Beyond the Gaussian:
A Higher-order Correlation Statistic for the Interstellar Medium

Douglas Finkbeiner
(including work by A. Saydjari)

1 April, 2021

We would like a way to characterize non-Gaussianity in physical systems.

(Many examples in astrophysics and cosmology)

Question 1: What do we mean by non-Gaussianity?

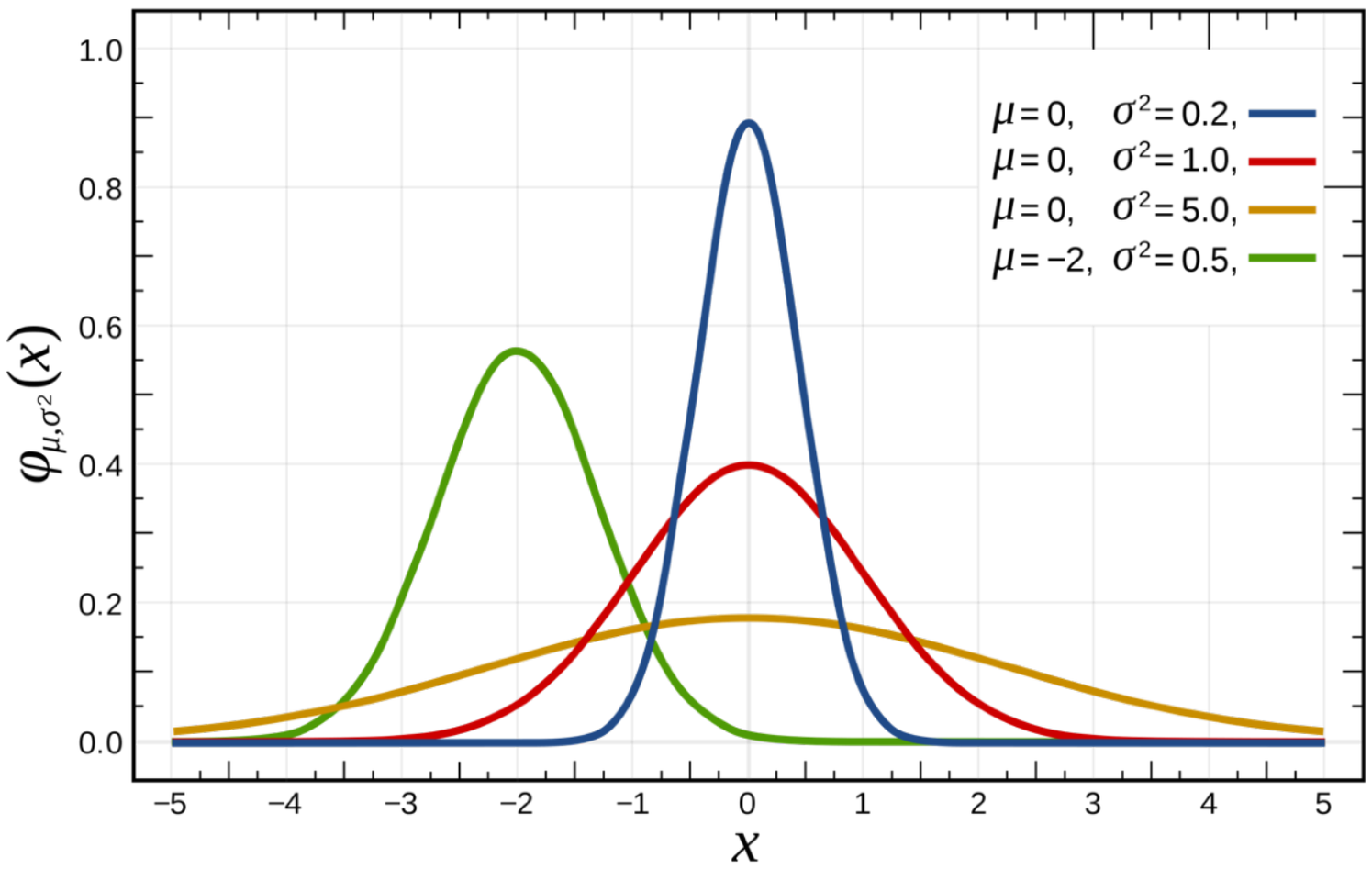
~~Question 1:~~ What do we mean by non-Gaussianity?

Question 0: What do we mean by Gaussianity?

Gaussian or normal distribution, 1D

$$\mathcal{N}(x\ ;\ \mu, \sigma) = \frac{1}{\sqrt{2\pi\sigma^2}} \exp\left[-\frac{1}{2}(x-\mu)^2/\sigma^2\right]$$

- Parameters: mean μ, variance σ^2
  (standard deviation σ)



https://medium.com/@SeoJaeDuk/archived-post-multivariate-gaussian-distributions-and-entropy-3-991578ca534c
and Wikipedia

But that isn't quite what we are talking about.
We are talking about the covariance structure.

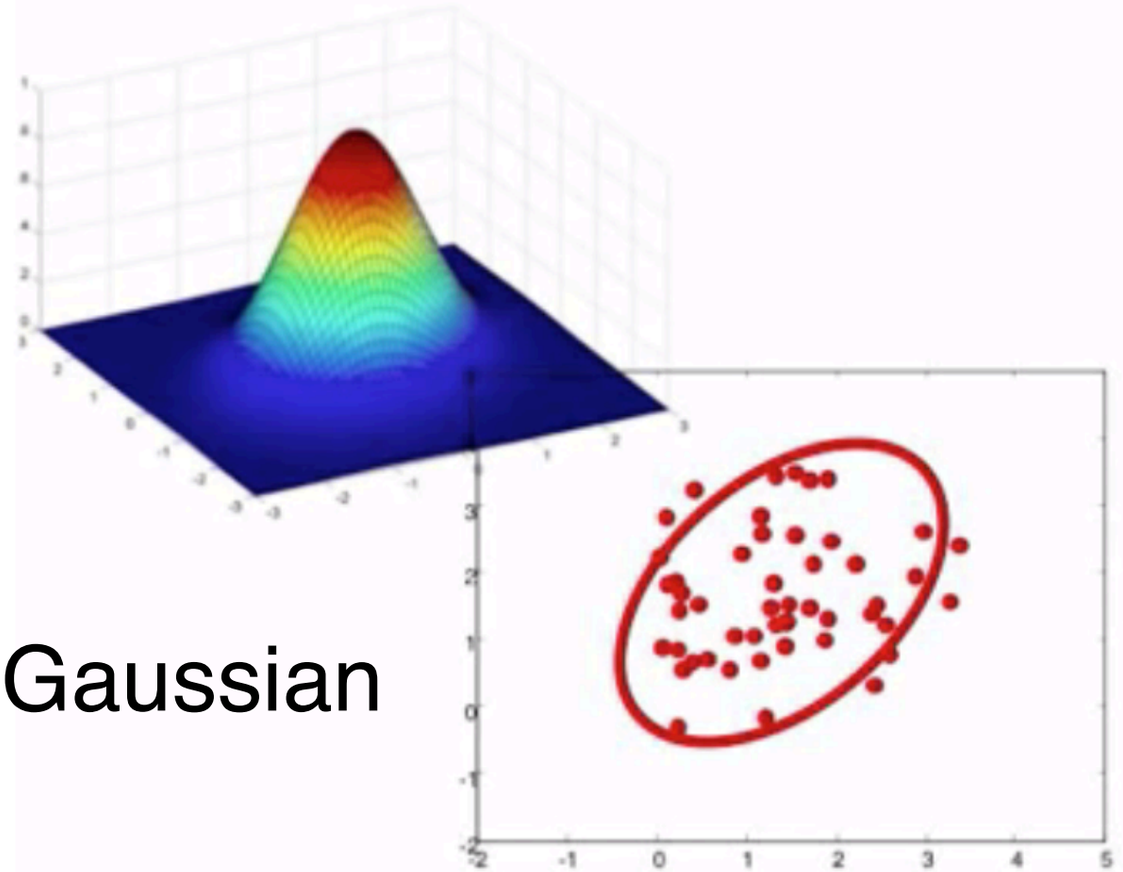
Multivariate Gaussian models

- Similar to univariate case

$$\mathcal{N}(\underline{x}\ ;\ \underline{\mu}, \Sigma) = \frac{1}{(2\pi)^{d/2}} |\Sigma|^{-1/2} \exp\left\{ -\frac{1}{2}(\underline{x} - \underline{\mu})\Sigma^{-1}(\underline{x} - \underline{\mu})^T \right\}$$

μ = length-d row vector
Σ = d x d matrix

$|\Sigma|$ = matrix determinant



2-D Gaussian

Maximum likelihood estimate:

$$\hat{\mu} = \frac{1}{m} \sum_j \underline{x}^{(j)}$$

$$\hat{\Sigma} = \frac{1}{m} \sum_j (\underline{x}^{(j)} - \underline{\hat{\mu}})^T (\underline{x}^{(j)} - \underline{\hat{\mu}})$$

(average of dxd matrices)

https://medium.com/@SeoJaeDuk/archived-post-multivariate-gaussian-distributions-and-entropy-3-991578ca534c

