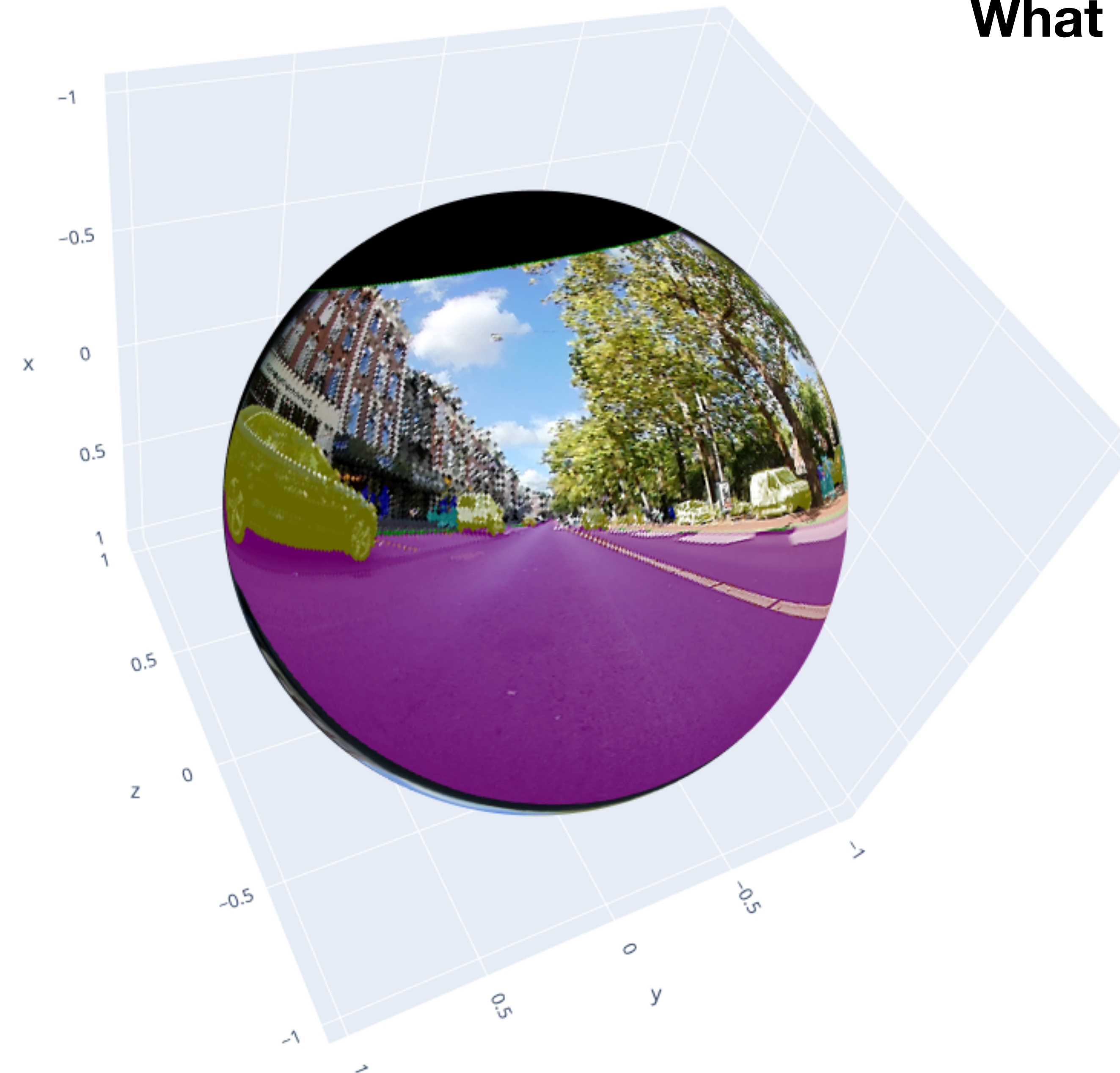
$$\mathcal{F} = \text{Ind}_H^G \rho = \{f : G \rightarrow V \mid f(gh) = \rho(h^{-1})f(g)\}$$

$$\left\{ \begin{array}{l} \text{Maps between layers} \\ \text{in the CNN} \end{array} \right\} \cong \left\{ \begin{array}{l} G\text{-equivariant linear maps} \\ \text{between feature spaces } \mathcal{F} \rightarrow \mathcal{F}' \end{array} \right\}$$

$$\text{Hom}_G(\mathcal{F}, \mathcal{F}')$$

(intertwining operators)

What if the input data is curved?



Cosmic microwave background radiation

[Image from the Woodscape dataset, projected onto a sphere]

Example: Spherical signals

By Gerken, Carlsson, Linander, Ohlsson, Petersson, **D.P.**

[ICML 2022, arXiv: 2202.03990]

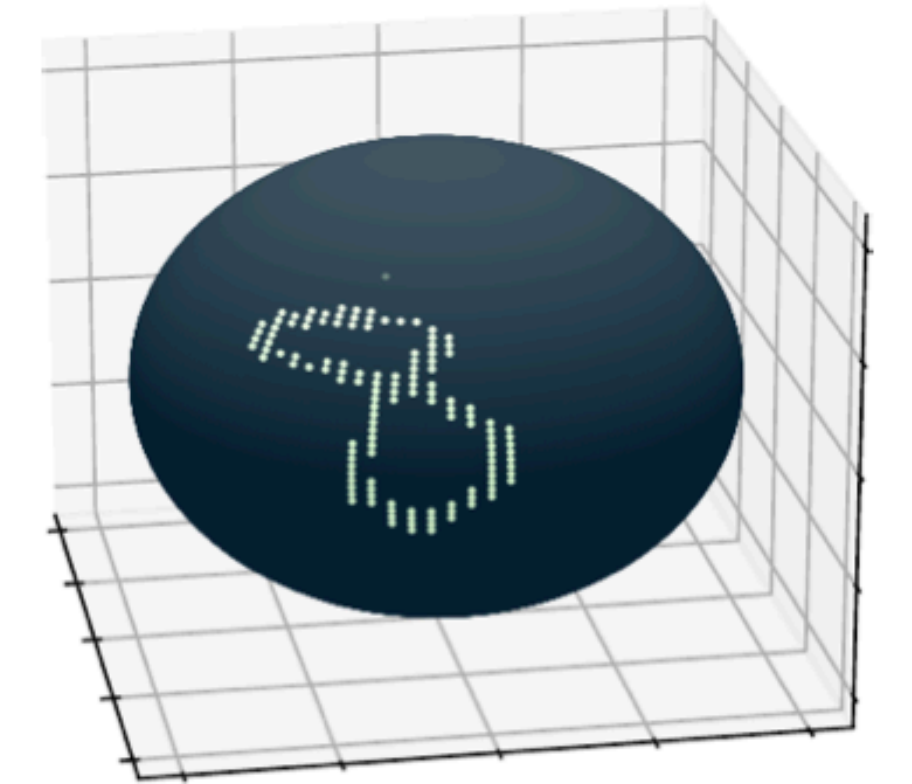
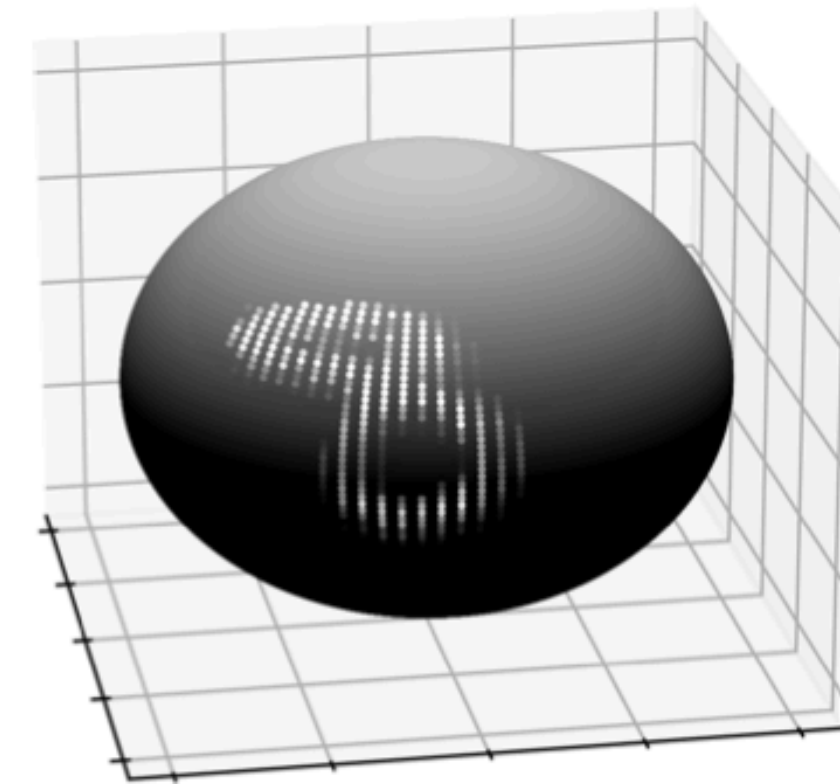
$$G = SO(3)$$

$$H = SO(2)$$

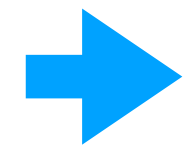
$$G/H \cong S^2$$

Feature maps

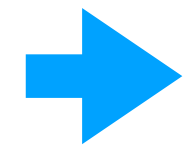
$$f : S^2 \rightarrow \mathbb{R}^K$$



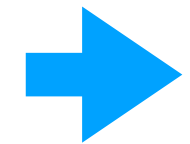
Relevant for :



Omnidirectional vision



Weather and climate data



Cosmology & astrophysics

$$(\kappa \star f)(R) = \int_{S^2} \kappa(R^{-1}x) f(x) dx$$

$$(\kappa \star f)(R) = \int_{SO(3)} \kappa(S^{-1}R) f(S) dS$$

HEAL-SWIN: A Vision Transformer On The Sphere

Oscar Carlsson^{*ab} Jan E. Gerken^{*a} Hampus Linander^a Heiner Spieß^c Fredrik Ohlsson^d

Christoffer Petersson^{ea}

Daniel Persson^a

Abstract

High-resolution wide-angle fisheye images are becoming more and more important for robotics applications such as autonomous driving. However, using ordinary convolutional neural networks or vision transformers on this data is problematic due to projection and distortion losses introduced when projecting to a rectangular grid on the plane. We introduce the HEAL-SWIN transformer, which combines the highly uniform Hierarchical Equal Area iso-Latitude Pixelation (HEALPix) grid used in astrophysics and cosmology with the Hierarchical Shifted-Window (SWIN) transformer to yield an efficient and flexible model capable of training on high-resolution, distortion-free spherical data. In HEAL-SWIN, the nested structure of the HEALPix grid is used to perform the patching and windowing operations of the SWIN transformer, enabling the network to process spherical representations with minimal computational overhead. We demonstrate the superior performance of our model on both synthetic and real automotive datasets, as well as a selection of other image datasets, for semantic segmentation, depth regression and classification tasks. Our code is publicly available¹.

