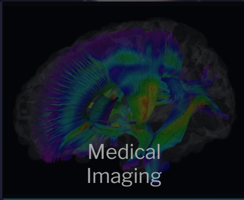
Geometric deep learning on the sphere for the physical sciences

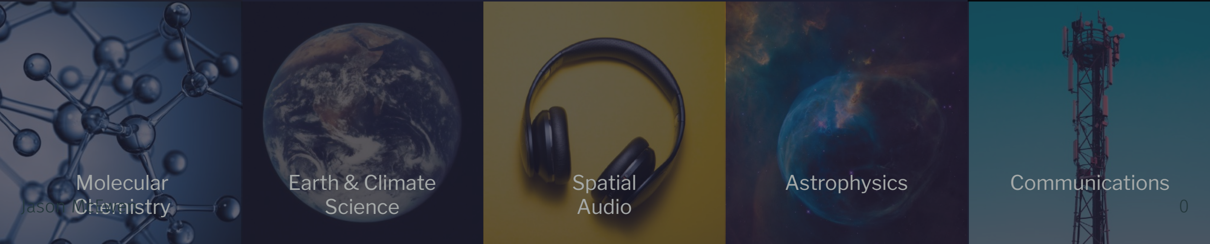
Jason McEwen
www.jasonmcewen.org

Mullard Space Science Laboratory (MSSL), UCL
CopernicAI

Maths4DL: Conference on Deep Learning for Computational Physics, July 2023

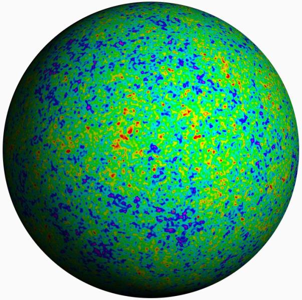
Data on the sphere arises in many applications





Cosmology and computer graphics

Whenever observe over angles, recover data on 2D sphere (or 3D rotation group).

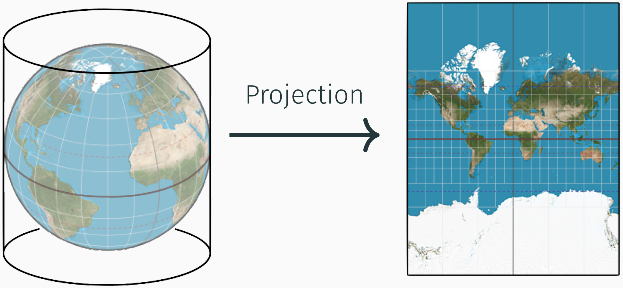


Cosmic microwave background

Computer graphics & vision

Why not standard (Euclidean) deep learning approaches?

Could project sphere to plane and then apply standard planar CNNs.



Why not standard (Euclidean) deep learning approaches?

Projection breaks symmetries and geometric properties of sphere.

$\Rightarrow$ Conformal, area-preserving projection does not exist.

Well-known that regular discretisation of the sphere does not exist (e.g. Tegmark 1996).

$\Rightarrow$ Not possible to discretise sphere in a manner that is invariant to rotations.

Why not standard (Euclidean) deep learning approaches?

Projection breaks symmetries and geometric properties of sphere.

$\Rightarrow$ Conformal, area-preserving projection does not exist.

Well-known that regular discretisation of the sphere does not exist (e.g. Tegmark 1996).

$\Rightarrow$ Not possible to discretise sphere in a manner that is invariant to rotations.

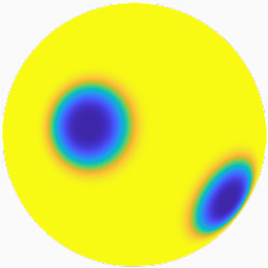
Capturing strict **rotational equivariance** with operations defined directly in discretised (pixel) space **not possible** due to structure of sphere.

Goals of geometric deep learning on the sphere

1. **Capture geometry and symmetry** of the sphere (rotational equivariance)
2. **Computationally scalable** to support high-resolution data

Categorization of spherical deep learning frameworks

Continuous

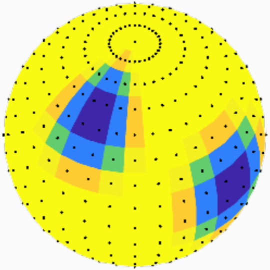


✓ Equivariant

✗ Not Scalable → ✓ Scalable

(Cohen et al. 2018, Esteves et al. 2018, Kondor et al. 2018, Cobb et al. 2021, McEwen et al. 2022, Price & McEwen in prep., Mousset et al. in prep., ...)

Discrete

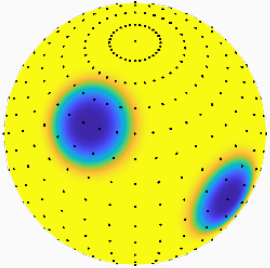


✗ Not Equivariant

✓ Scalable

(Jiang et al. 2019, Zhang et al. 2019, Perraudin et al. 2019, Cohen et al. 2019, ...)