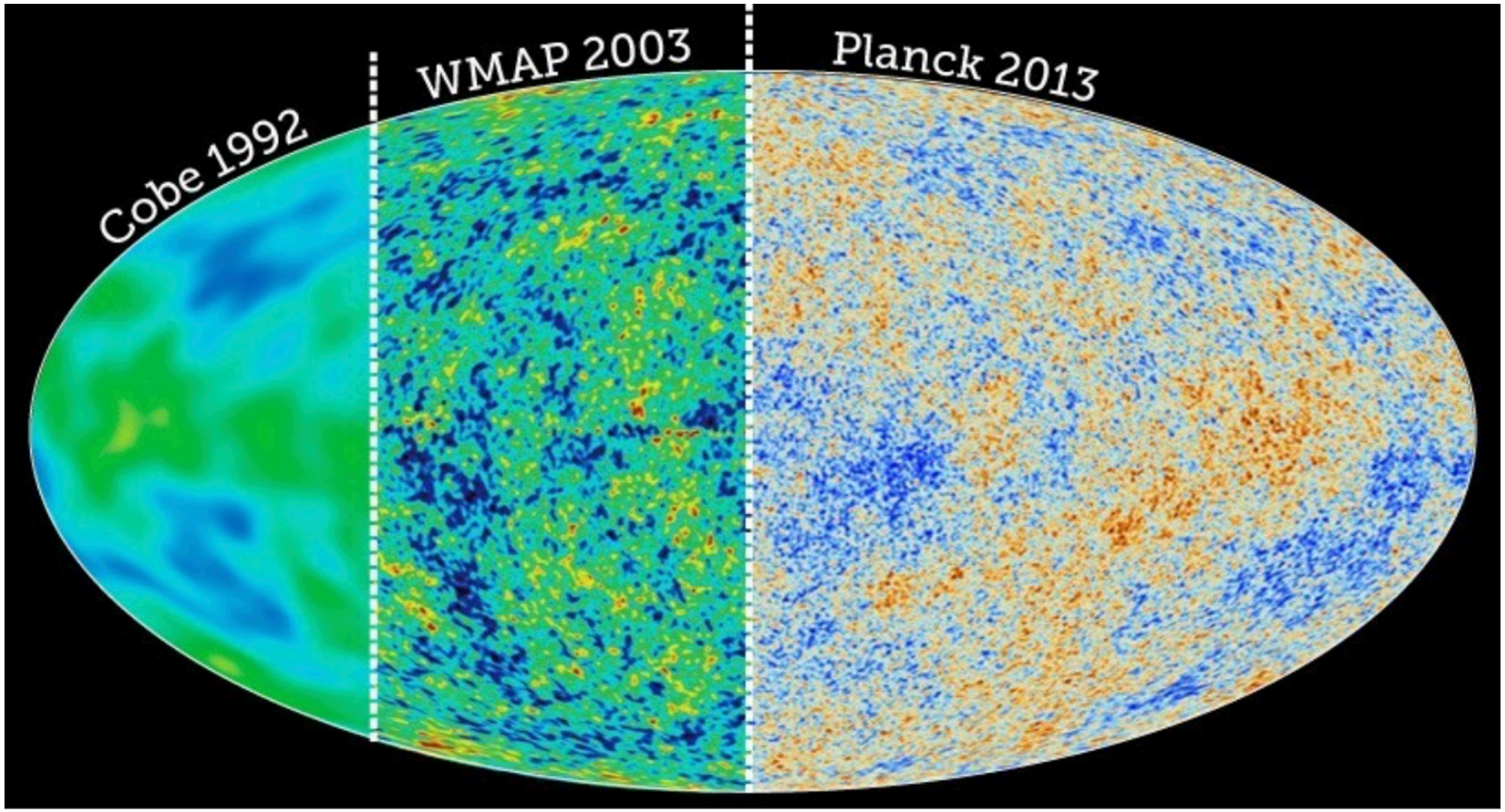
There is a correspondence between a covariance matrix and a Gaussian.

This can be done in high-dimension vector spaces like
- every pixel in an image
- every wavelength in a spectrum

The statistical “Gaussian process” that generates the data may be stationary (invariant under shifts and/or time)

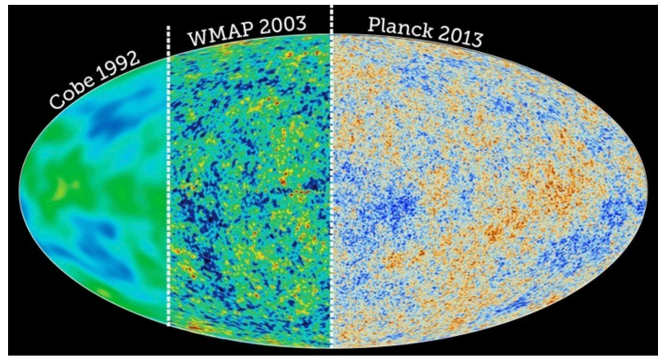
Or non-stationary.

The Cosmic Microwave Background (CMB) intensity corresponds to a stationary Gaussian process.

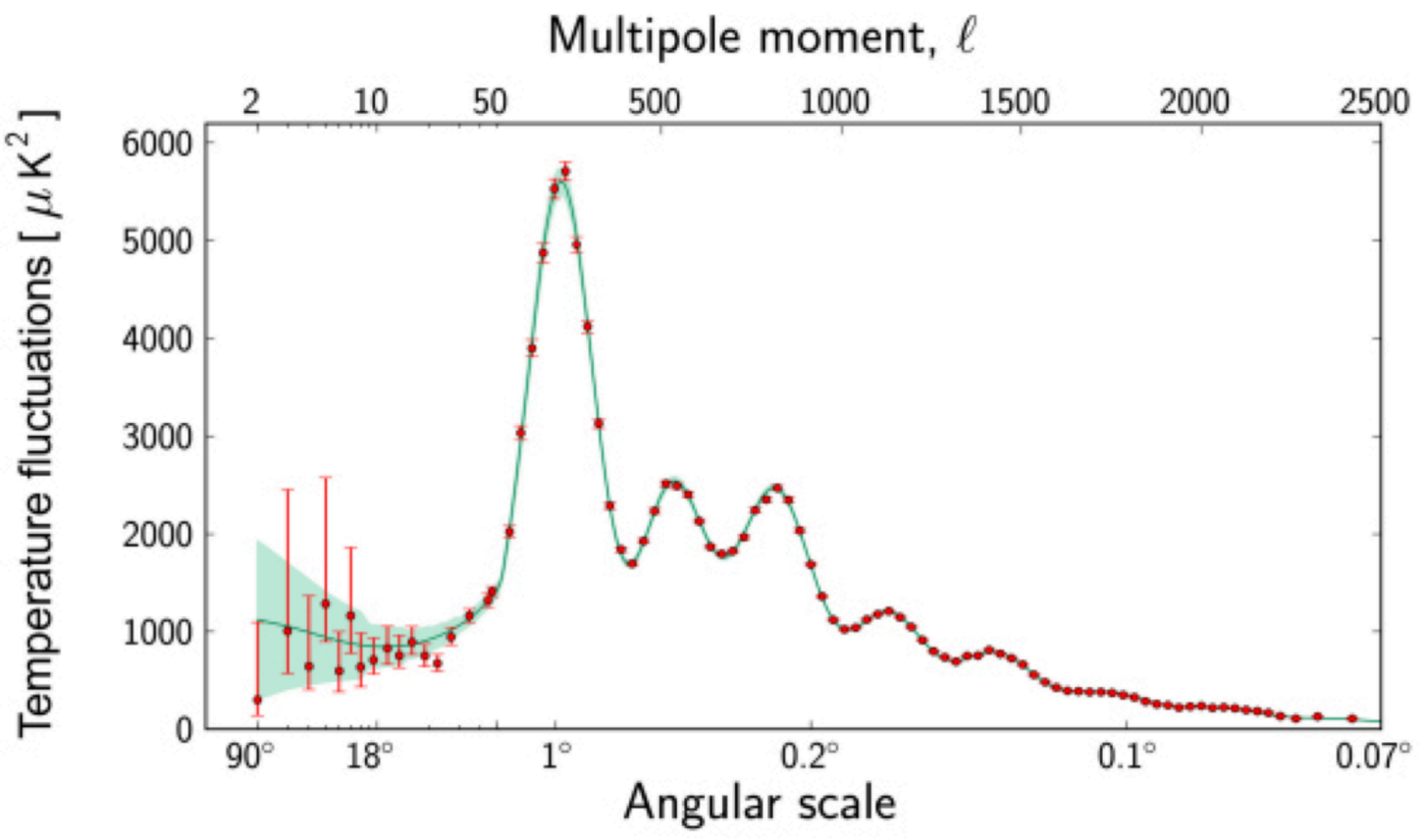


This Gaussian process is characterized by pixel-pixel covariance that depends only on *the angular distance between the pixels*.

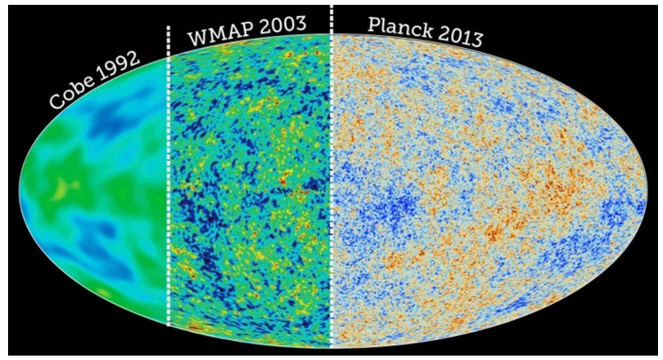
The CMB: A stationary Gaussian Process



The pixel-pixel covariance is captured by a 2-point correlation function, or (related via a Fourier transform) the power spectrum.