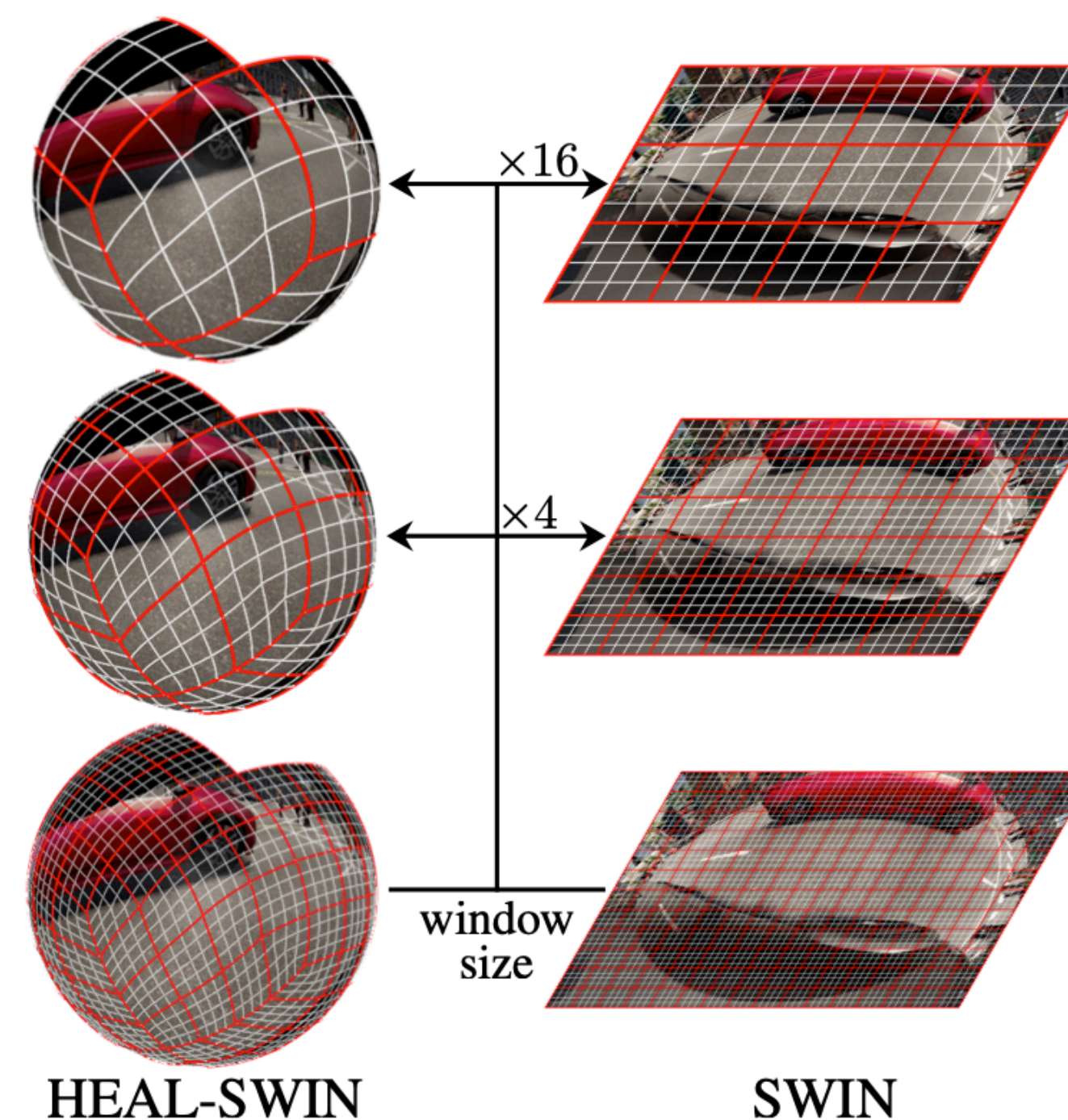


Figure 1. Our HEAL-SWIN model uses the nested structure of the HEALPix grid to lift the windowed self-attention of the SWIN model onto the sphere.

PEAR: Equal Area Weather Forecasting on the Sphere

Hampus Linander
VERSES AI Research Lab
Los Angeles, USA
and

Department of Mathematical Sciences
Chalmers University of Technology
hampus.linander@chalmers.se

Christoffer Petersson
Zenseact
Gothenburg, Sweden
and

Department of Mathematical Sciences
Chalmers University of Technology
christoffer.petersson@zenseact.com

Daniel Persson
Department of Mathematical Sciences
Chalmers University of Technology
University of Gothenburg, Sweden
daniel.persson@chalmers.se

Jan E. Gerken
Department of Mathematical Sciences
Chalmers University of Technology
University of Gothenburg, Sweden
gerken@chalmers.se

Abstract

Machine learning methods for global medium-range weather forecasting have recently received immense attention. Following the publication of the Pangu Weather model, the first deep learning model to outperform traditional numerical simulations of the atmosphere, numerous models have been published in this domain, building on Pangu’s success. However, all of these models operate on input data and produce predictions on the Driscoll–Healy discretization of the sphere which suffers from a much finer grid at the poles than around the equator. In contrast, in the Hierarchical Equal Area iso-Latitude Pixelization (HEALPix) of the sphere, each pixel covers the same surface area, removing unphysical biases. Motivated by a growing support for this grid in meteorology and climate sciences, we propose to perform weather forecasting with deep learning models which natively operate on the HEALPix grid. To this end, we introduce Pangu Equal Area (PEAR), a transformer-based weather forecasting model which operates directly on HEALPix-features and outperforms the corresponding model on Driscoll–Healy without any computational overhead. *Code will be made available.*