Discrete-Continuous (DISCO)



✓ Equivariant

✓ Scalable

(Ocampo, Price & McEwen 2023)

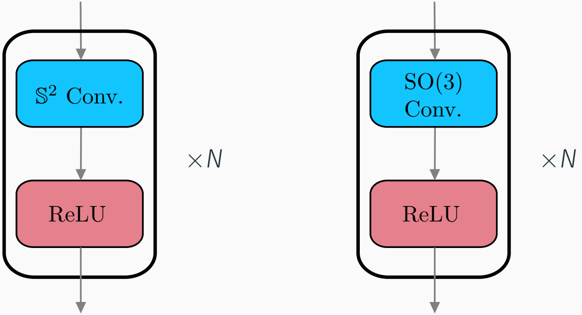
Outline

1. Spherical CNNs
2. Scattering networks on the sphere
3. Emulation of cosmic strings

Spherical CNNs

Spherical CNN

Spherical CNNs constructed by analog of Euclidean CNNs but using convolution on the sphere (and rotation group) and pointwise non-linear activations functions, e.g. ReLU (Cohen et al. 2018; Esteves et al. 2018).



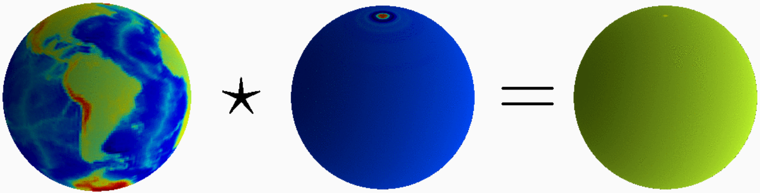
Convolution of signals on the sphere

Convolution of signals in spatial domain

Convolution of two signals $f, \psi \in L^2(\mathbb{S}^2)$ is given by

$$(f \star \psi)(\rho) = \langle f, R\psi \rangle = \int_{\mathbb{S}^2} \mathrm{d}\mu(\omega)\, f(\omega)\, \psi^*(\rho^{-1}\omega), \qquad \text{for } \omega \in \mathbb{S}^2, \rho \in \mathrm{SO}(3),$$

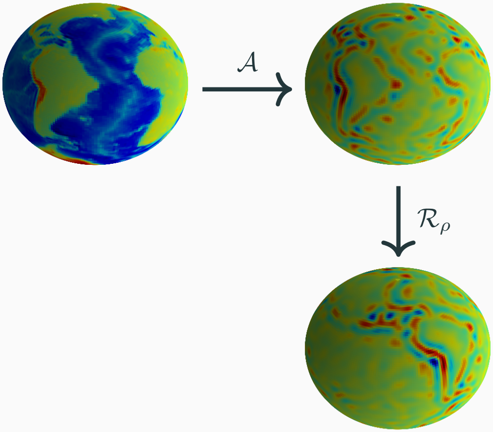
where $d\mu(\omega)$ denotes the Haar measure on $\mathbb{S}^2$ and $\cdot^*$ complex conjugation.



Convolution is rotationally equivariant

Convolution is rotationally equivariant (if computed continuously):

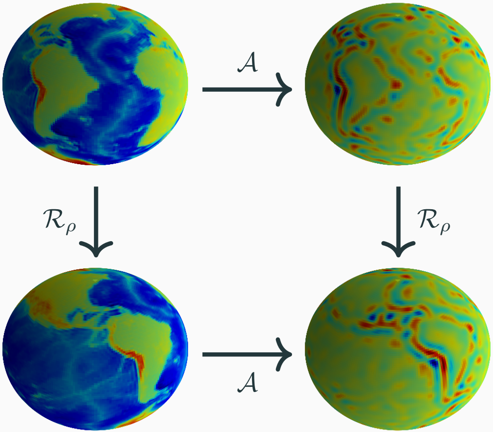
$$((\mathcal{R}_\rho f) \star \psi)(\rho') = (\mathcal{R}_\rho (f \star \psi))(\rho').$$



Convolution is rotationally equivariant

Convolution is rotationally equivariant (if computed continuously):

$$((\mathcal{R}_\rho f) \star \psi)(\rho') = (\mathcal{R}_\rho (f \star \psi))(\rho').$$



Fourier representations

Since sphere is **compact manifold**, Fourier space is discrete and **sampling theorems** can be leveraged to compute Fourier representations exactly for bandlimited signals on the sphere and rotation group (e.g. McEwen & Wiaux 2011, McEwen et al. 2011).

$\Rightarrow$ Provides access to underlying continuous signals and symmetries of sphere.

Differentiable and accelerated Fourier transforms on $\mathbb{S}^2$ and SO(3)

Tests passing | docs passing | codecov 96% | License MIT | arXiv xxxx.xxxxx | all contributors 6 | code style black

Differentiable and accelerated spherical transforms with JAX

`S2FFT` is a JAX package for computing Fourier transforms on the sphere and rotation group. It leverages autodiff to provide differentiable transforms, which are also deployable on modern hardware accelerators (e.g. GPUs and TPUs).

Github: `https://github.com/astro-informatics/s2fft`

Docs: `https://astro-informatics.github.io/s2fft`

Paper: Price & McEwen, in prep.

Harmonic space tensor product activations

Previous approach introduces non-linearity by pointwise activations in spatial domain.

- Computationally **costly** since it requires repeated harmonic transforms.
- Introduces **small equivariance error** due to irregular pixelisation.

Harmonic space tensor product activations

Previous approach introduces non-linearity by pointwise activations in spatial domain.

- Computationally **costly** since it requires repeated harmonic transforms.
- Introduces **small equivariance error** due to irregular pixelisation.

Alternatively introduce non-linearity in the harmonic domain in an equivariant manner.

Harmonic space tensor product activations

Previous approach introduces non-linearity by pointwise activations in spatial domain.