The CMB: A stationary Gaussian Process



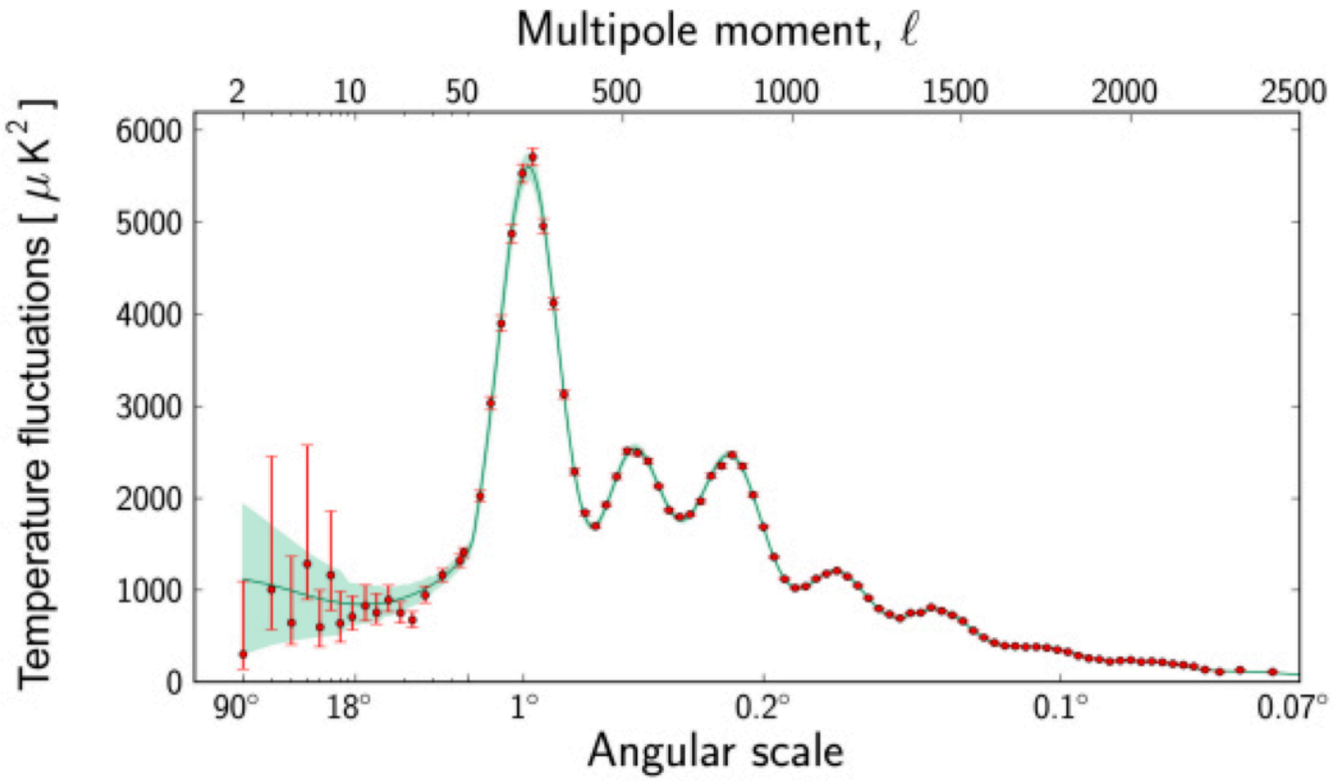
This is *not* a Gaussian process merely because the measurement noise in each pixel is Gaussian.

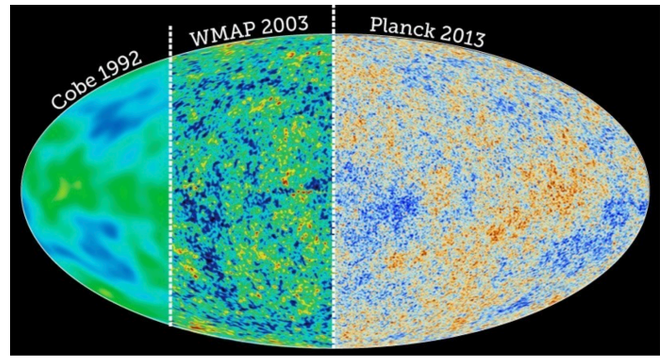
It is because the *spatial covariance structure* of the map is Gaussian.

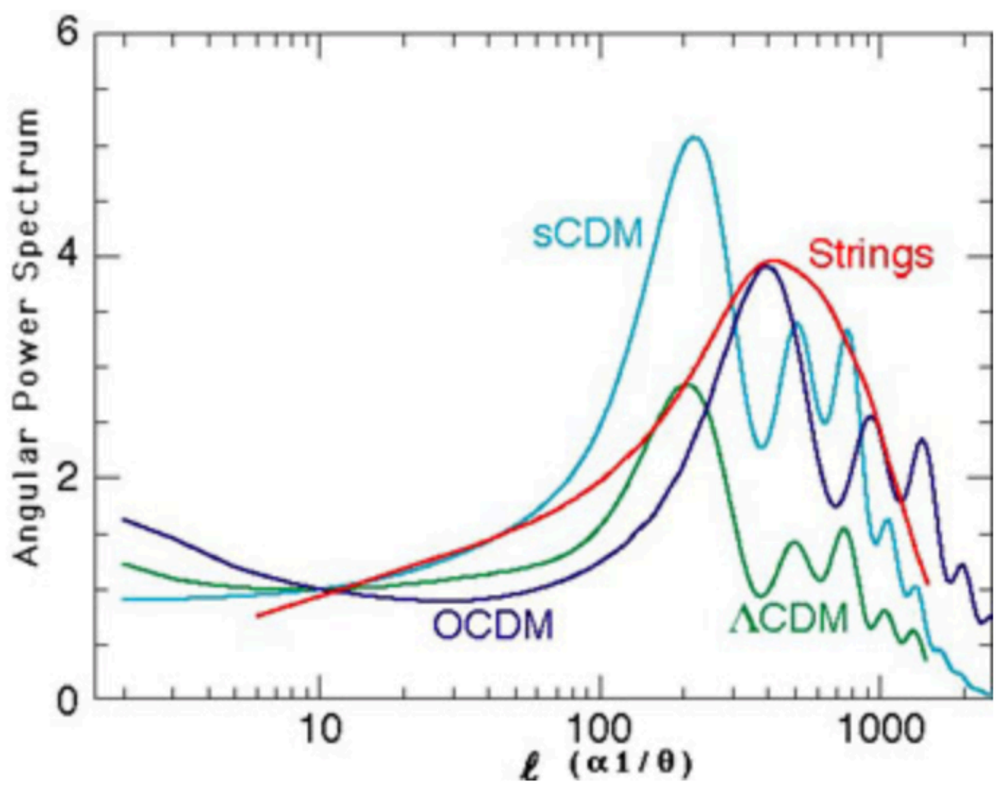
Therefore, the power spectrum contains "everything you need to know."

It is where observation meets theory.

We need a space where theory and observation meet…







Question 1: What do we mean by non-Gaussianity?

The information not captured by a power spectrum (or 2-point correlation function)

Question 2: What kinds of systems exhibit non-Gaussianity?

Subtle non-Gaussianity: systems that are *almost* Gaussian, but subtle deviations are interesting. Inflation, novel theories of gravity, "stringy physics"…

- CMB
- Large-scale structure
- …

Strong non-Gaussianity: systems with non-linear couplings that dominate the dynamics.