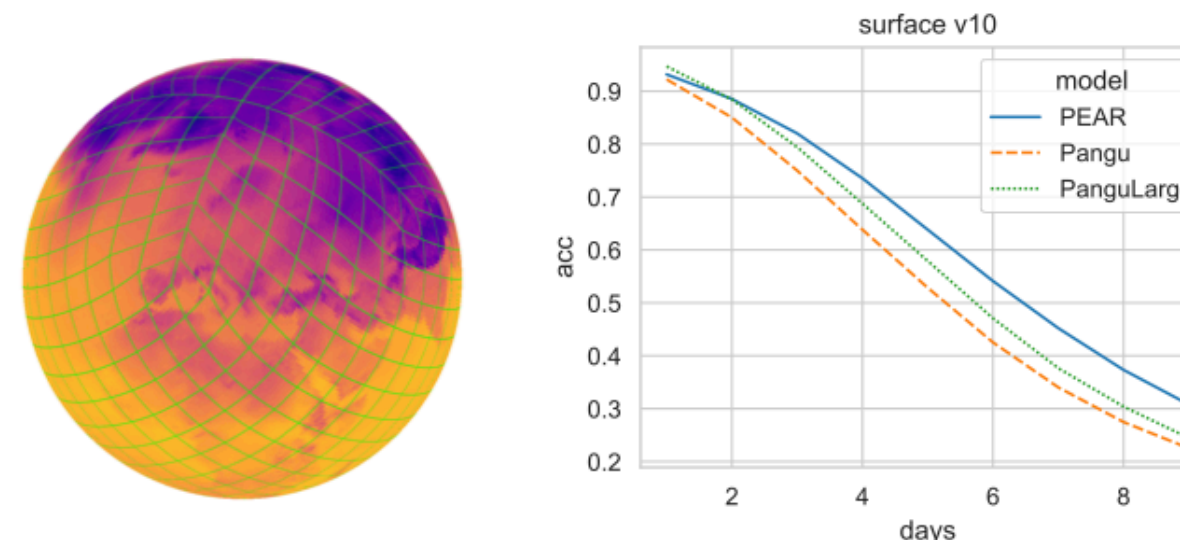
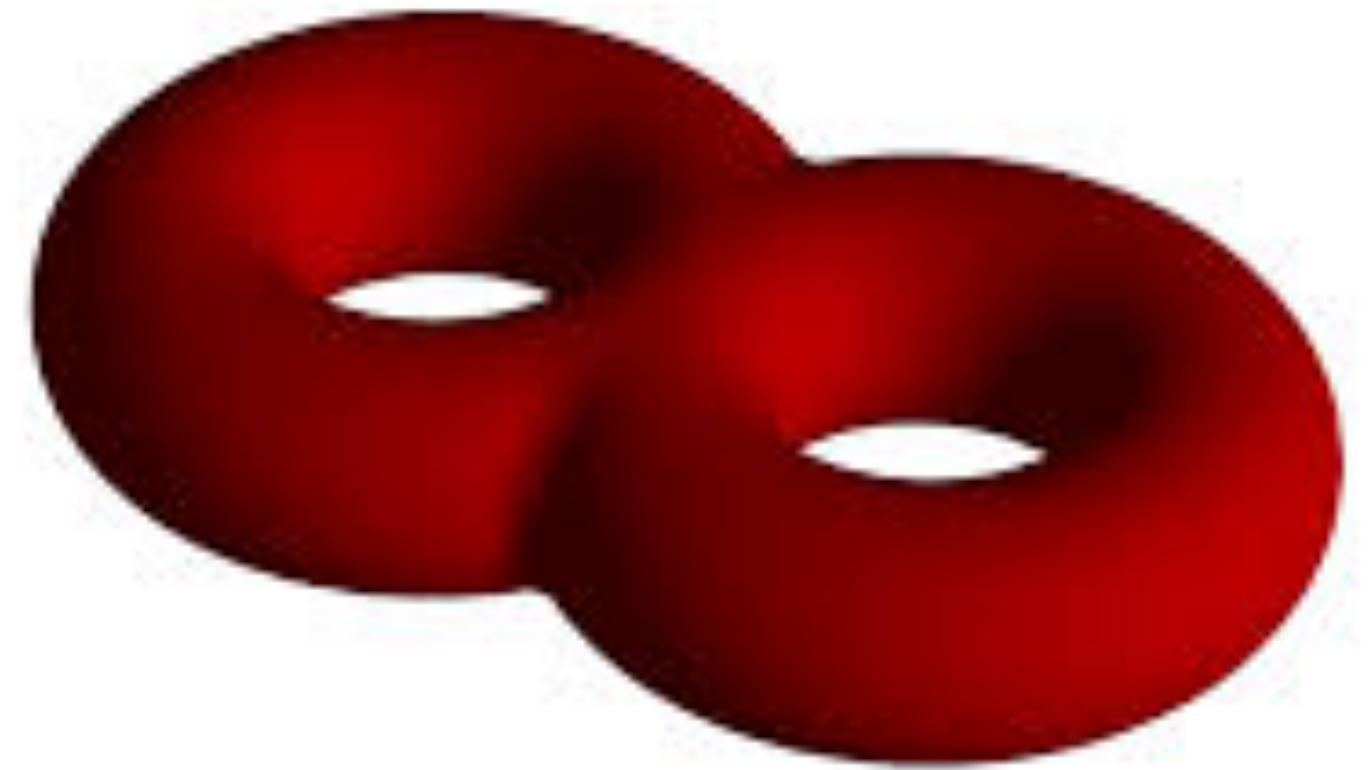
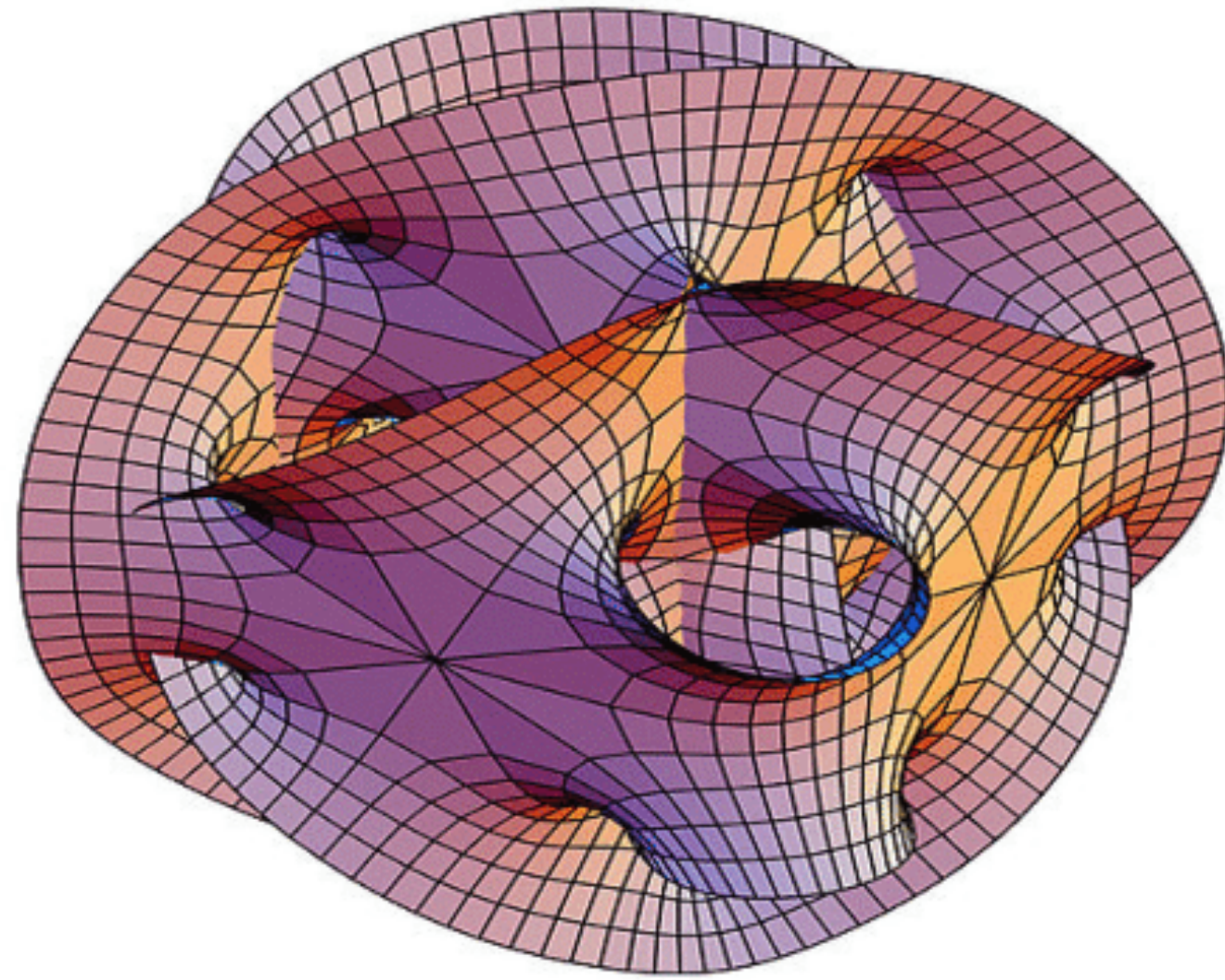
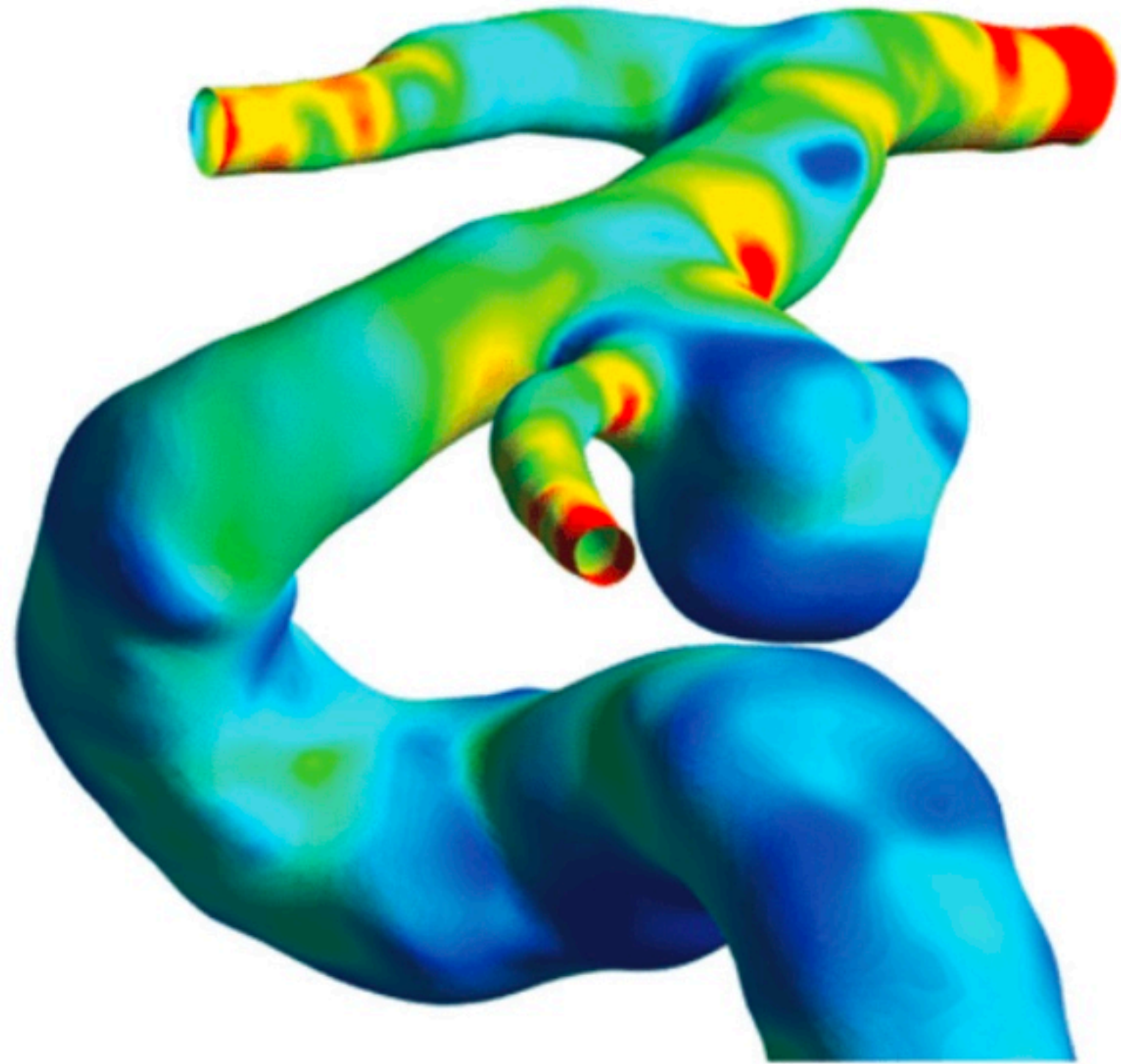
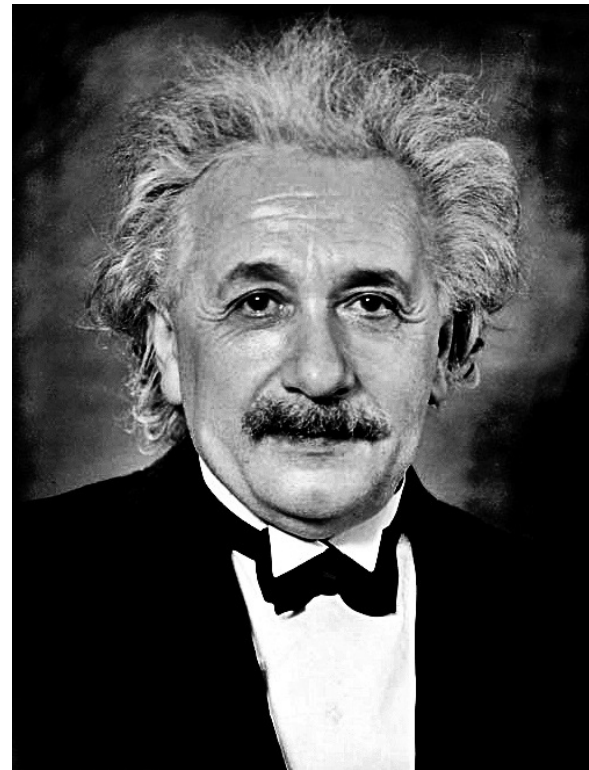
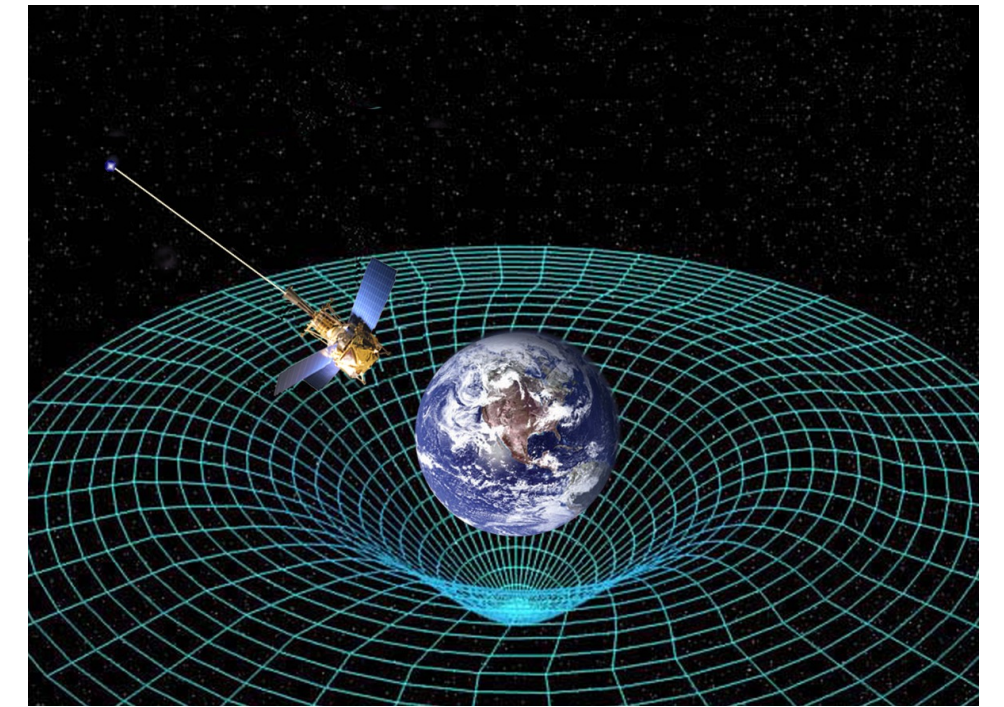
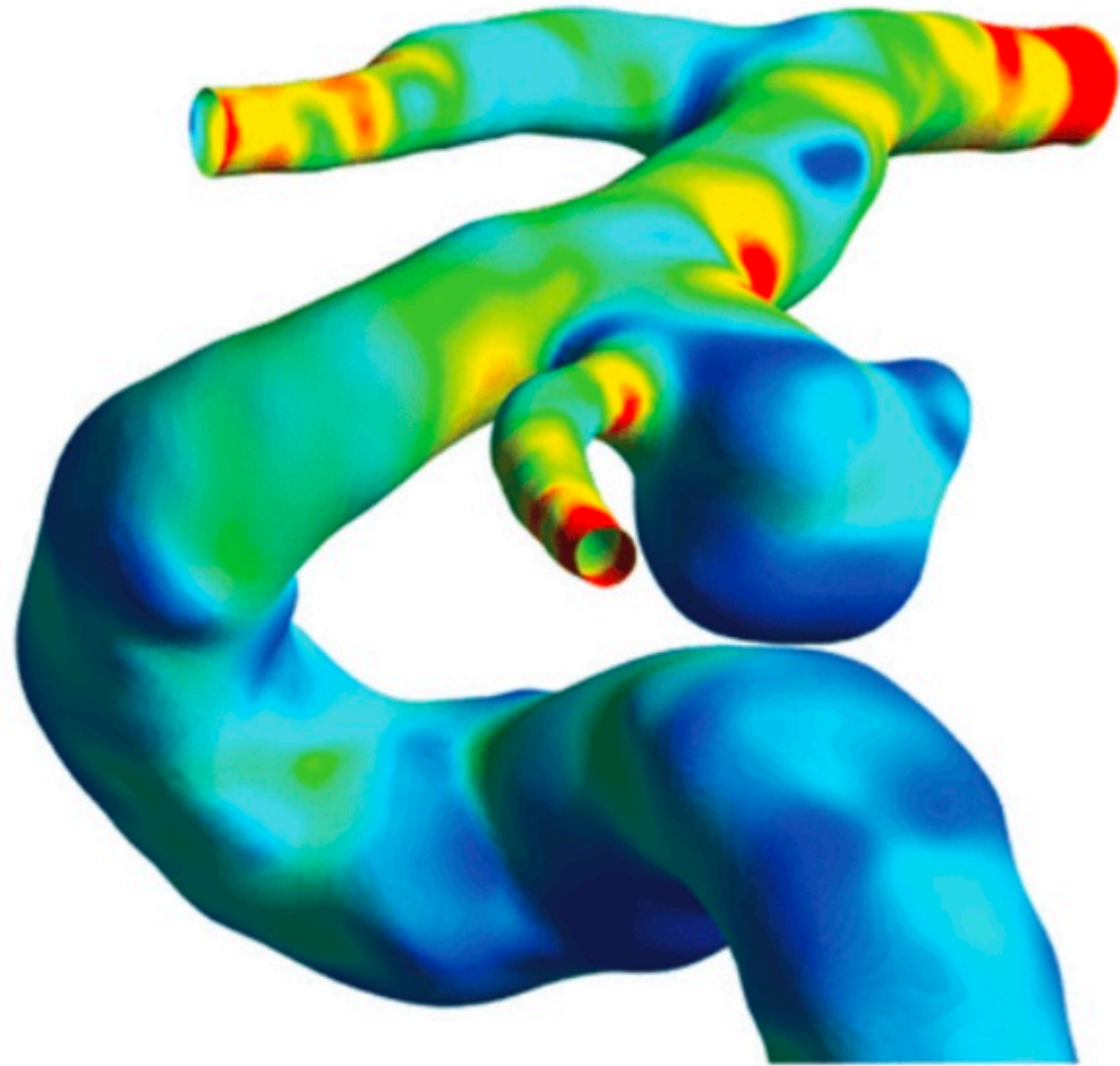