- Computationally **costly** since it requires repeated harmonic transforms.
- Introduces **small equivariance error** due to irregular pixelisation.

Alternatively introduce non-linearity in the harmonic domain in an equivariant manner.

Consider irreducible representations of the rotation group SO(3) and leverage the decomposability of the tensor product between these representations (Thomas et al. 2018, Kondor et al. 2018).

$\Rightarrow$ **Clebsch-Gordan** decomposition (cf. coupling of angular momenta in quantum mechanics).

Efficient generalized spherical CNNs

Consider the s-th layer of a generalized spherical CNN to take the form of a triple (Cobb et al. 2021; arXiv:2010.11661)

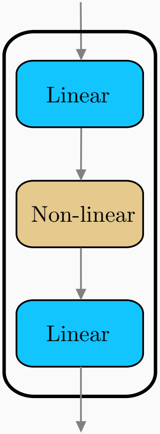
$$\mathcal{A}^{(s)} = (\mathcal{L}_1, \mathcal{N}, \mathcal{L}_2),$$

such that

$$\mathcal{A}^{(s)}(f^{(s-1)}) = \mathcal{L}_2(\mathcal{N}(\mathcal{L}_1(f^{(s-1)}))),$$

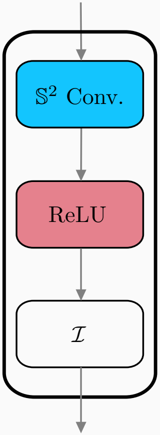
where

- $\mathcal{L}_1, \mathcal{L}_2 : \mathcal{F}_L \rightarrow \mathcal{F}_L$ are spherical convolution operators,
- $\mathcal{N} : \mathcal{F}_L \rightarrow \mathcal{F}_L$ is a non-linear, spherical activation operator.

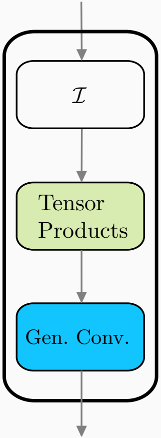


Efficient generalised spherical CNNs

- Build on other **influential equivariant spherical CNN** constructions:
  - Cohen et al. (2018)
  - Esteves et al. (2018)
  - Kondor et al. (2018)
- Encompass other frameworks as special cases.
- General framework supports hybrids models.
- Significant efficiency improvements.



Cohen et al. (2018), Esteves et al. (2018)



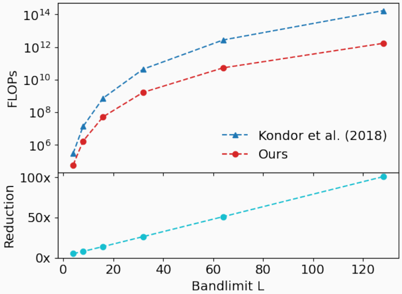
Kondor et al. (2018)

Contributions to improve efficiency

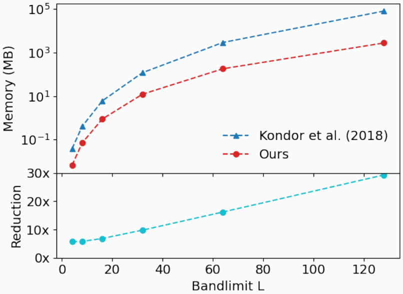
1. Channel-wise structure
2. Constrained generalized convolutions
3. Optimized degree mixing sets
4. Efficient sampling theory on the sphere and rotation group
   (McEwen & Wiaux 2011; McEwen et al. 2015)

Computational cost and memory requirements

- **State-of-the-art (SOTA) performance** on many benchmark problems.
- Considerable **computational savings** in FLOPs and memory.



Computational cost



Memory requirements

Discrete-continuous (DISCO) spherical convolution

Scalable and Equivariant Spherical CNNs by Discrete-Continuous (DISCO) Convolutions
(Ocampo, Price & McEwen 2023; arXiv:2209.13603)

Follows by a careful hybrid representation of the spherical convolution:

- some components left continuous, to facilitate accurate rotational equivariance;
- while other components are discretized, to yield scalable computation.

Discrete-continuous (DISCO) spherical convolution

DISCO spherical convolution

Spherical convolution can be carefully approximated by the DISCO representation

$$(f \star \psi)(R) = \int_{\mathbb{S}^2} f(\omega)\psi(R^{-1}\omega)\mathrm{d}\omega \approx \sum_i f[\omega_i]\psi(R^{-1}\omega_i)q(\omega_i),$$

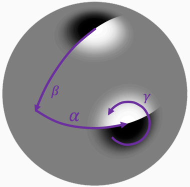
for spherical signal and filter kernel $f, \psi : \mathbb{S}^2 \rightarrow \mathbb{R}$, with spherical coordinates $\omega \in \mathbb{S}^2$, where, for now, we consider 3D rotations $R \in \mathrm{SO}(3)$.

- Appeal to sampling theorem on the sphere with quadrature weights $q : \mathbb{S}^2 \rightarrow \mathbb{R}$ (McEwen & Wiaux 2011; arXiv:1110.6298):
  - $\Rightarrow$ all information content of signal captured by samples $\{f[\omega_i]\}_i$;
  - $\Rightarrow$ continuous integral evaluated accurately by quadrature (exact for sufficient sampling).
- Filter ψ and rotation R treated continuously to avoid any discretization artefacts.