- Hydrodynamics
- Interstellar dust
- Basically everything else in the universe

Astro2020 Science White Paper

Primordial Non-Gaussianity

Thematic Areas: Cosmology and Fundamental Physics

Principal Author:
Name: P. Daniel Meerburg

Lots of people are interested in these topics…

https://arxiv.org/pdf/1903.04409.pdf

Authors/Endorsers[a]**:**
Daniel Green[1], Muntazir Abidi[2], Mustafa A. Amin[3], Peter Adshead[4], Zeeshan Ahmed[5], David Alonso[6], Behzad Ansarinejad[7], Robert Armstrong[8], Santiago Avila[9], Carlo Baccigalupi[10,11,12], Tobias Baldauf[2], Mario Ballardini[13], Kevin Bandura [14,15], Nicola Bartolo[16], Nicholas Battaglia[17], **Daniel Baumann**[18], Chetan Bavdhankar[19], José Luis Bernal[20,21], Florian Beutler[22], Matteo Biagetti[18], Colin Bischoff[23], Jonathan Blazek[24,25], J. Richard Bond[26], Julian Borrill[27], François R. Bouchet[28], Philip Bull[29], Cliff Burgess[30], Christian Byrnes[31], Erminia Calabrese[32], John E. Carlstrom[33,34,35], Emanuele Castorina[36], Anthony Challinor[37,2,38], Tzu-Ching Chang[39], Jonás Chaves-Montero[35], Xingang Chen[40], Christophe Yèche[41], Asantha Cooray[42], William Coulton[38,37], Thomas Crawford[33,34], **Elisa Chisari**[43], Francis-Yan Cyr-Racine[44,45], Guido D'Amico [46], Paolo de Bernardis [47,48], Axel de la Macorra[49], Olivier Doré[39], Adri Duivenvoorden[50], Joanna Dunkley[51], **Cora Dvorkin**[44], Alexander Eggemeier [7], Stephanie Escoffier[52], Tom Essinger-Hileman[53], Matteo Fasiello[22], Simone Ferraro[27], **Raphael Flauger**[1], Andreu Font-Ribera[54], **Simon Foreman**[55], Oliver Friedrich[38], Juan García-Bellido[9], Martina Gerbino[56], Vera Gluscevic[57], **Garrett Goon**[2], Krzysztof M. Górski[39], Jon E. Gudmundsson[50], Nikhel Gupta [58], Shaul Hanany[59], Will Handley,[38,60], Adam J. Hawken[61], J. Colin Hill[62,63], Christopher M. Hirata[25], Renée Hložek[64,65], Gilbert Holder[4], Dragan Huterer[56], Marc Kamionkowski[66], Kirit S. Karkare[33,34], Ryan E. Keeley[67], William Kinney[68], Theodore Kisner[27], Jean-Paul Kneib[24], Lloyd Knox[69], Savvas M. Koushiappas[70], Ely D. Kovetz[71], Kazuya Koyama[22], Benjamin L'Huillier[67], Ofer Lahav[54], Massimiliano Lattanzi[72], **Hayden Lee**[44], Michele Liguori[16], **Marilena Loverde**[73], Mathew Madhavacheril[51], Juan Maldacena[62], **M.C. David Marsh**[74], Kiyoshi Masui[75], Sabino Matarrese[76], Liam McAllister[17], Jeff McMahon[56], Matthew McQuinn[77], Joel Meyers[78], Mehrdad Mirbabayi[79], **Azadeh Moradinezhad Dizgah**[44,80], Pavel Motloch[26], Suvodip Mukherjee[28], Julian B. Muñoz[44], Adam D. Myers[81], Johanna Nagy[64], Pavel Naselsky[82], Federico Nati[83], Newburgh[84], Alberto Nicolis[85], Michael D. Niemack[17] Gustavo Niz[86], Andrei Nomerotski[87], Lyman Page[51], **Enrico Pajer**[2], Hamsa Padmanabhan[26,88], Gonzalo A. Palma[89], Hiranya V. Peiris[54,50], Will J. Percival [90,91,30], Francesco Piacentni[47,48], **Guilherme L. Pimentel**[18], Levon Pogosian[92], Chanda Prescod-Weinstein[93], Clement Pryke[59], Giuseppe Puglisi[46,94], Benjamin Racine[40], Radek Stompor[95], Marco Raveri[34,33], Mathieu Remazeilles[96], Graça Rocha[39], Ashley J. Ross[97], Graziano Rossi[98], John Ruhl[99], Misao Sasaki[100], Emmanuel Schaan[27,101], Alessandro Schillaci[102], Marcel Schmittfull[62], Neelima Sehgal[103], Leonardo Senatore[94], Hee-Jong Seo[104], Huanyuan Shan[105], Sarah Shandera[106], Blake D. Sherwin[2,38], **Eva Silverstein**[46], Sara Simon[56], **Anže Slosar**[87], Suzanne Staggs[51], Glenn Starkman[99], Albert Stebbins[107], Aritoki Suzuki[27], Eric R. Switzer[53], Peter Timbie[108], Andrew J. Tolley[109], Maurizio Tomasi[110], Matthieu Tristram[111], Mark Trodden[83], Yu-Dai Tsai[107], Cora Uhlemann[2], Caterina Umiltà[23], Alexander van Engelen[26], M. Vargas-Magaña[49], Abigail Vieregg[33], **Benjamin Wallisch**[62,1], David Wands[22], **Benjamin Wandelt**[28], Yi Wang[112], Scott Watson[113], Mark Wise[102], W. L. K. Wu[34], Zhong-Zhi Xianyu[44], Weishuang Xu[44], Siavash Yasini[114], Sam Young[115], Duan Yutong[116], Matias Zaldarriaga[62], Michael Zemcov[117], Gong-Bo Zhao[118,22], Yi Zheng[119], Ningfeng Zhu[83]

[a]Names in bold indicate significant contribution.

Astro2020 Science White Paper

Primordial Non-Gaussianity

Thematic Areas: Cosmology and Fundamental Physics

Principal Author:
Name: P. Daniel Meerburg

- **Inflaton self-interactions** Non-gaussanity can arise from non-linear dynamics during single-field inflation. In the most well-studied case, these interactions also cause the fluctuations to propagate with a speed slower than the speed of light. Both a detection or an exclusion of such a signature provides a unique window into the mechanism behind inflation.
- **Additional light fields** Light degrees of freedom are excited from the vacuum with an amplitude set by the Hubble scale. When this degree of freedom is not the inflaton, these fluctuations freeze-out and describe isocurvature (entropy) fluctuations. These isocurvature modes may eventually convert into isocurvature perturbations, during inflation or reheating. These conversion processes induce correlations between modes that are necessarily non-Gaussian.
- **Additional heavy fields** Heavy degrees of freedom (e.g. particles with mass on the order of the Hubble scale during inflation, or larger) are excited during inflation but are diluted quickly after horizon crossing. However, when the inflaton couples to these additional degrees of freedom, their fluctuations can still correlate the adiabatic modes producing non-Gaussianity.

https://arxiv.org/pdf/1903.04409.pdf

These are important questions that many people are working to address. My group tends to focus on interstellar dust.

We see it in absorption

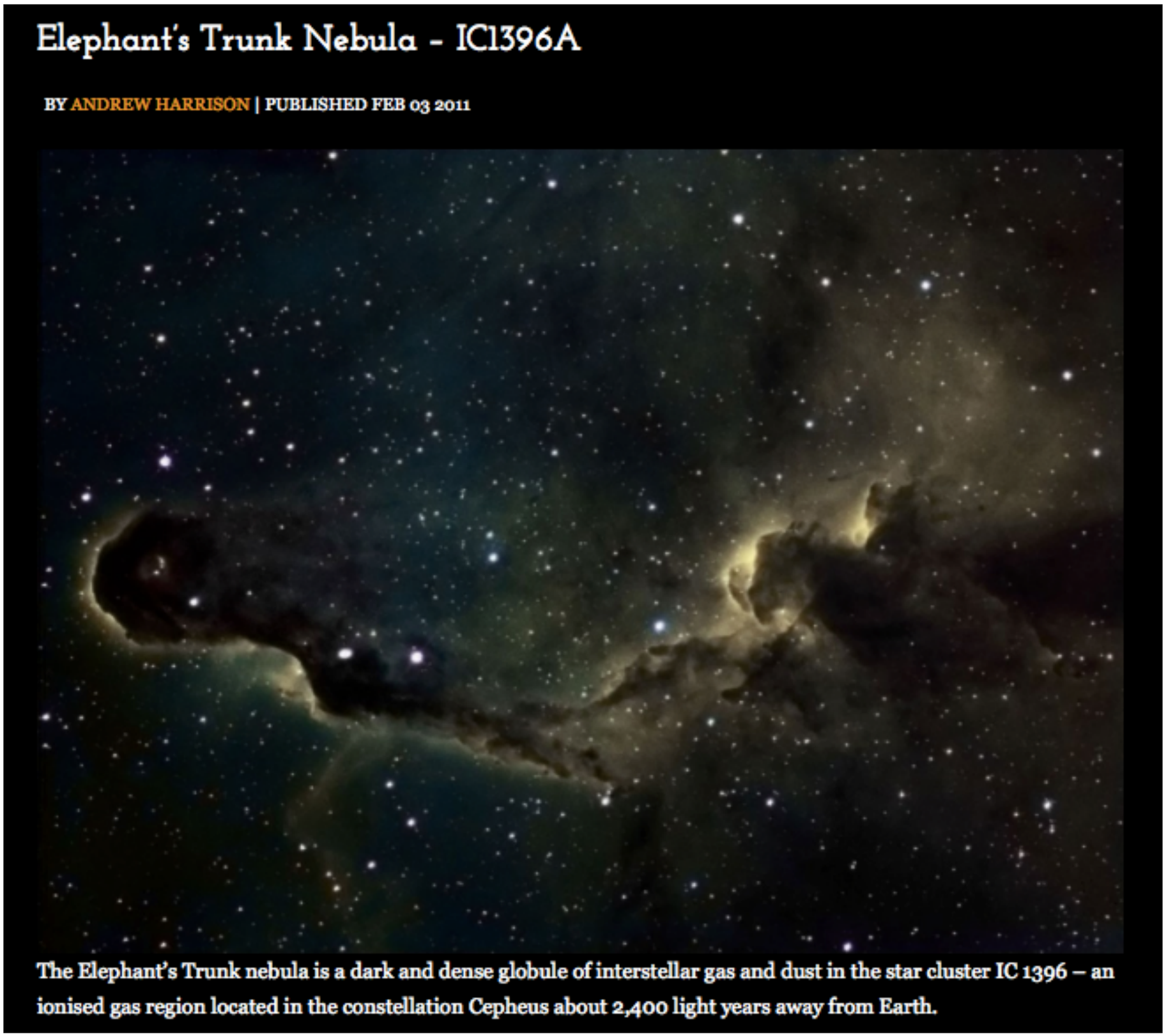
The "Black Cloud" B68
(VLT ANTU + FORS1)

ESO PR Photo 20a/99 (30 April 1999)

… and reflection

Rigel and the Witch Head Nebula
Image Credit & Copyright: Rogelio Bernal Andreo (Deep Sky Colors)

… and sometimes both.



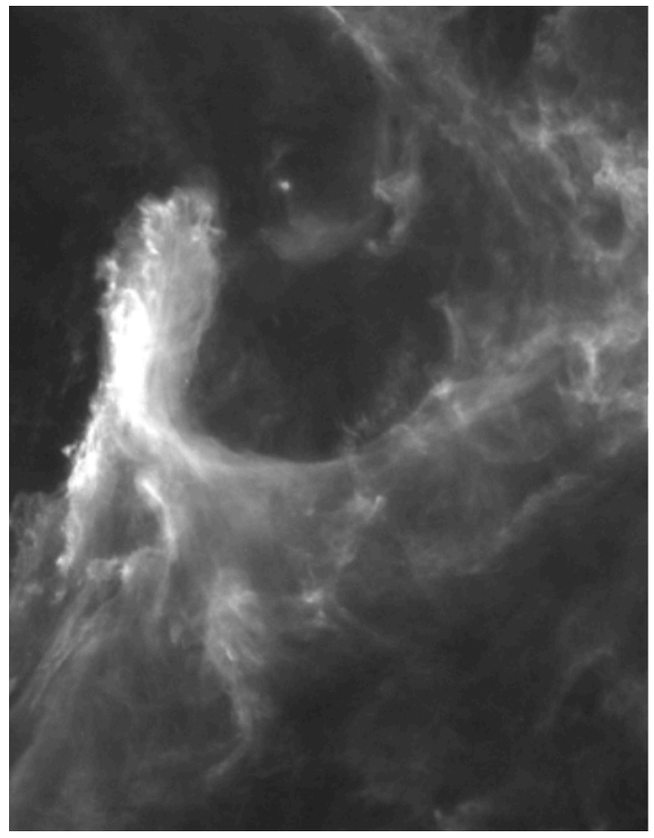
It has been mapped in exquisite detail in 2-D,
and more recently in 3-D.