Figure 1: Left: Predicted surface level temperature from PEAR. Green lines show the HEALPix cell boundaries at 3 levels of course-graining above the model resolution. Right: Anomaly correlation coefficient (ACC) (higher is better) for surface level northward wind component with forecast horizon up to 10 days. PEAR outperforms the almost 8 times larger PanguLarge at longer forecast horizons.

Geometric deep learning: Deep learning on manifolds



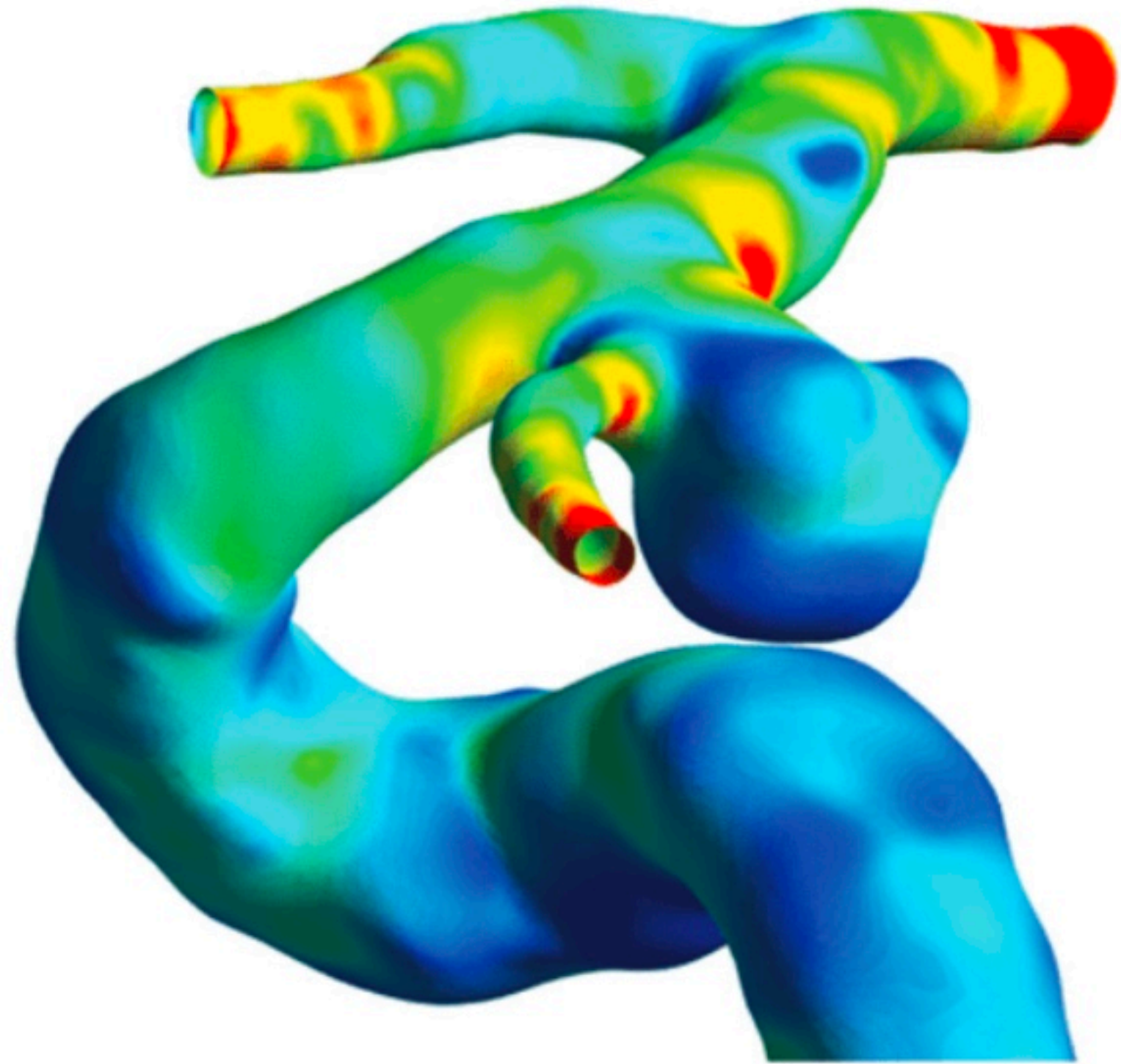
Geometric deep learning



Recall Einstein's

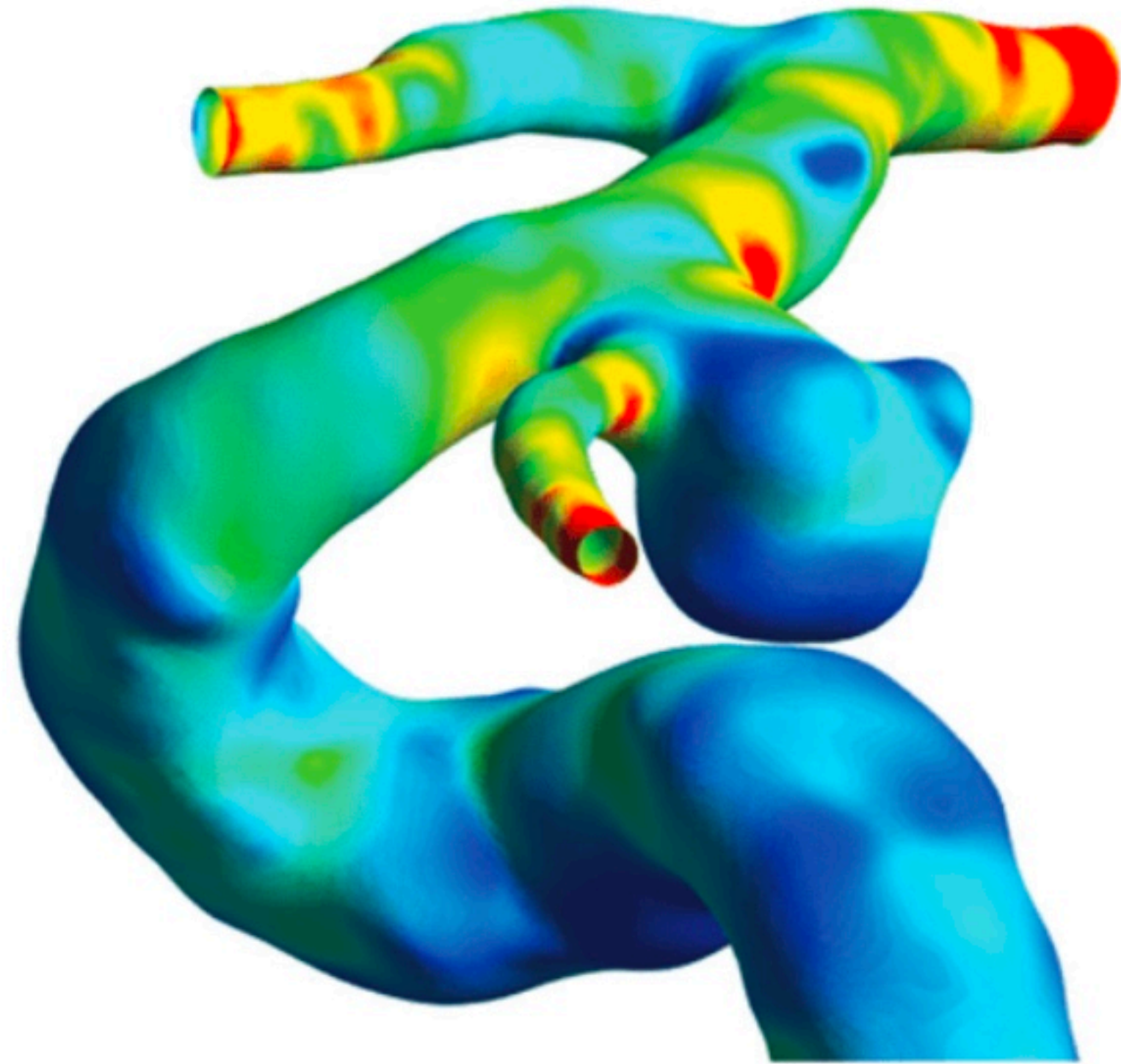
Principle of general covariance:
The laws of physics should take the same form independently of which coordinate system we use to represent them