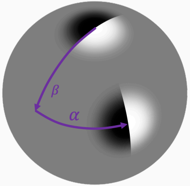
Restricting rotations to SO(3)/SO(2)

While the DISCO spherical convolution is already efficient, we seek further computational savings by restricting the space of rotations to quotient space SO(3)/SO(2).

- **Analogous to Euclidean planar CNNs**, where filters are translated across the image but are *not* rotated in the plane.
- However, as the space SO(3)/SO(2) is **not a group**, when restricting rotations in this manner important differences to the usual setting arise.



$$R = Z(\alpha)Y(\beta)Z(\gamma) \in \mathrm{SO}(3)$$



$$R = Z(\alpha)Y(\beta) \in \mathrm{SO}(3)/\mathrm{SO}(2) \simeq \mathbb{S}^2$$

Rotational equivariance for rotations $R \in \mathrm{SO}(3)$

DISCO spherical convolution $f \star \psi$ for rotations $Q, R \in \mathrm{SO}(3)$ satisfies SO(3) rotational equivariance.

Only holds since SO(3) exhibits a **group structure** and so $Q^{-1}R \in \mathrm{SO}(3)$.

Asymptotic rotational equivariance for rotations $R \in \mathrm{SO}(3)/\mathrm{SO}(2)$

DISCO spherical convolution $f \circledast \psi$ for rotations $Q, R \in \mathrm{SO}(3)/\mathrm{SO}(2)$ **does not satisfy rotational equivariance** (in contrast to the Euclidean setting).

But DISCO spherical convolution $f \circledast \psi$ does satisfy asymptotic SO(3) equivariance as $\beta \rightarrow 0$, where $Q = Z(\alpha)Y(\beta)Z(\gamma)$.

Asymptotic SO(3) equivariance of significant practical use since content in spherical signals often orientated and similar content appears at similar latitudes, particularly for 360° panoramic photos and video.

Computationally scalable DISCO spherical convolution

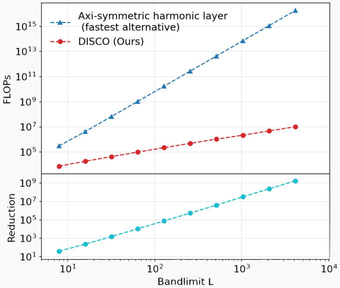
DISCO convolution affords a computationally scalable implementation.

1. Sparse tensor representation.
2. Memory compression.
3. Custom sparse gradients.

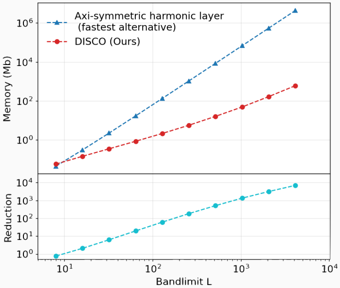
Linear scaling in number of pixels on the sphere $O(N) = O(L^2)$ for both computational cost and memory usage.

Computational cost and memory requirements

- **State-of-the-art (SOTA) performance** on many dense-prediction benchmark problems.
- Dramatic **computational savings** in FLOPs and memory.



Computational cost



Memory requirements

For 4k spherical image, 10^9 saving in computational cost and 10^4 saving in memory usage.

Scattering networks on the sphere

Scattering networks

Scattering networks inspired by CNNs but designed rather than learned filters (Mallat 2012).

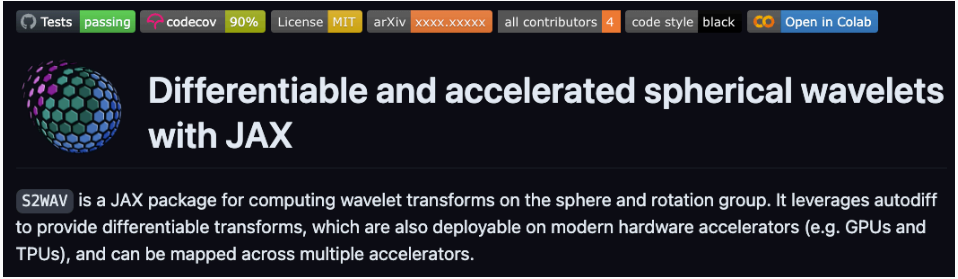
Scattering networks

Scattering networks inspired by CNNs but designed rather than learned filters (Mallat 2012).

$\Rightarrow$ Scattering networks on the sphere follows by direct analogue of Mallat's Euclidian construction (McEwen et al. 2022; arXiv:2102.02828)

1. Scalable
2. Rotationally equivariant
3. Stable and locally invariant representation (i.e. effective representation space)

Adopt scale-discretized wavelets on the sphere (e.g. McEwen et al. 2018, McEwen et al. 2015).



Github: `https://github.com/astro-informatics/s2wav`

Docs: `https://astro-informatics.github.io/s2wav`

Paper: Price, Polanska, Whitney & McEwen, in prep.

Scattering transform on the sphere

Spherical scattering propagator for scale j:

$$U[j]f = |f \star \psi_j|.$$

Modulus function is adopted for the activation function since non-expansive. Acts to mix signal content to low frequencies.

Scattering transform on the sphere

Spherical scattering propagator for scale j:

$$U[j]f = |f \star \psi_j|.$$

Modulus function is adopted for the activation function since non-expansive. Acts to mix signal content to low frequencies.

Spherical cascade of propagators:

$$U[p]f = |||f \star \psi_{j_1}| \star \psi_{j_2}| \dots \star \psi_{j_d}|,$$

for the path $p = (j_1, j_2, \dots, j_d)$ with depth d.