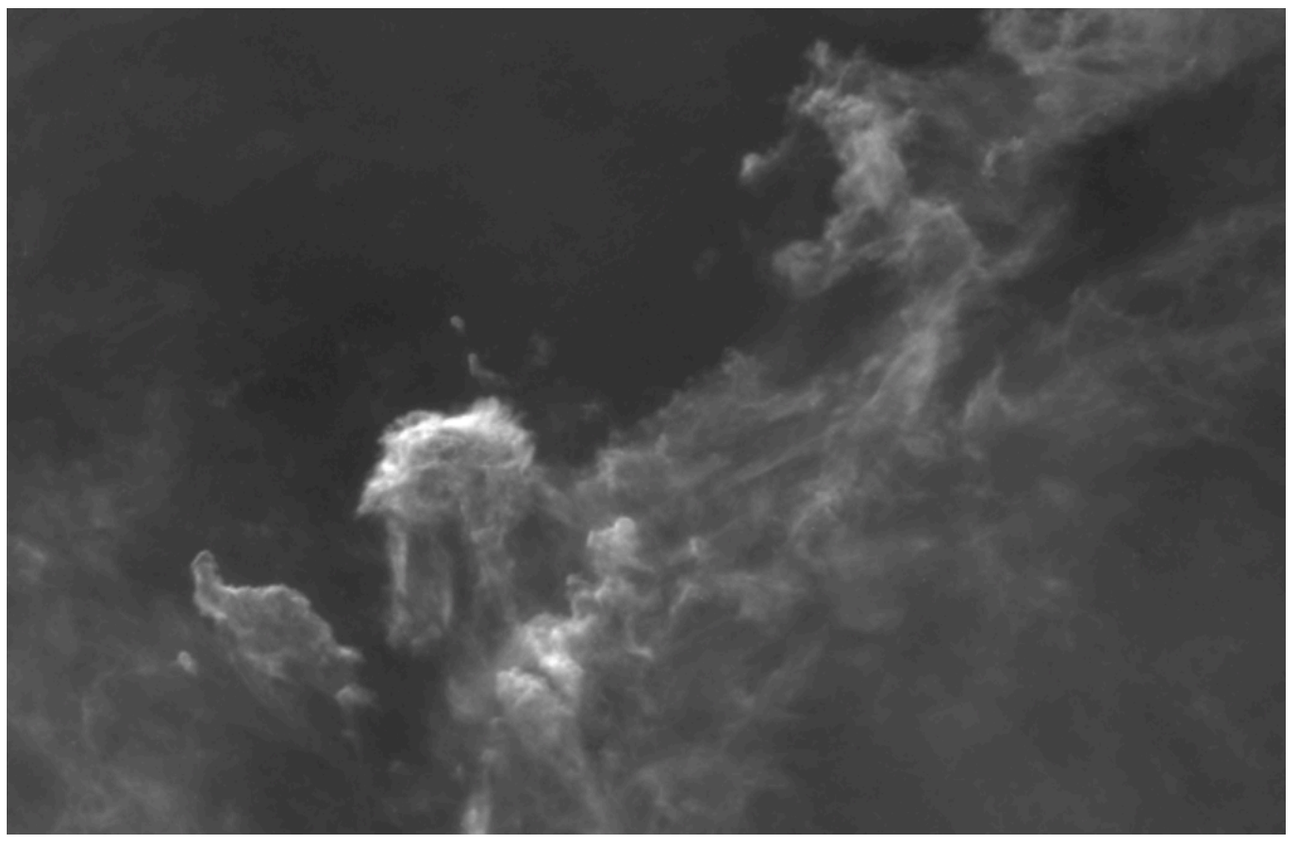
But what do we want to learn from it?

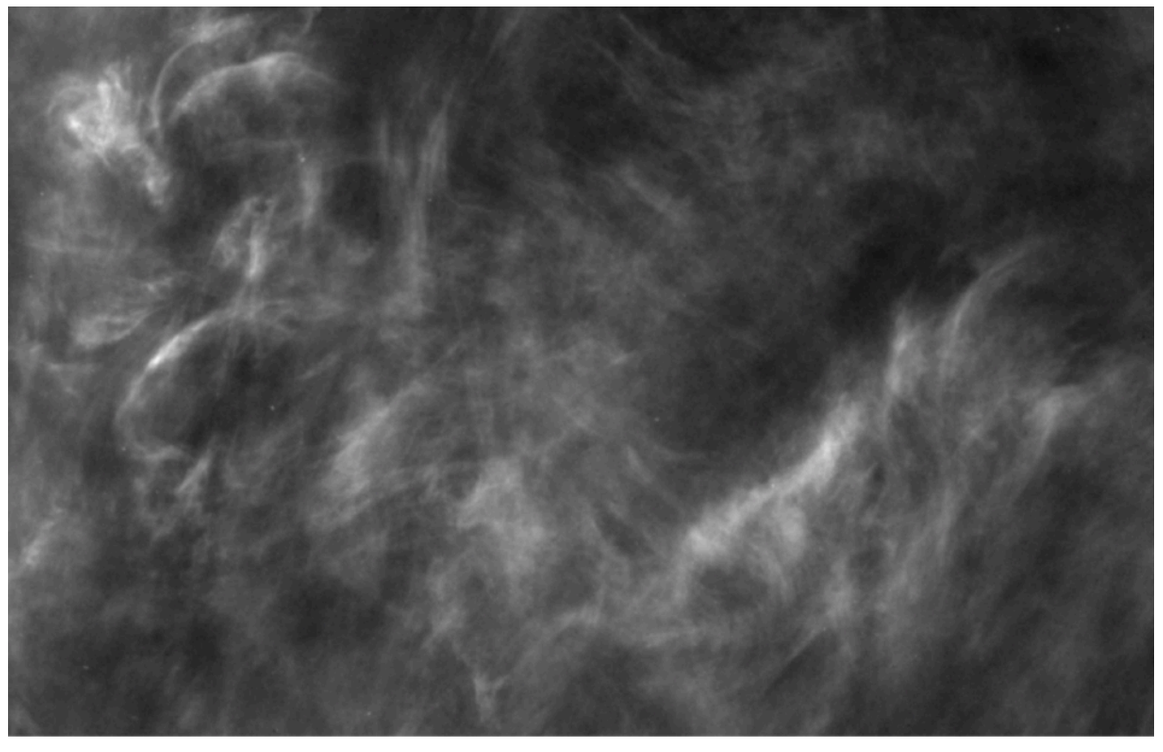
What are the physical processes happening?
What are the parameters?
(Density, B-Field, Mach numbers, etc.)

Can we do the same for the interstellar medium?

It has “filaments”
“Phase coherence”
“Non-Gaussian structure”







WSSA, Meisner & Finkbeiner (2015), see wssa.skymaps.info

We would like a statistic that is

- translation invariant
- rotation invariant
- robust to projection
- robust to noise
- robust to differences between telescopes (PSF, etc.)
- robust to different spectral regimes (e.g. comparing dust emission with HI or CO)

Why not CNNs?
- quantity and quality of training data
- want to understand in terms of physics

-> Middle-ground between CNN and Gaussian statistics

Classic paper on wavelet scattering transform (WST)

Invariant Scattering Convolution Networks

Joan Bruna and Stéphane Mallat
CMAP, Ecole Polytechnique, Palaiseau, France

Abstract—A wavelet scattering network computes a translation invariant image representation, which is stable to deformations and preserves high frequency information for classification. It cascades wavelet transform convolutions with non-linear modulus and averaging operators. The first network layer outputs SIFT-type descriptors whereas the next layers provide complementary invariant information which improves classification. The mathematical analysis of wavelet scattering networks explain important properties of deep convolution networks for classification.