Geometric deep learning - A unified framework for deep learning

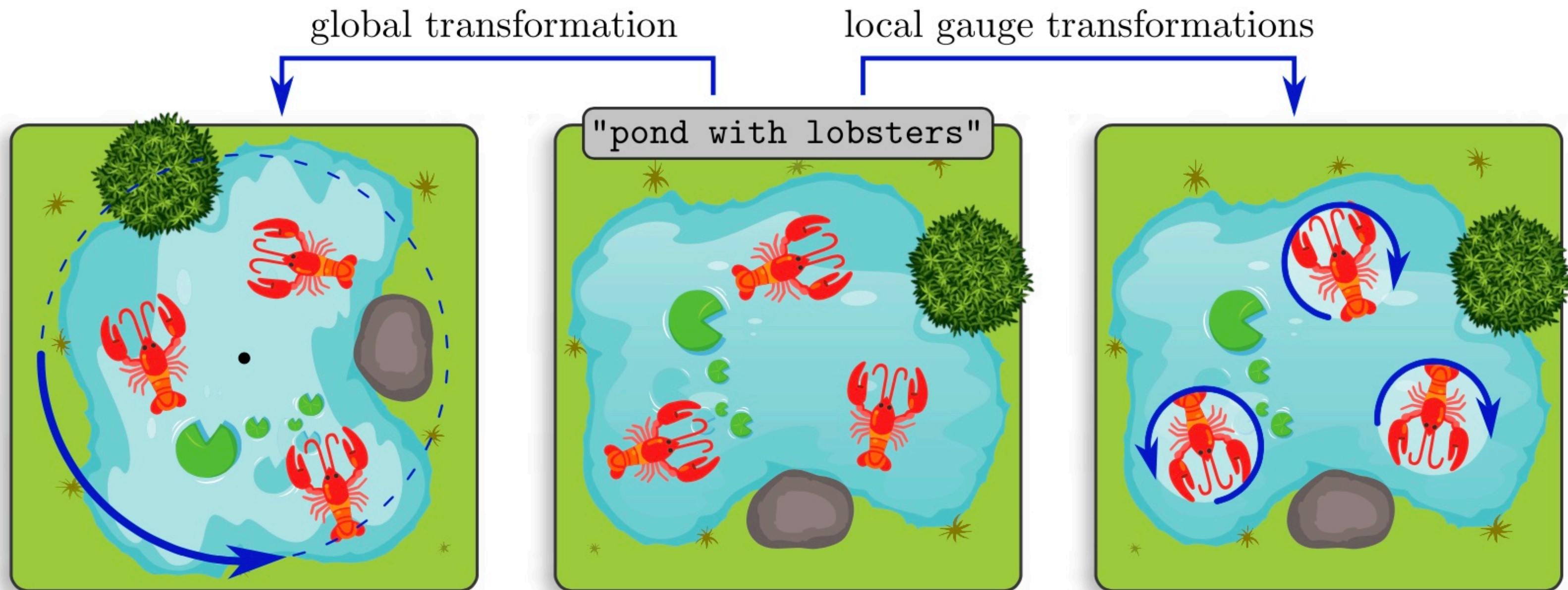


Principle of geometric deep learning:
The equations governing neural networks should be equivariant with respect to all **local** and **global** symmetries of the input data

Geometric deep learning - A unified framework for deep learning

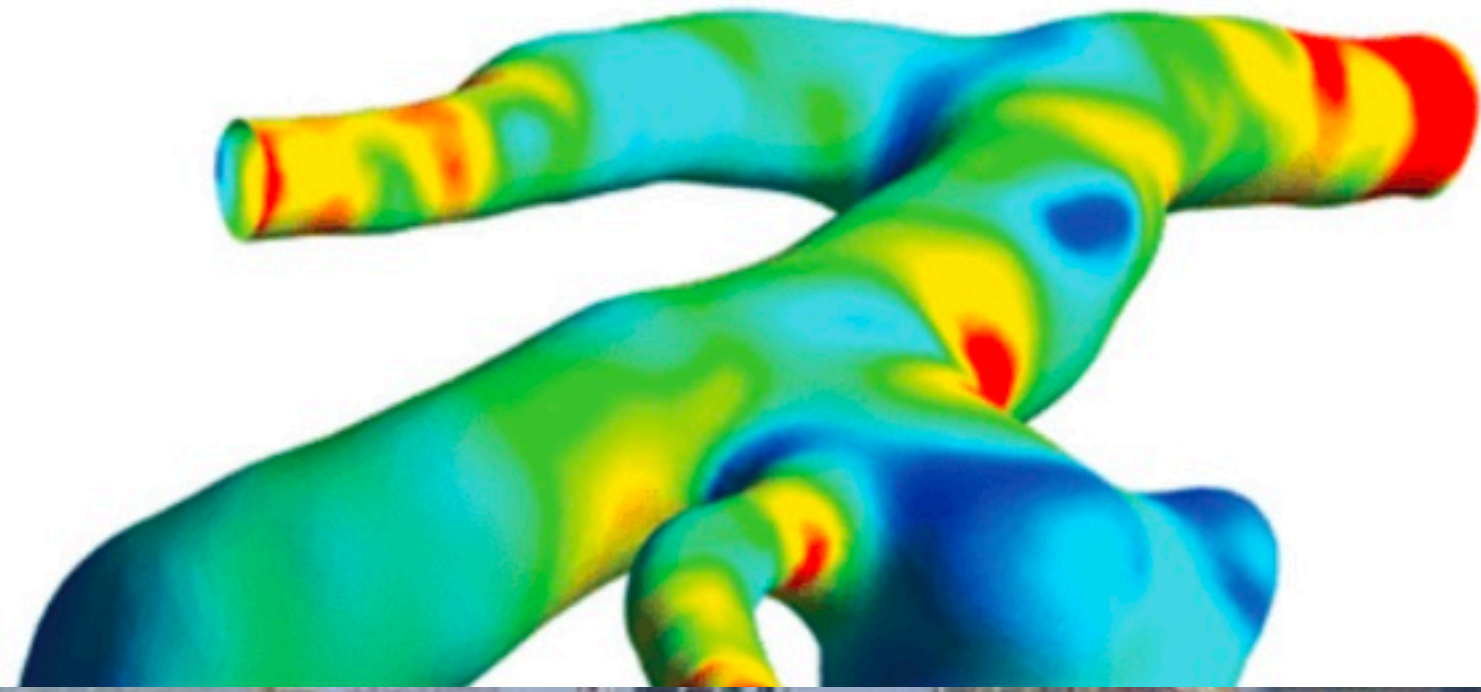


Principle of geometric deep learning:
The equations governing neural networks should be equivariant with respect to all **local** and **global** symmetries of the input data



[Images from: Weiler, Forré, Verlinde, Welling (2023)]

Geometric deep learning - A unified framework for deep learning

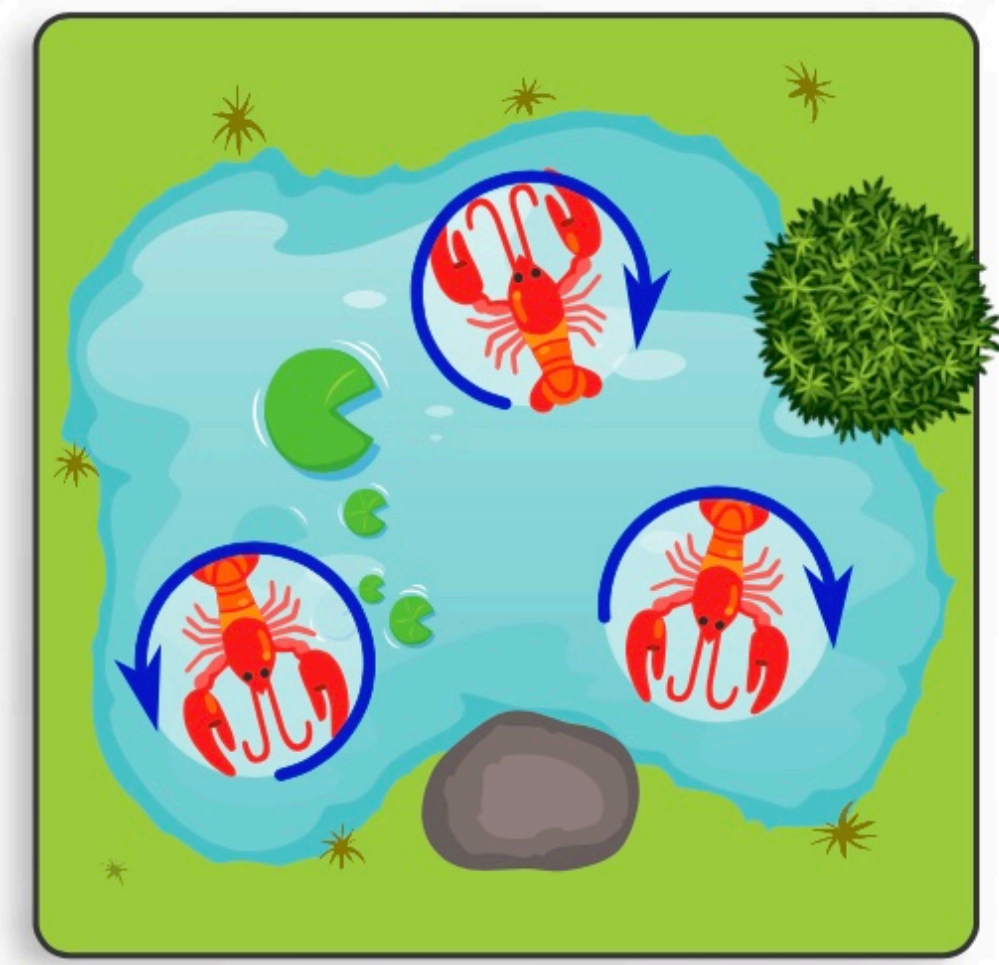
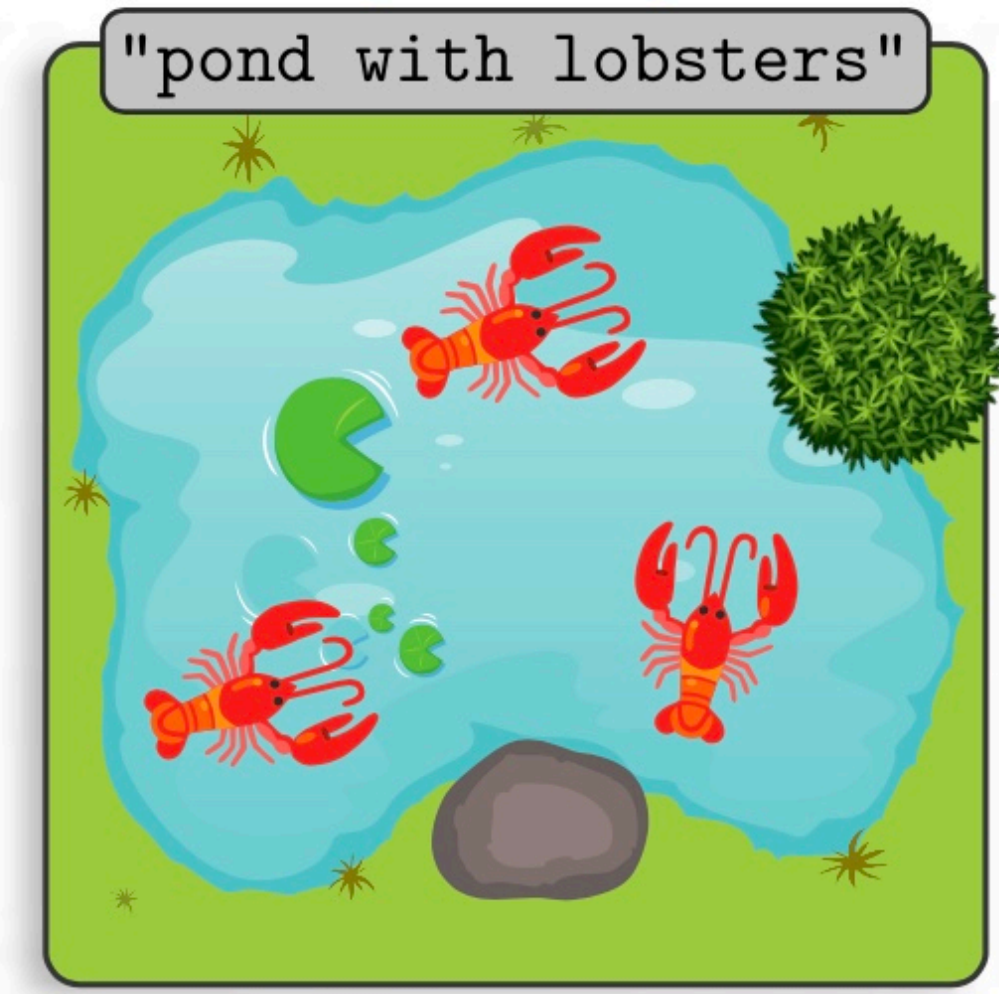