Scattering transform on the sphere

Spherical scattering propagator for scale j:

$$U[j]f = |f \star \psi_j|.$$

Modulus function is adopted for the activation function since non-expansive. Acts to mix signal content to low frequencies.

Spherical cascade of propagators:

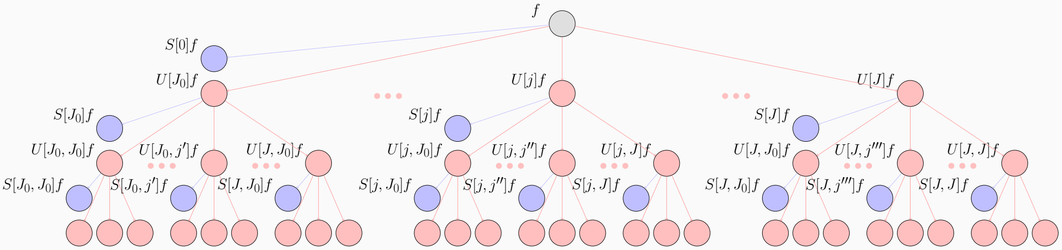
$$U[p]f = |||f \star \psi_{j_1}| \star \psi_{j_2}| \dots \star \psi_{j_d}|,$$

for the path $p = (j_1, j_2, \dots, j_d)$ with depth d.

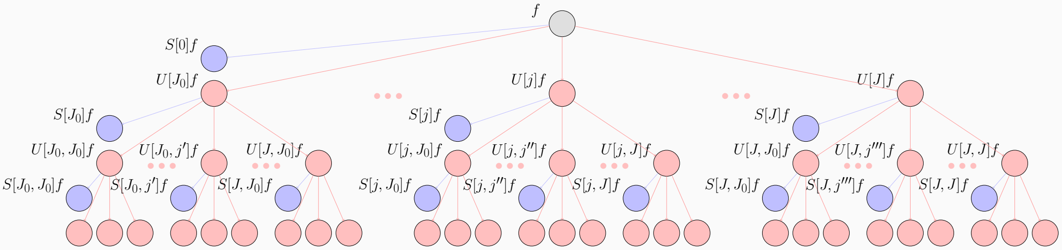
Scattering coefficients:

$$S[p]f = |||f \star \psi_{j_1}| \star \psi_{j_2}| \dots \star \psi_{j_d}| \star \phi.$$

Spherical scattering network is collection of scattering transforms for a number of paths: $\mathcal{S}_{\mathbb{P}} f = \{S[p]f : p \in \mathbb{P}\}$, where the general path set $\mathbb{P}$ denotes the infinite set of all possible paths $\mathbb{P} = \{p = (j_1, j_2, \ldots, j_d) : J_0 \leq j_i \leq J, 1 \leq i \leq d,\ d \in \mathbb{N}_0\}$.

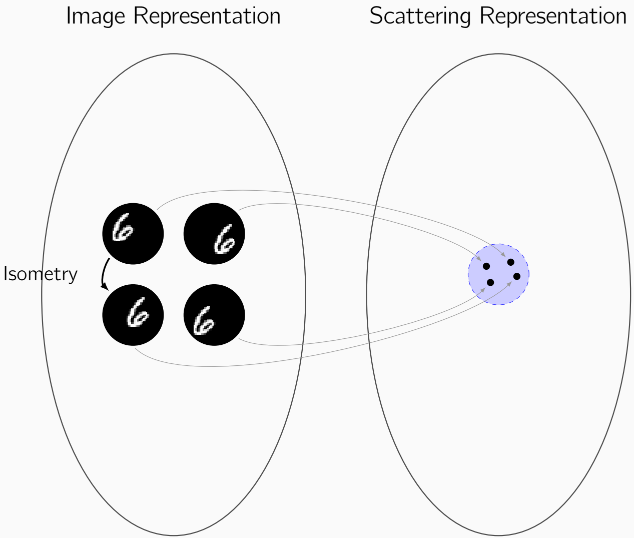


Spherical scattering network is collection of scattering transforms for a number of paths: $\mathcal{S}_{\mathbb{P}} f = \{S[p]f : p \in \mathbb{P}\}$, where the general path set $\mathbb{P}$ denotes the infinite set of all possible paths $\mathbb{P} = \{p = (j_1, j_2, \ldots, j_d) : J_0 \leq j_i \leq J, 1 \leq i \leq d,\ d \in \mathbb{N}_0\}$.



Scattering networks are rotationally equivariant (since the spherical wavelet transform and modulus operator are rotationally equivariant).

Isometric invariance



Isometric invariance

Theorem (Isometric Invariance)

Let $\zeta \in \mathrm{Isom}(\mathbb{S}^2)$*, then there exists a constant* C *such that for all* $f \in \mathrm{L}^2(\mathbb{S}^2)$*,*

$$\|\mathcal{S}_{\mathbb{P}_D} f - \mathcal{S}_{\mathbb{P}_D} V_\zeta f\|_2 \leq C L^{5/2} (D+1)^{1/2} \, \lambda^{J_0} \, \|\zeta\|_\infty \|f\|_2 .$$

(**Proof:** Follows by straightforward extension of proof of Perlmutter et al. 2020.)