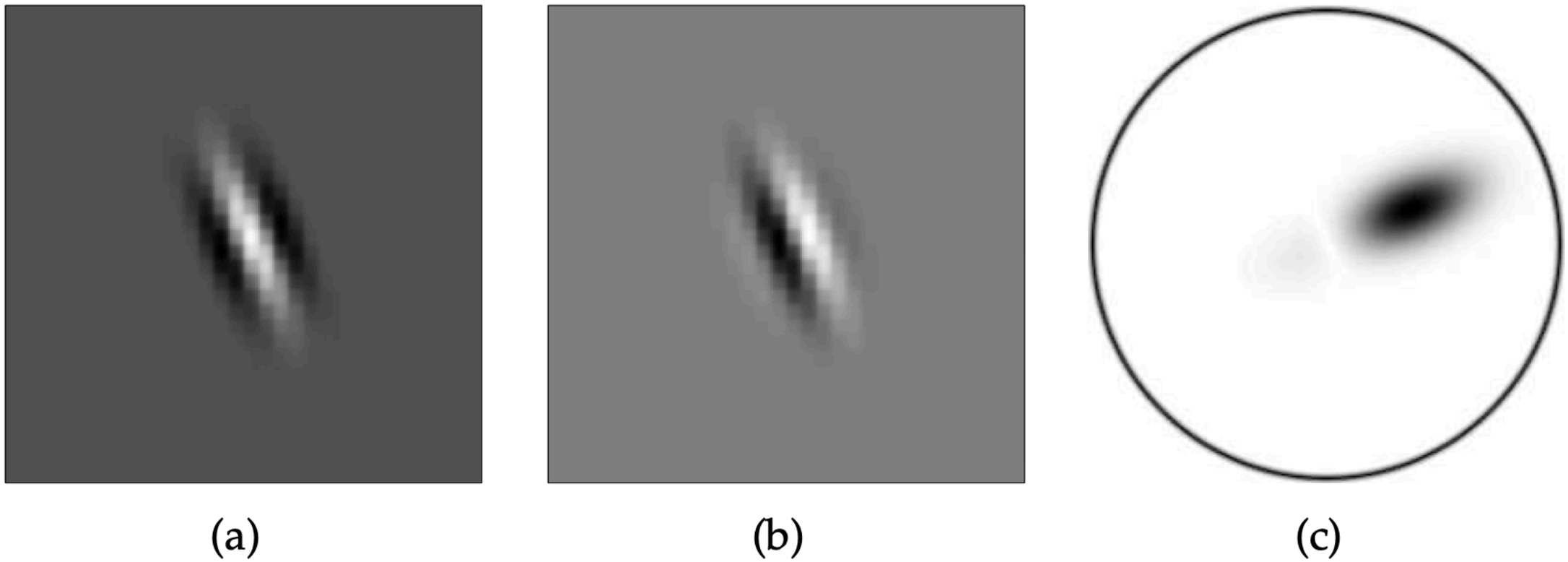
A scattering representation of stationary processes incorporates higher order moments and can thus discriminate textures having same Fourier power spectrum. State of the art classification results are obtained for handwritten digits and texture discrimination, with a Gaussian kernel SVM and a generative PCA classifier.

Bruna & Mallat, 1203.1513v2

The Morlet Wavelet

A complex sinusoidal wave windowed by a Gaussian

$$\exp(i\vec{k}\cdot\vec{x})\,G(\vec{x})$$



(a) (b) (c)

(a): Real part of ψ. (b): Imaginary part of ψ. (c): Fourier modulus $|\hat{\psi}|$.

Bruna & Mallat, 1203.1513v2

The Wavelet Scattering Transform

The wavelet, $\psi_{j,l}(\vec{x})$, is indexed on j,l

j: log(size) (e.g. 6 scales for 256x256 image)

l: angle (e.g. 8 angles in half-plane)

Convolve input image, I, with all wavelets.

Take the modulus

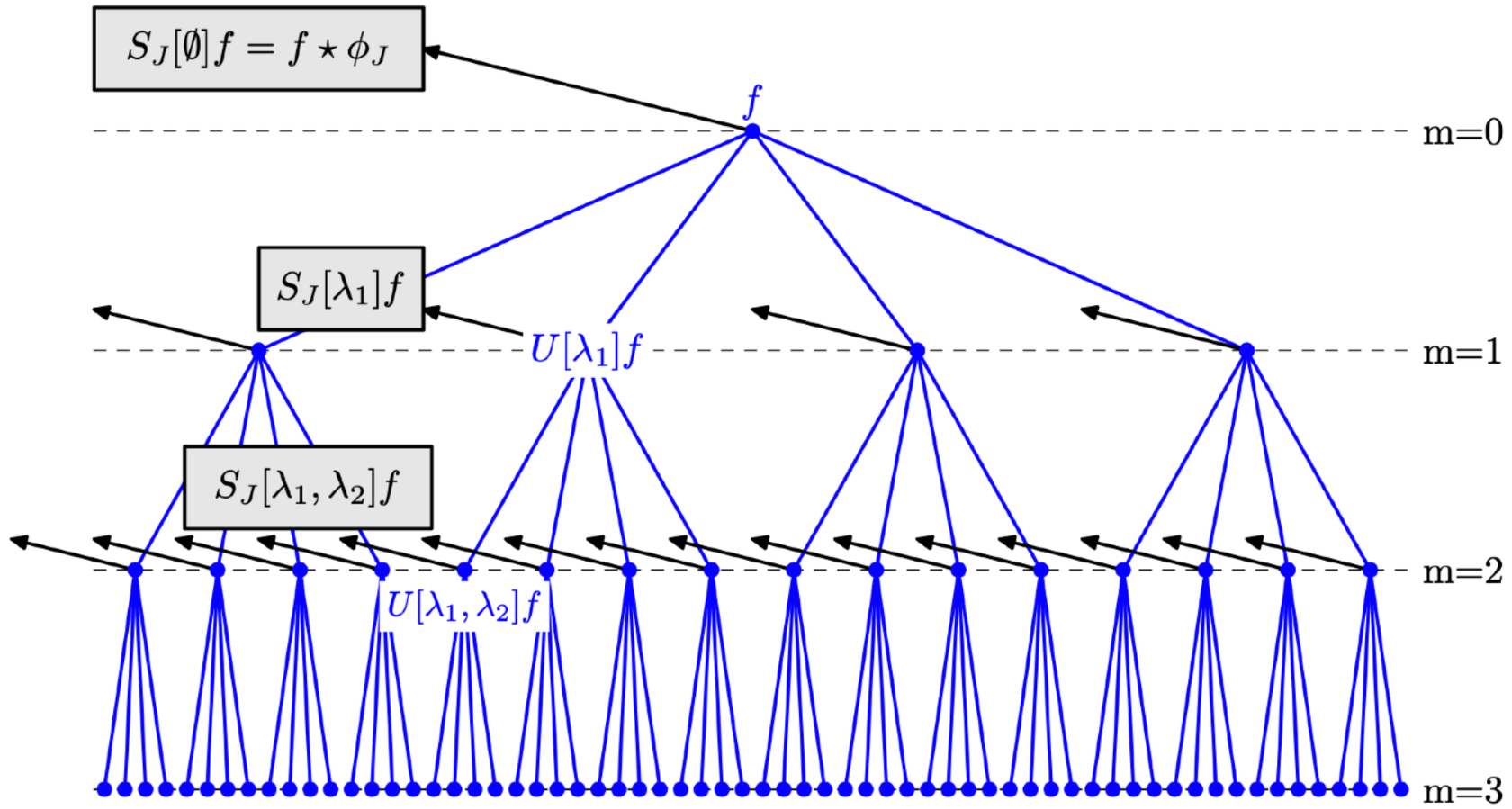
Repeat m times… (usually m=2)

$$S_1(j_1, l_1) = \frac{1}{\mu_1} \int |I \star \psi_{j_1,l_1}|(\vec{x})\ d^2\vec{x}$$

$$S_2(j_1, l_1, j_2, l_2) = \frac{1}{\mu_2} \int ||I \star \psi_{j_1,l_1}| \star \psi_{j_2,l_2}|(\vec{x})\ d^2\vec{x}$$

Bruna & Mallat, 1203.1513v2

The Wavelet Scattering Transform



$\lambda_1 \equiv (j_1, l_1)$

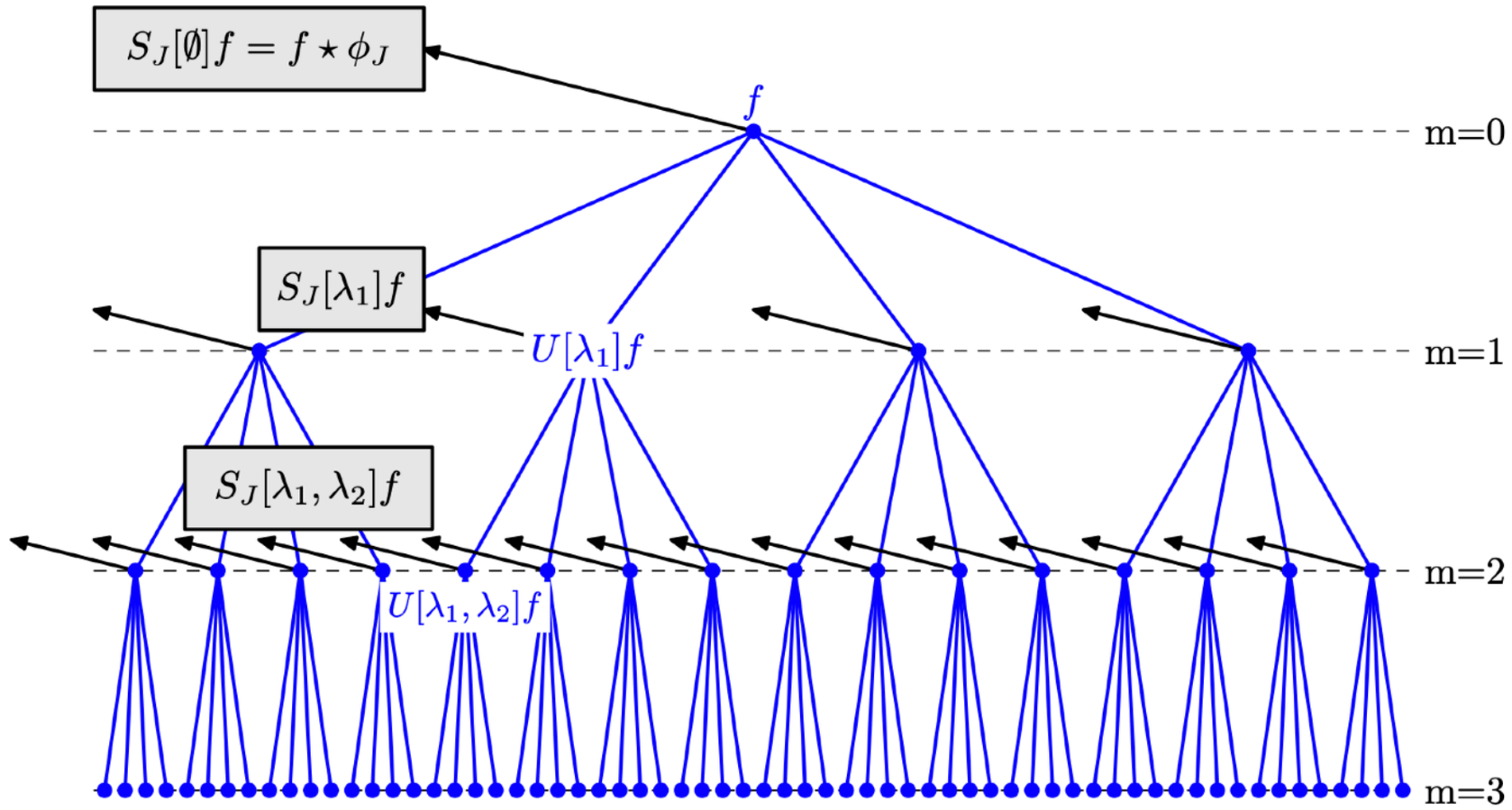
$\lambda_2 \equiv (j_2, l_2)$

$$S_1(j_1, l_1) = \frac{1}{\mu_1} \int |I \star \psi_{j_1,l_1}|(\vec{x})\ d^2\vec{x}$$

$$S_2(j_1, l_1, j_2, l_2) = \frac{1}{\mu_2} \int ||I \star \psi_{j_1,l_1}| \star \psi_{j_2,l_2}|(\vec{x})\ d^2\vec{x}$$