Principle of geometric deep learning:
The equations governing neural networks should be equivariant with respect to all **local** and **global** symmetries of the input data

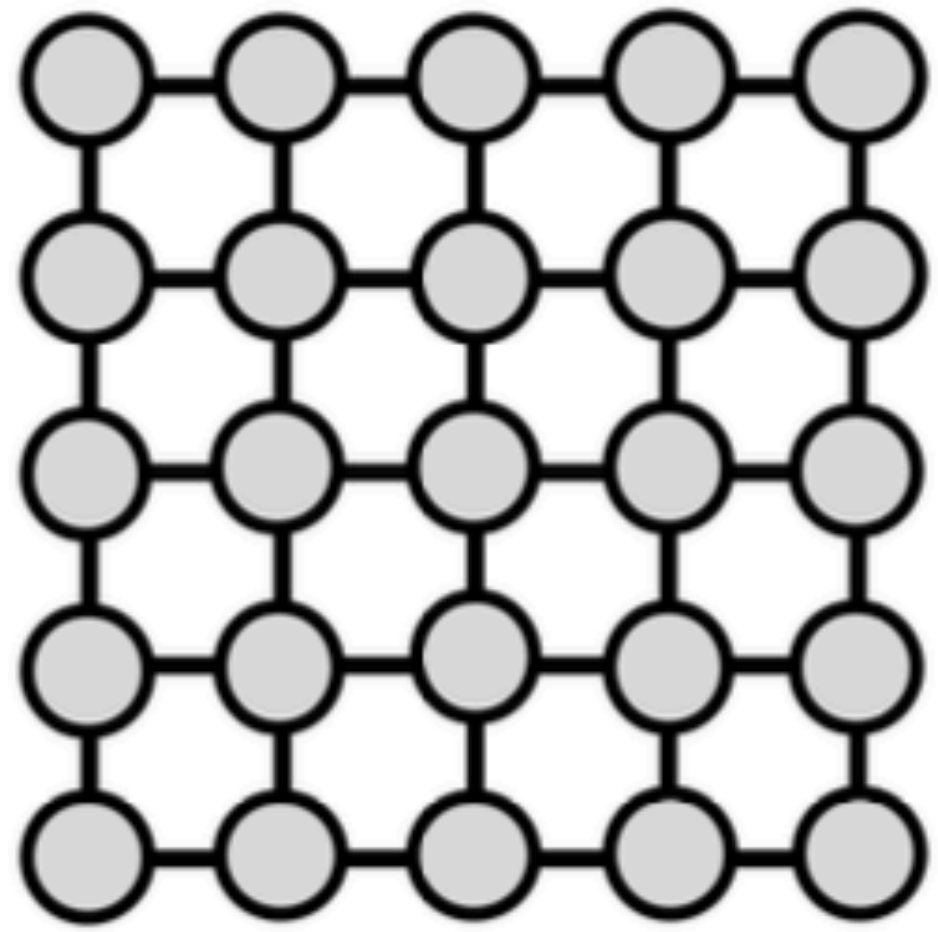


formation

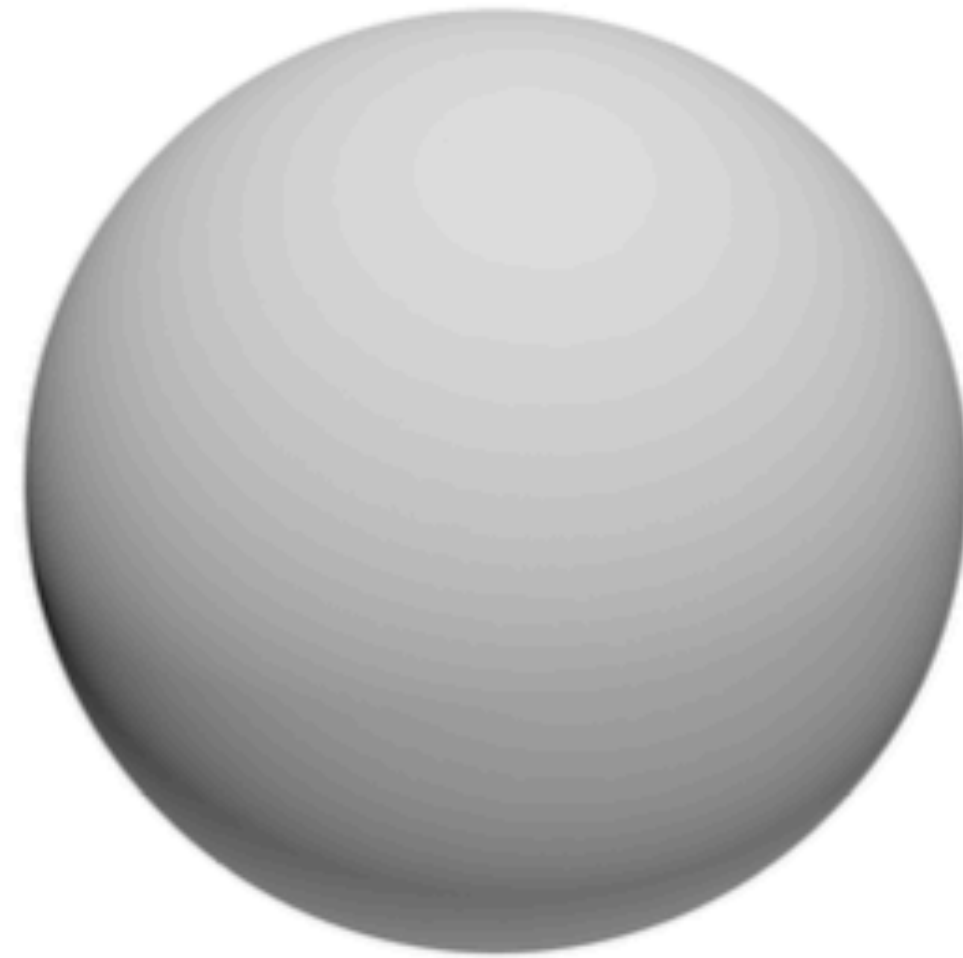
local gauge transformations



from: Weiler, Forré, Verlinde, Welling (2023)]



Grids



Groups



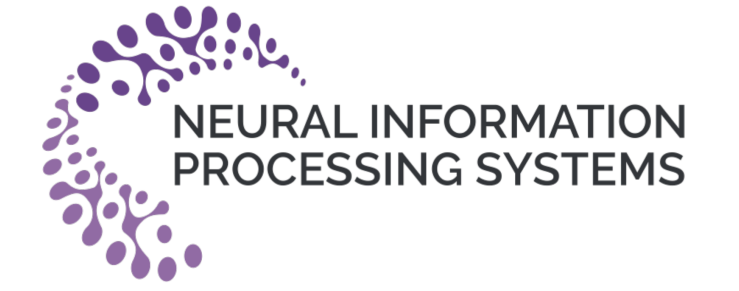
Graphs



Geodesics & Gauges

Learning Chern Numbers Of Multiband Topological Insulators With Gauge Equivariant Neural Networks

Longde Huang^{a,1}, Oleksandr Balabanov², Hampus Linander^{1,4}, Mats Granath³, Daniel Persson¹, Jan E. Gerken^{b,1}



BACKGROUND & AIMS

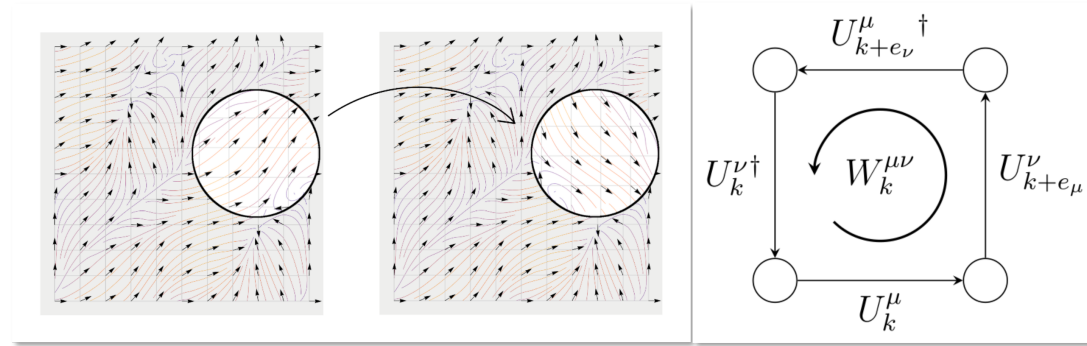
Topological insulators:

- Predicting their topological invariants (**Chern numbers**) is a key challenge.
- “**Multiband**”: matrix-valued symmetry; usual models can't even learn simple related problems like determinant.
- “**Gauge**”: local symmetry; grows **exponentially** with lattice size.