Isometric invariance

Theorem (Isometric Invariance)

Let $\zeta \in \mathrm{Isom}(\mathbb{S}^2)$, then there exists a constant C such that for all $f \in \mathrm{L}^2(\mathbb{S}^2)$,

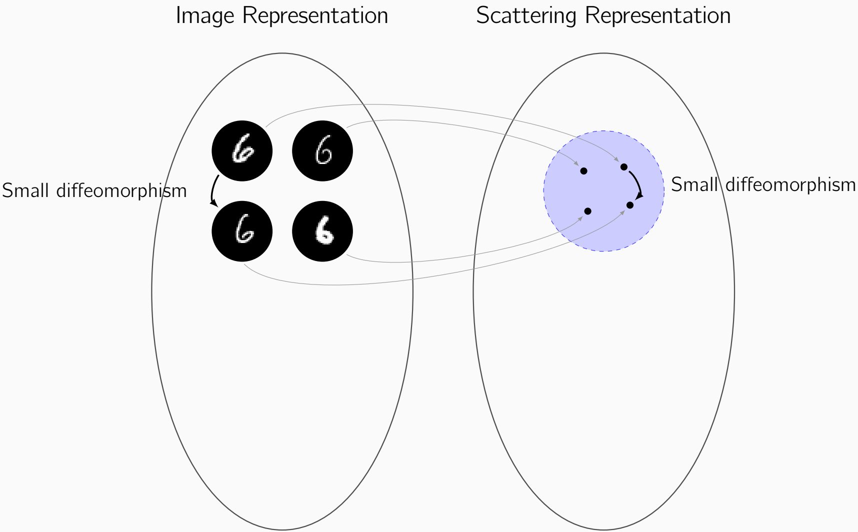
$$\|\mathcal{S}_{\mathbb{P}_D} f - \mathcal{S}_{\mathbb{P}_D} V_\zeta f\|_2 \leq C L^{5/2} (D+1)^{1/2} \lambda^{J_0} \|\zeta\|_\infty \|f\|_2 .$$

Difference in representation

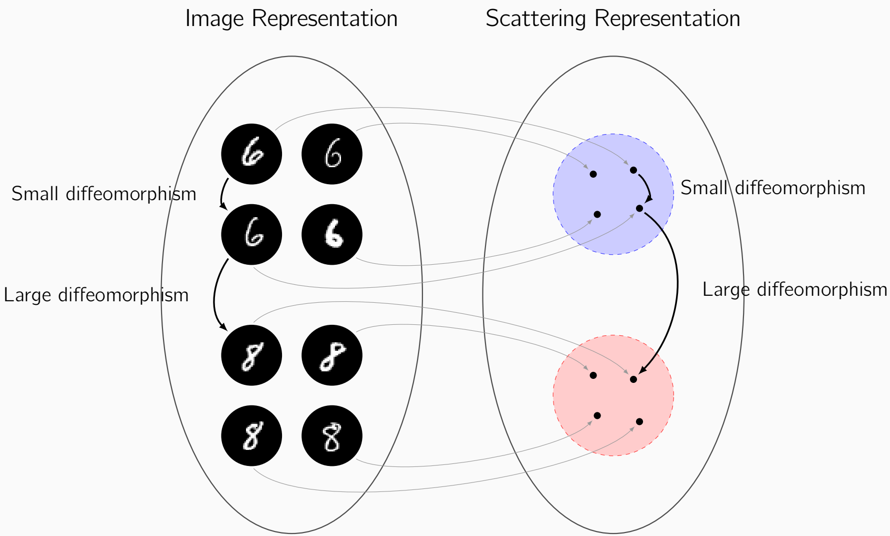
Scattering network representation is invariant to isometries up to a scale.

(**Proof:** Follows by straightforward extension of proof of Perlmutter et al. 2020.)

Stability to diffeomorphisms



Stability to diffeomorphisms



Stability to diffeomorphisms

Theorem (Stability to Diffeomorphisms)

Let $\zeta \in \mathrm{Diff}(\mathbb{S}^2)$. *If* $\zeta = \zeta_1 \circ \zeta_2$ *for some isometry* $\zeta_1 \in \mathrm{Isom}(\mathbb{S}^2)$ *and diffeomorphism* $\zeta_2 \in \mathrm{Diff}(\mathbb{S}^2)$, *then there exists a constant* C *such that for all* $f \in \mathrm{L}^2(\mathbb{S}^2)$,

$$\|\mathcal{S}_{\mathbb{P}_D} f - \mathcal{S}_{\mathbb{P}_D} V_\zeta f\|_2 \ \leq CL^2 \left[L^2 \ \|\zeta_2\|_\infty \ + L^{1/2}(D+1)^{1/2} \lambda^{J_0} \ \|\zeta_1\|_\infty \ \right] \|f\|_2 .$$

(**Proof:** Follows by straightforward extension of proof of Perlmutter et al. 2020.)