Bruna & Mallat, 1203.1513v2

The Wavelet Scattering Transform



$\lambda_1 \equiv (j_1, l_1)$

$\lambda_2 \equiv (j_2, l_2)$

$$S_1(j_1, l_1) = \frac{1}{\mu_1} \int |I \star \psi_{j_1,l_1}|(\vec{x})\ d^2\vec{x}$$

~ power spectrum

$$S_2(j_1, l_1, j_2, l_2) = \frac{1}{\mu_2} \int ||I \star \psi_{j_1,l_1}| \star \psi_{j_2,l_2}|(\vec{x})\ d^2\vec{x}$$

Non-linear coupling of various scales and angles

Bruna & Mallat, 1203.1513v2

WST / CNN similarity

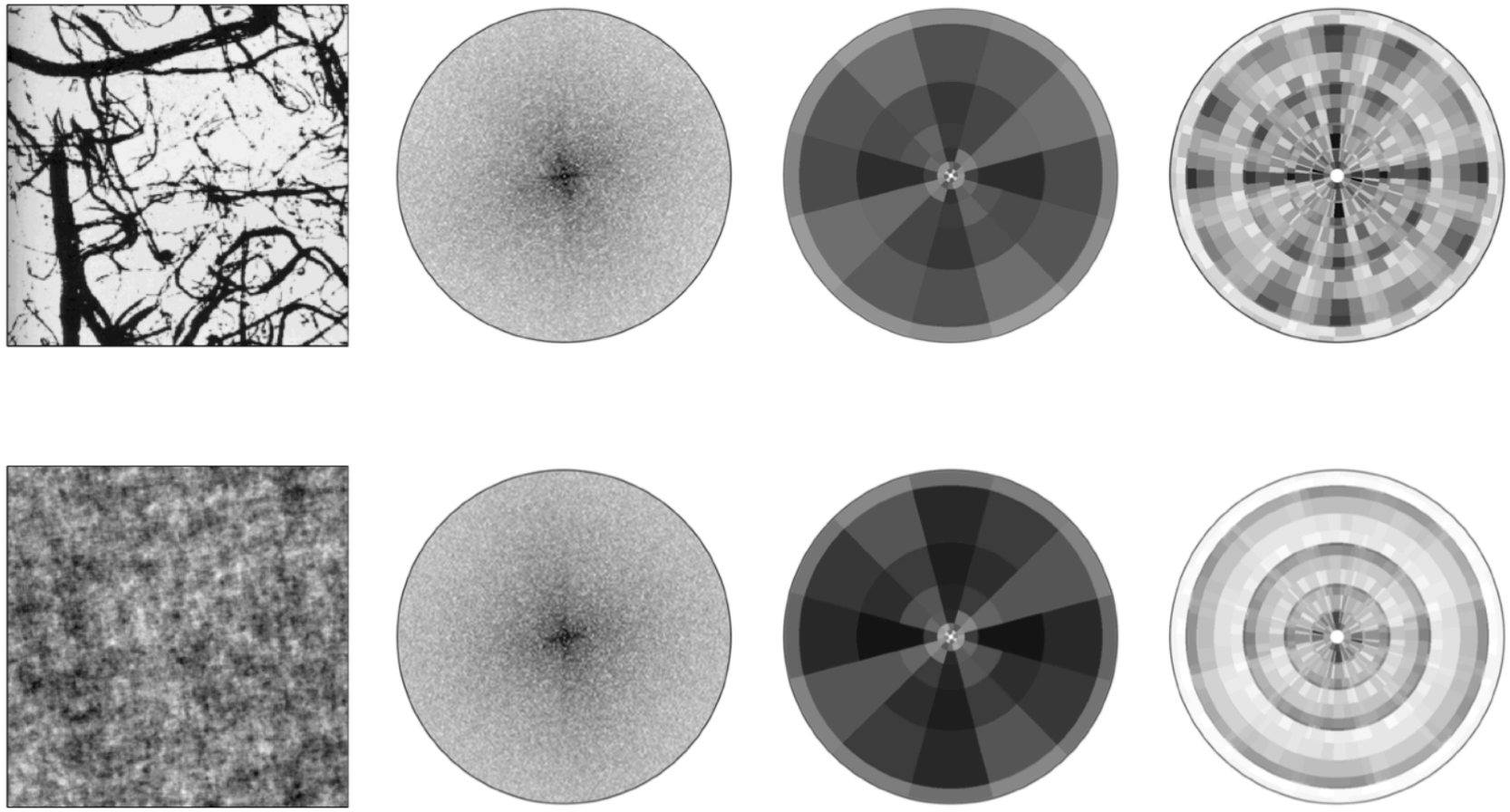
The WST bears some resemblance to a convolutional neural network (CNN). Both involve some notion of

1. Convolve with filters
2. Put through a non-linear function; output the sum
3. Convolve each image in (2) with each filter
4. Repeat *M* times…

Unlike a CNN, WST sees all size scales at each step.
Instead of a sigmoid or Relu, *modulus* is the non-linear function.

A CNN must *learn* its filters. For WST the filters are fixed.

WST is sensitive to non-Gaussian texture



(a) (b) (c) (d)

Fig. 5. Two different textures having the same Fourier power spectrum. (a) Textures $X(u)$. Top: Brodatz texture. Bottom: Gaussian process. (b) Same estimated power spectrum $\widehat{R}X(\omega)$. (c) Nearly same scattering coefficients $S_J[p]X$ for $m = 1$ and 2^J equal to the image width. (d) Different scattering coefficients $S_J[p]X$ for $m = 2$.

Bruna & Mallat, 1203.1513v2

WST is sensitive to non-Gaussian texture

This is brilliant…

We have a statistic that
- is fast and easy to compute
- is translation invariant (and somewhat rotation invariant)
- requires no "learning"
- S_1 is sensitive to the Power spectrum (i.e. Gaussian)
- S_2 is sensitive to non-Gaussianity

Now what?