Equivariance: models with hard-coded symmetry.

- Previous models: success only for $M = N = 1$.
- LGE-CNN: numerically unstable.

Goal: To develop a gauge equivariant network that can learn multiband Chern numbers efficiently.



Left: An illustration of gauge symmetry on a lattice. Any unitary rotation on a local patch (a single lattice site when discretized) does not change the Chern number of the underlying topological insulator. Right: An illustration of the input data. Intuitively, it is an integration over a 2D infinitesimal area.

PROBLEM FORMULATION

Input: a series of unitary matrix data on a lattice Λ^{2n} . For $n = 1$:

$$(W_k)_{k \in \Lambda^2} \in U(M)^{|\Lambda^2|},$$

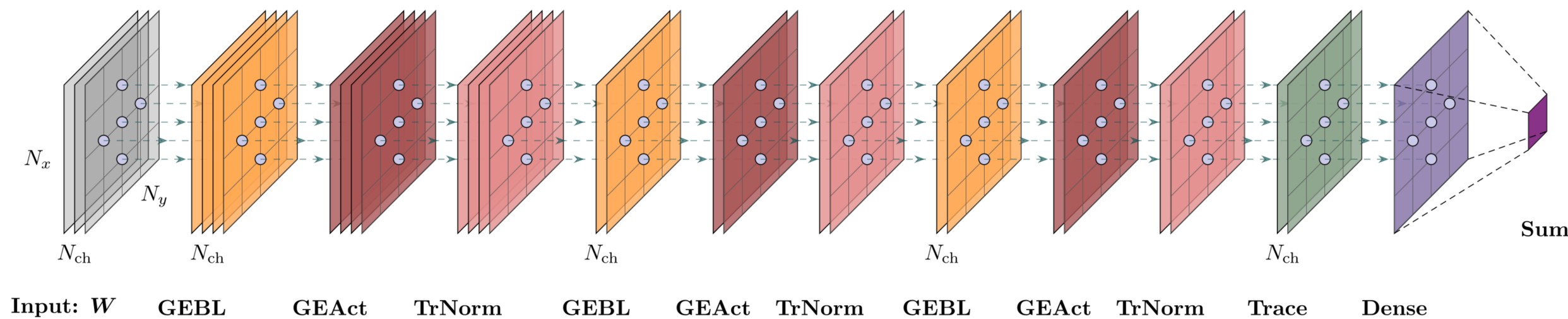
Shape: $N_x \times N_y \times M \times M \times 2$

Objective: Chern number (an integer).

$$C_1 = \frac{1}{2\pi} \sum_{k \in \Lambda^2} \text{Im Tr log } W_k$$

Gauge Group: invariance under **conjugation** at **each lattice site**.

$$(W_k) \sim (\Omega_k W_k \Omega_k^T), \forall (\Omega_k) \in U(M)^{|\Lambda^2|}$$



TOY EXAMPLE: DETERMINANTS

The determinant is:

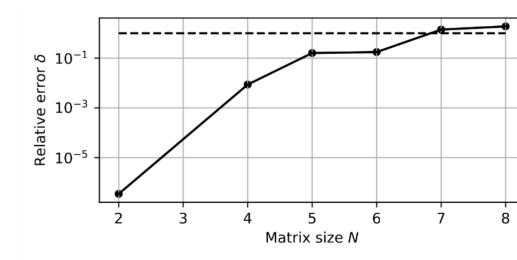
- A simpler, **polynomial** function that is **invariant under conjugation**.

Polynomial MLP: residual MLPs that have higher order layers

$$A_{ij}^{\text{out}} = \sum_{k_1, \dots, k_{2R}} \theta_{ij}^{k_1 \dots k_{2R}} A_{k_1 k_2}^{\text{in}} \dots A_{k_{2R-1} k_{2R}}^{\text{in}}, R \leq 4$$

MLP cannot learn determinants for larger matrices!

⇒ **Equivariant architecture needed.**



Best relative error of predicted matrix determinants for polynomial architectures as a function of increasing matrix size. The ablations include layers up to order 4. Dashed line indicates relative error of a mean predictor. Architectures considered include layers of order ≤ 4 and depth \leq , containing terms of up to order 16 by composition.

MODEL

GEBLNet: a purely local gauge equivariant network.

- GEBL: Gauge Equivariant Bilinear layer extracting higher order terms.

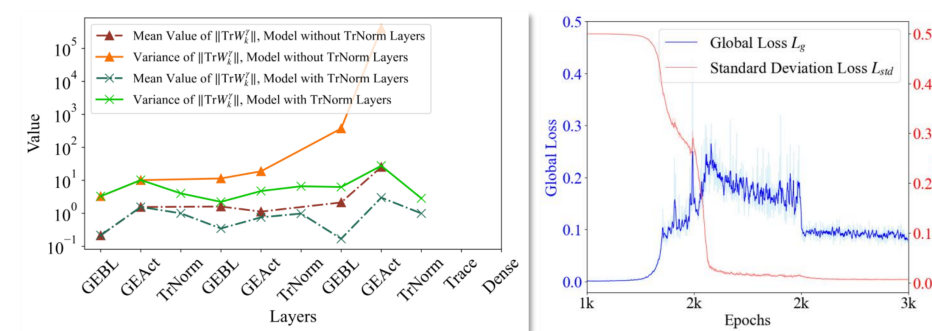
$$W_k^Y \mapsto \alpha_{\gamma\mu\nu} W_k^\mu W_k^\nu$$

- GEAct: Gauge Equivariant Activation layer.

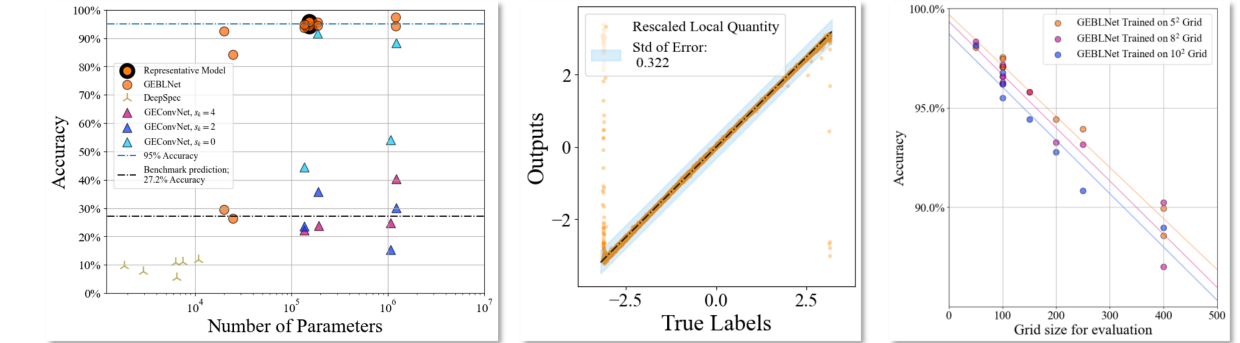
$$W_k^Y \mapsto \text{ReLU}(\text{Re Tr } W_k^Y) W_k^Y$$

- TrNorm: Key architectural novelty, channel-wise trace normalization:

$$W_k^Y \mapsto 1/(\text{mean}_y |\text{Tr } W_k^Y|) W_k^Y$$



Left: Layerwise statistics during training, with or without TrNorm layers. TrNorm layer helps restraining the rapidly accumulated trace of the features and stabilizing the training dynamic. Right: The loss curve when training on zero samples with TrNorm. Here the Standard Deviation Loss enforces that, for each sample, the local outputs maintain a minimum level of variance across lattice points.



Left: Comparison of model complexity and accuracy across different architectures. Each marker represents a training run with variations in models, learnable parameters, and random seeds. Middle: Comparison between the true local quantities and the learned local quantities. Right: Comparison of model accuracy across different grid sizes. A simple model trained on small grids performs equally well with those trained on larger grids.

RESULTS

Universal Approximation Theorem:

- Generalizes to arbitrary functions invariant under the joint unitary conjugation.

$$f(h_1, \dots, h_k) = f(gh_1g^{-1}, \dots, gh_kg^{-1}), \forall g, h_1, \dots, h_k \in U(M)$$

High expressivity: learn the Chern numbers for high number of bands.

- $N = 1$: up to $M = 7$ bands with an average accuracy of 95%.
- Captures local quantities accurately

High generalization abilities:

- Able to generalize to larger grids with minor loss of accuracy.
- Able to generalize to higher dimensions; local MAE around 0.23 for $N = 2$.

Learning on trivial data:

- Recognize nonzero Chern numbers when trained on only topologically trivial dataset.

KEY TAKEAWAY

Model: Introduced the TrNorm layer, solving the numerical instability

Theory: Proved a universal approximation theorem.

Experiments: Model showing high expressivity and generalization ability.

Limitation:

- Synthetic dataset, yet to test on real experimental data
- Higher dimensional Chern number: lacking accurate approximation.
- Other tasks: yet to try on problems emerged from different disciplines.

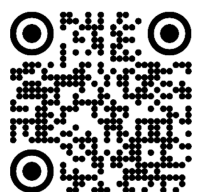
¹Department of Mathematical Sciences, Chalmers University of Technology and University of Gothenburg, Gothenburg, Sweden

²Department of Physics, Stockholm University, Stockholm, Sweden

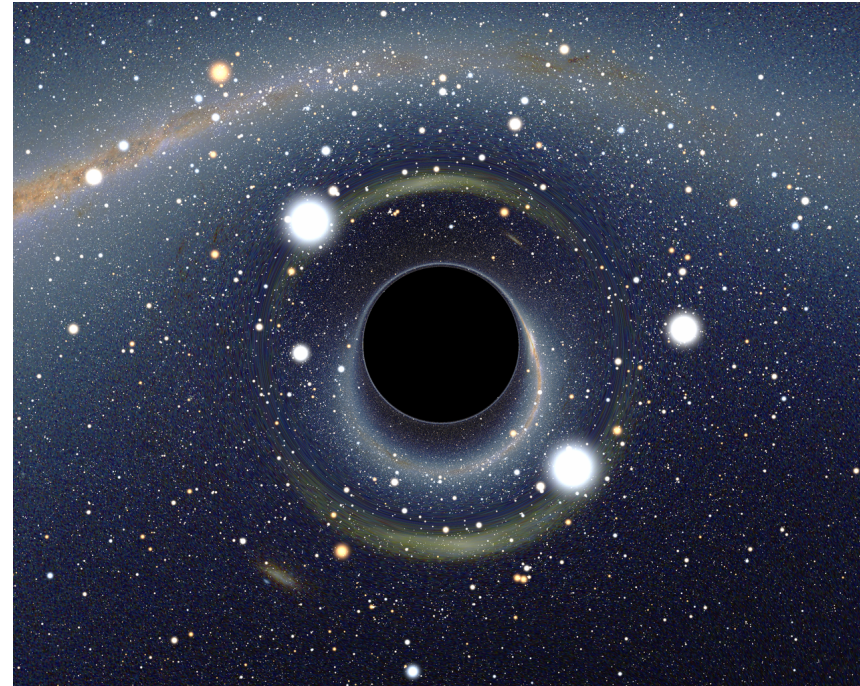
³Department of Physics, University of Gothenburg, Gothenburg, Sweden