Stability to diffeomorphisms

Theorem (Stability to Diffeomorphisms)

Let $\zeta \in \mathrm{Diff}(\mathbb{S}^2)$*. If* $\zeta = \zeta_1 \circ \zeta_2$ *for some isometry* $\zeta_1 \in \mathrm{Isom}(\mathbb{S}^2)$ *and diffeomorphism* $\zeta_2 \in \mathrm{Diff}(\mathbb{S}^2)$*, then there exists a constant* C *such that for all* $f \in \mathrm{L}^2(\mathbb{S}^2)$*,*

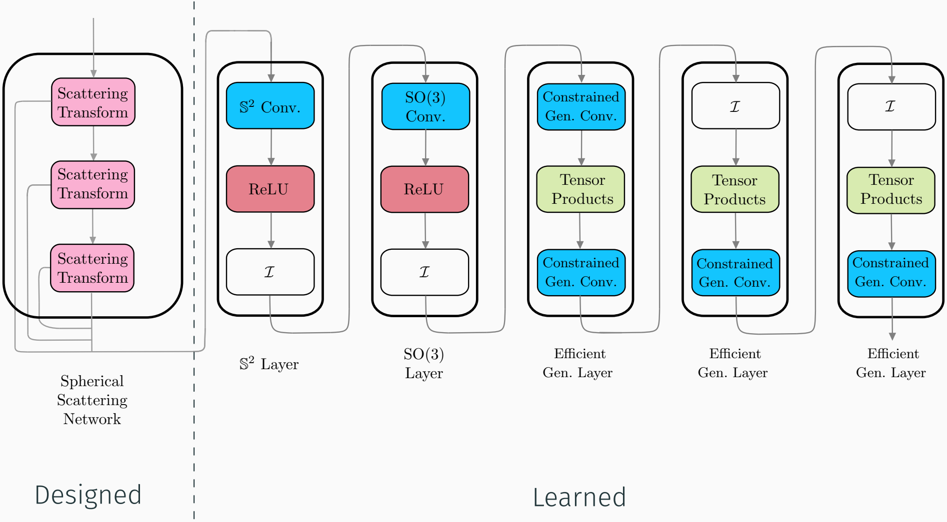
$$\|\mathcal{S}_{\mathbb{P}_D} f - \mathcal{S}_{\mathbb{P}_D} V_\zeta f\|_2 \leq C L^2 \left[L^2 \, \|\zeta_2\|_\infty + L^{1/2} (D+1)^{1/2} \lambda^{J_0} \, \|\zeta_1\|_\infty \right] \|f\|_2 .$$

Difference in representation

Scattering network representation is stable to small diffeomorphisms about isometry.

(**Proof:** Follows by straightforward extension of proof of Perlmutter et al. 2020.)

Scalable and rotationally equivariant spherical CNNs



Spherical scattering covariance

Spherical scattering covariances (Mousset, Price, Allys, McEwen, in prep.)

Captures non-Gaussian properties of (cosmological) fields very well.

Scattering statistics considered:

1. $S_1[j]\,f = \mathbb{E}\big[\, |f \star \psi_j| \,\big]$.
2. $P_{00}[j]\,f = \mathbb{E}\big[\, |f \star \psi_j|^2 \,\big]$.
3. $C_{01}[j_1, j_2]\,f = \mathrm{Cov}\big[\, f \star \psi_{j_2}, |f \star \psi_{j_1}| \star \psi_{j_2} \,\big]$.
4. $C_{11}[j_1, j_2, j_3]\,f = \mathrm{Cov}\big[\, |f \star \psi_{j_1}| \star \psi_{j_3}, |f \star \psi_{j_2}| \star \psi_{j_3} \,\big]$.

Emulation of cosmic strings

Cosmic strings

Symmetry breaking **phase transitions** in the early Universe $\rightarrow$ **topological defects**.

Cosmic strings **well-motivated** phenomenon that arise when axial or cylindrical symmetry is broken $\rightarrow$ **line-like discontinuities** in the fabric of the Universe.

Cosmic strings

Symmetry breaking **phase transitions** in the early Universe $\rightarrow$ **topological defects**.

Cosmic strings **well-motivated** phenomenon that arise when axial or cylindrical symmetry is broken $\rightarrow$ **line-like discontinuities** in the fabric of the Universe.

Observed string-like topological defects in other media, e.g. ice and liquid crystal.

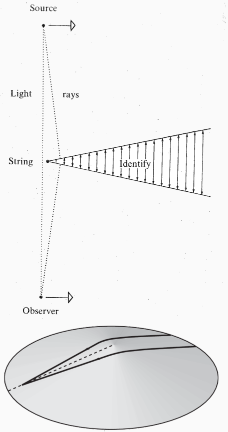
Detection of cosmic strings would open a new window into the physics of the Universe!

Optical microscope **photograph** of liquid crystal after temperature quench. [Credit: Chuang et al. 1991].

Observational signatures of cosmic strings

Spacetime about a cosmic string is **canonical**, with a three-dimensional wedge removed (Vilenkin 1981).

Strings moving transverse to the line of sight induce line-like discontinuities in the CMB (Kaiser & Stebbins 1984).

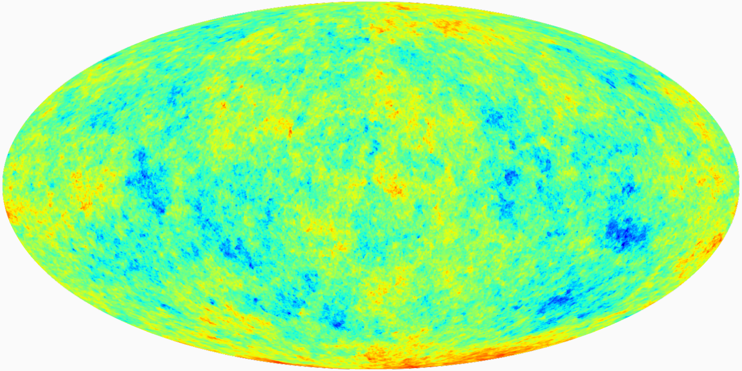


Spacetime around a cosmic string. [Credit: Kaiser & Stebbins 1984].