⁴VERSES AI Research Lab, Los Angeles, USA

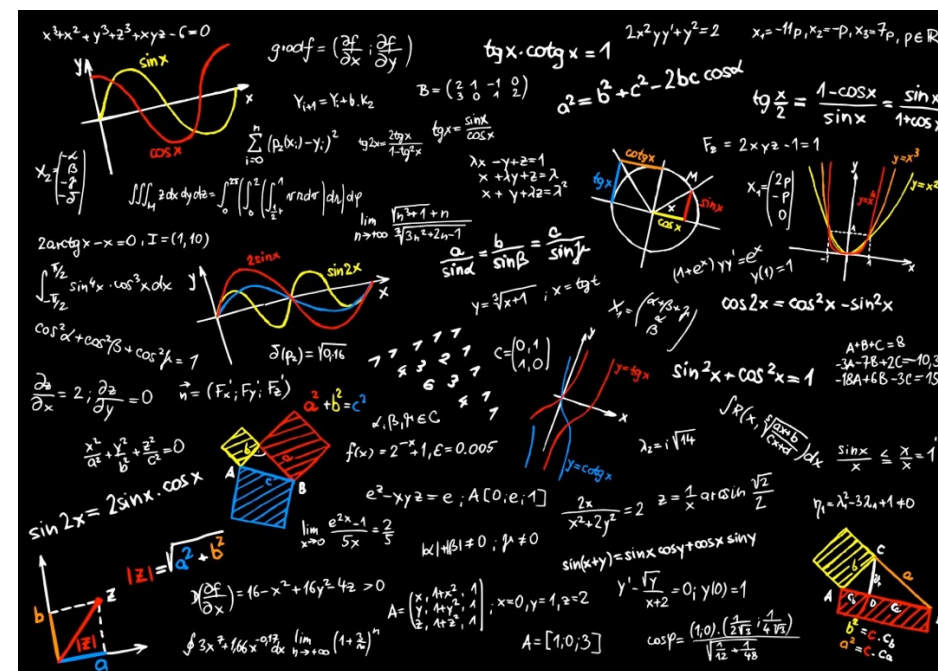
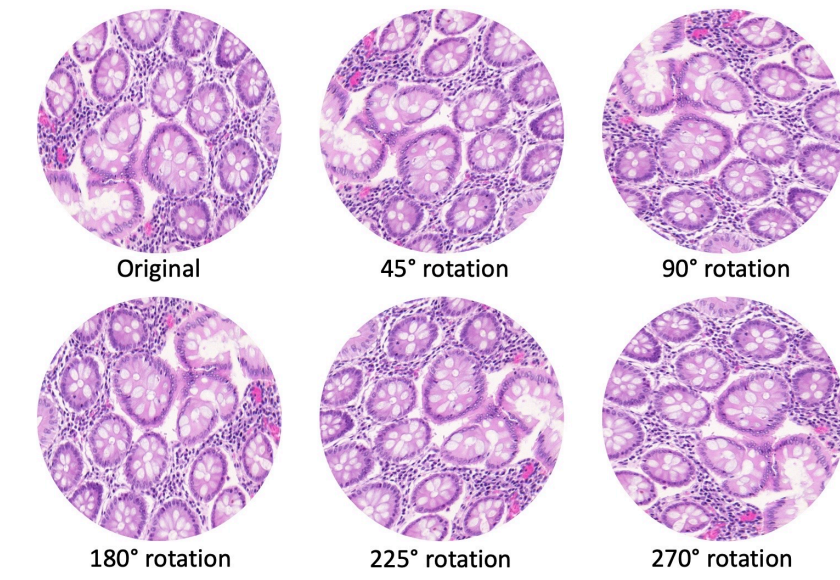
Contact: longde@chalmers.se^a, gerken@chalmers.se^b



Physics

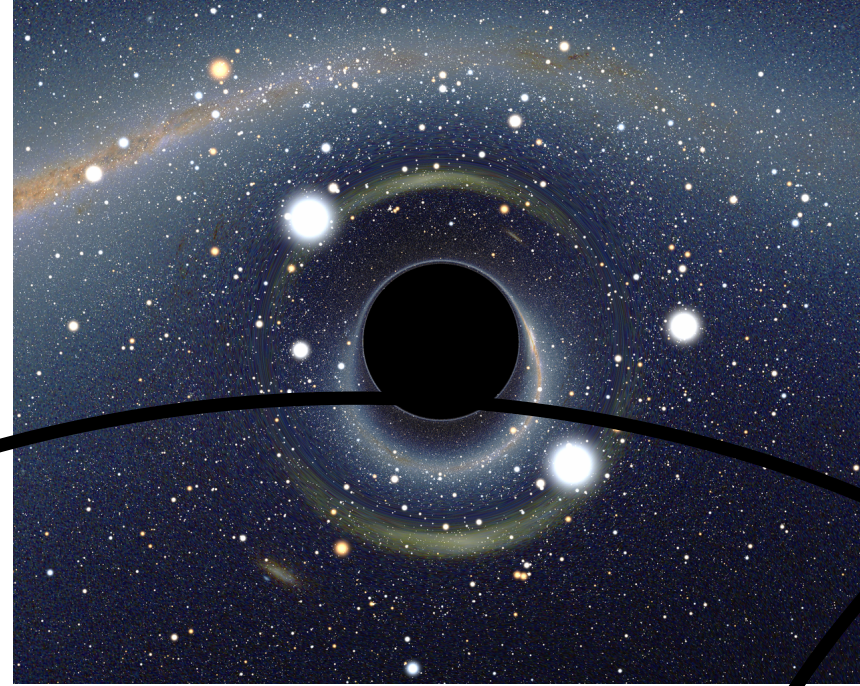


AI

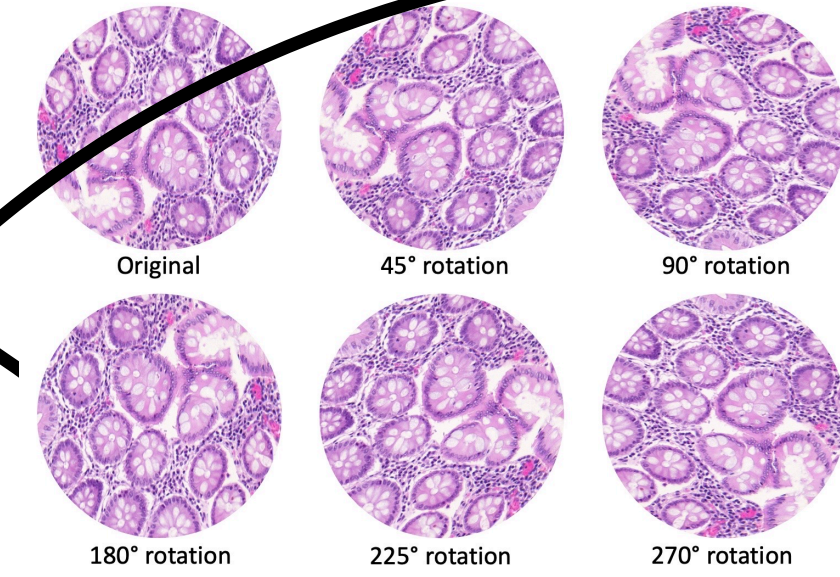


Mathematics

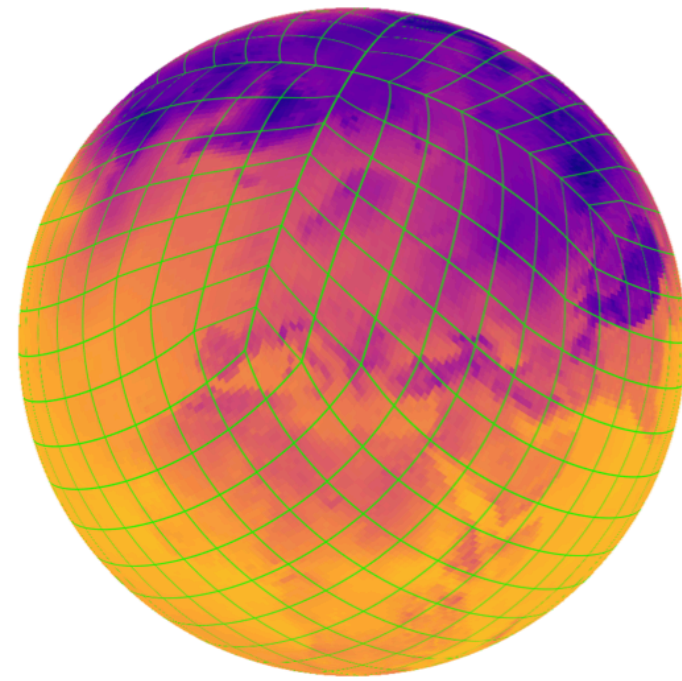
Physics



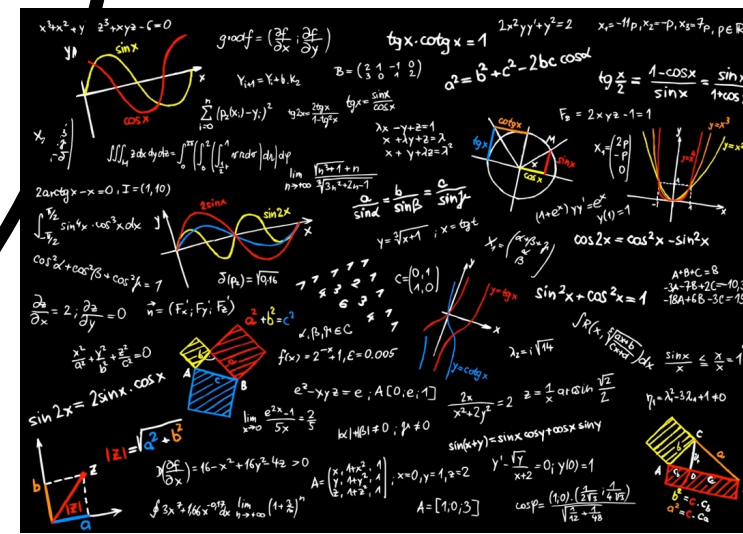
AI



Earth science



Mathematics



Biology/Chemistry



A vibrant, multi-colored galaxy with a central black hole, surrounded by a complex network of stars and filaments. The galaxy's core is a bright, swirling vortex of blue, purple, and pink, transitioning into a dark, swirling ring of orange and red. The surrounding space is filled with a dense network of thin, golden-brown filaments and numerous small, bright stars of various colors. The overall scene is set against a dark, starry background.

TACK!