Simulation of cosmic strings

Contact between theory and data via high-resolution simulations (Ringeval et al. 2012).

Need to **simulate full physics**, evolving a network of strings through cosmic time, and then ray-trace CMB photons through the string network.



A single simulation requires 800,000 CPU hours on a supercomputer.

In total there are three full-sky string maps in existence.

Emulation to match scattering covariance

Instead of simulating full physics we emulate with a generative model.

Emulation to match scattering covariance

Instead of simulating full physics we emulate with a generative model.

Emulate by **matching scattering covariance statistics** $\mathcal{S}(f)$ with a (single) simulation:

$$\min_{f_{\text{emu}}} \|\mathcal{S}(f_{\text{emu}}) - \mathcal{S}(f_{\text{sim}})\|^2.$$

Emulation to match scattering covariance

Instead of simulating full physics we emulate with a generative model.

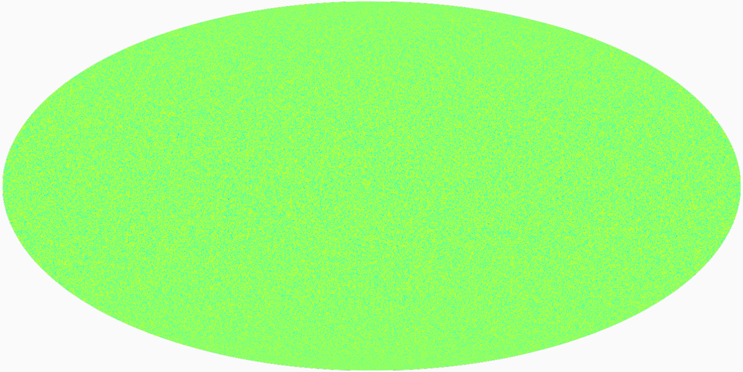
Emulate by **matching scattering covariance statistics** $\mathcal{S}(f)$ with a (single) simulation:

$$\min_{f_{\text{emu}}} \|\mathcal{S}(f_{\text{emu}}) - \mathcal{S}(f_{\text{sim}})\|^2 .$$

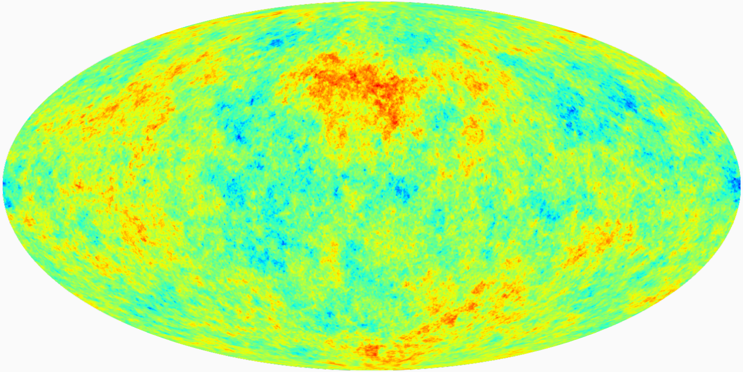
Solve using L-BFGS algorithm in JAX.

Leverage **automatic differentiable** of spherical harmonic transforms (`s2fft`), wigner transforms (`s2fft`), wavelet transform (`s2wav`), and spherical scattering computation (coming very soon!).

Emulation of cosmic strings



Emulation of cosmic strings

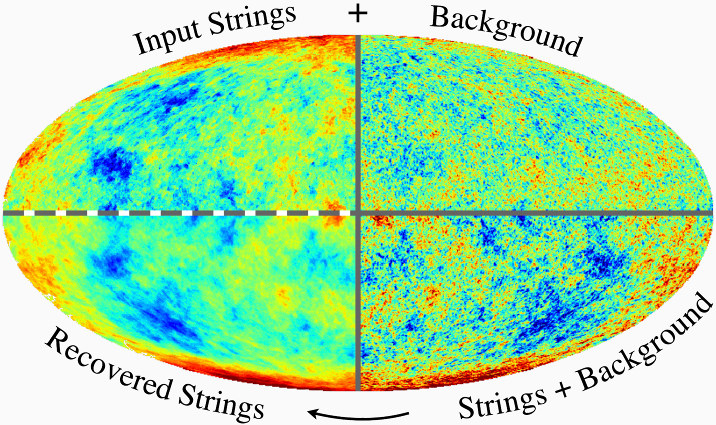


Computation time: 800,000 CPU hours on supercomputer $\rightarrow \mathcal{O}(1)$ hours on A100 GPU.

Still work in progress (detailed statistical validation in progress).

Cosmic strings

Ability to rapidly emulate CMB string maps opens up many new analyses to search for evidence of cosmic strings, such as Bayesian inference, simulation-based inference (cf. McEwen et al. 2018; Planck Collaboration XXV 2014).



Summary

- **Data on the sphere prevalent** (cosmology, climate, geophysics, computer graphics, ...).
- Require geometric deep learning on the sphere to encode symmetries and geometric properties (rotational equivariance); e.g. spherical CNNs, spherical scattering networks.
- Need to carefully design approaches to ensure **computationally scalable**.
- **Differentiable programming** critical: accelerated and differentiable codes for generalised Fourier transforms on the sphere and rotation group (`s2fft`), spherical wavelet transforms (`s2wav`), scattering networks (coming!), spherical CNNs (coming!).
- Rich and active field with **many potential new applications!**