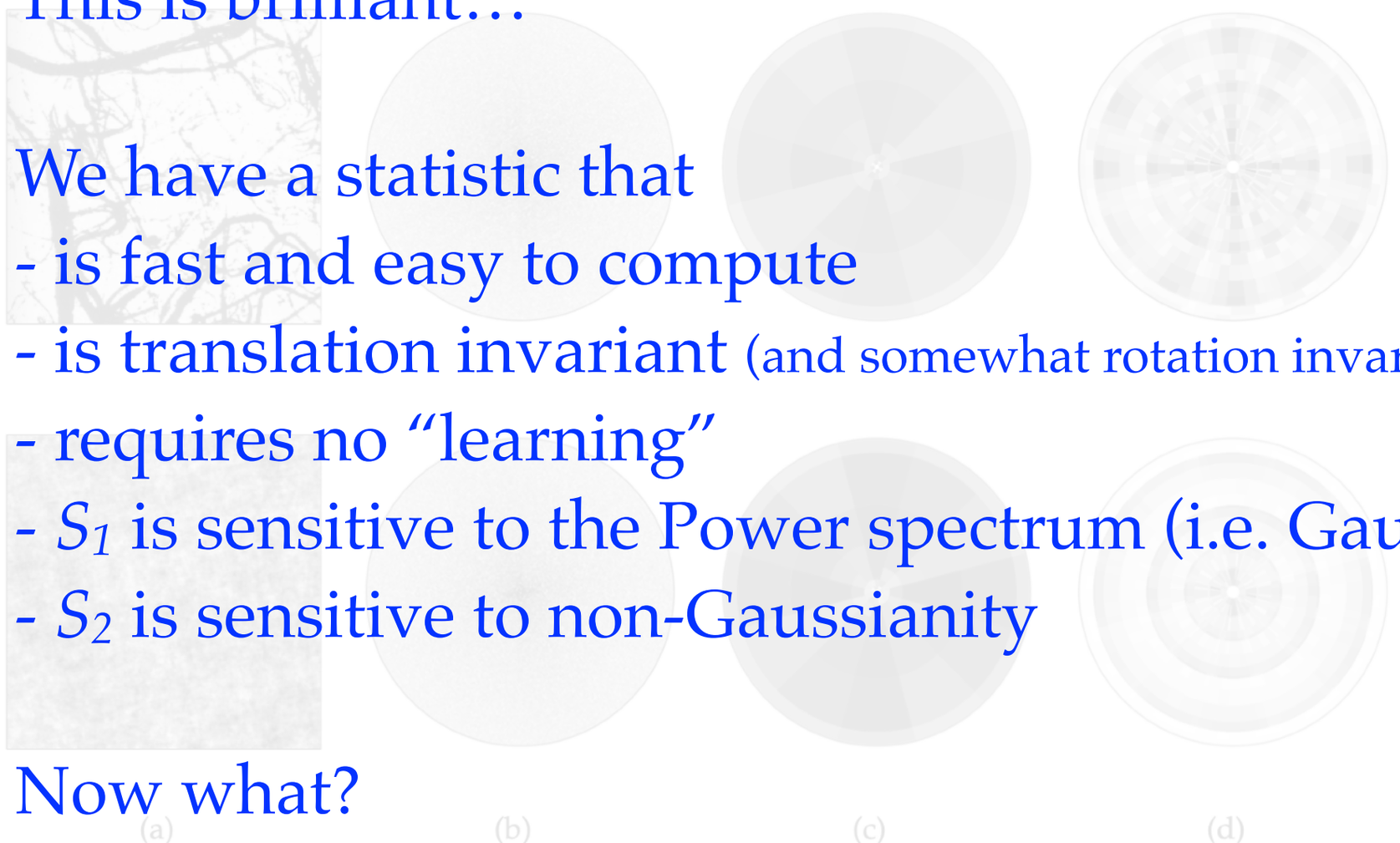
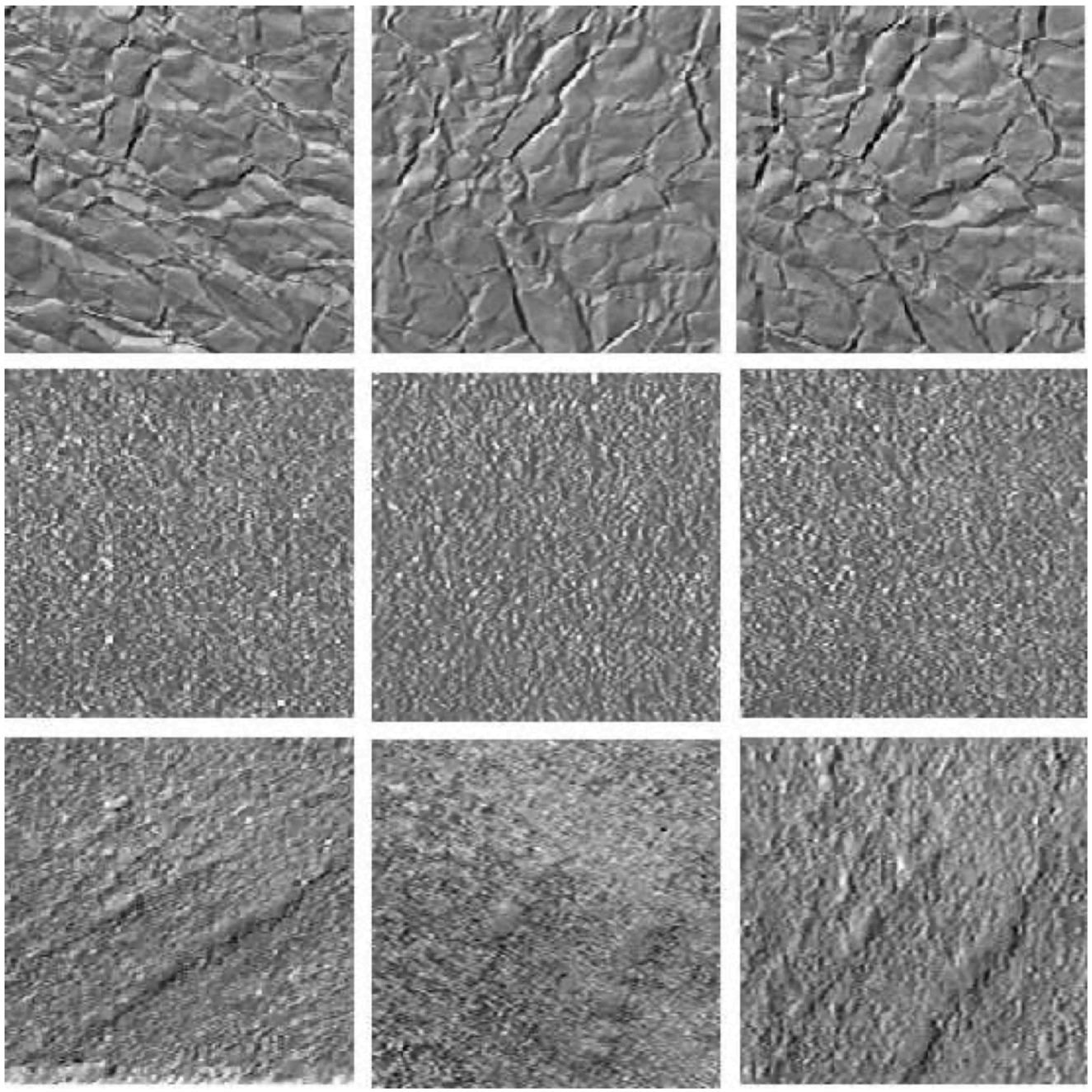


Fig. 5. Two different textures having the same Fourier power spectrum. (a) Textures $X(u)$. Top: Brodatz texture. Bottom: Gaussian process. (b) Same estimated power spectrum $\widehat{R}X(\omega)$. (c) Nearly same scattering coefficients $S_J[p]X$ for $m = 1$ and 2^J equal to the image width. (d) Different scattering coefficients $S_J[p]X$ for $m = 2$.

WST classifies textures



If WST can do this, can it classify ISM density fields?

Bruna & Mallat, 1203.1513v2

ISM dust is non-Gaussian

Let's apply WST to these images and see how they differ.

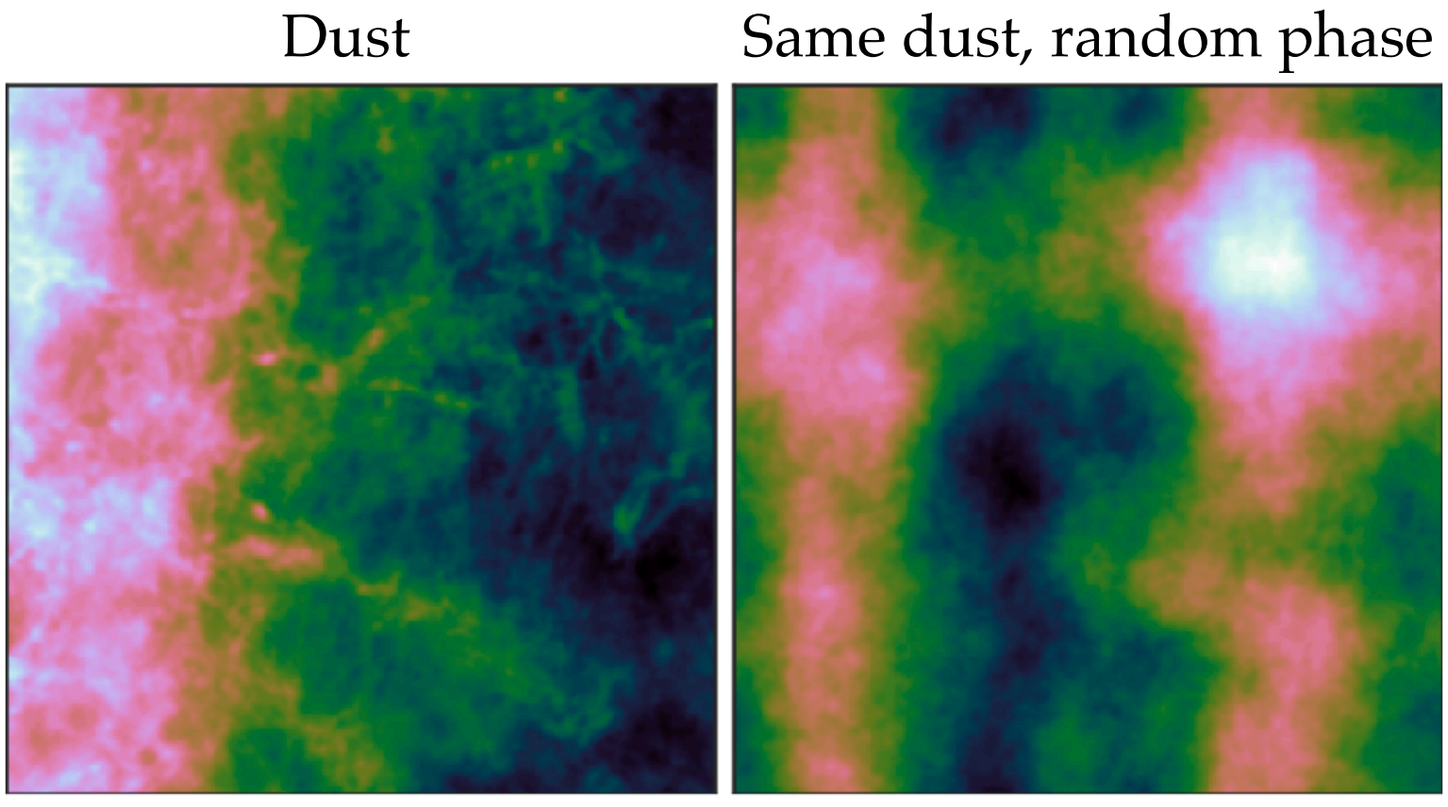


Figure 2. *Left*: Representative 256^2 image of the 2D SFD dust map (log color scale). *Right*: Same as left, but with random phase in the Fourier domain (linear color scale).

Saydjari+ 2010.11963

Dimensionality reduction

WST is clearly sensitive to something of interest, but outputs 100s or 1000s of parameters.

How do extract a few useful parameters?

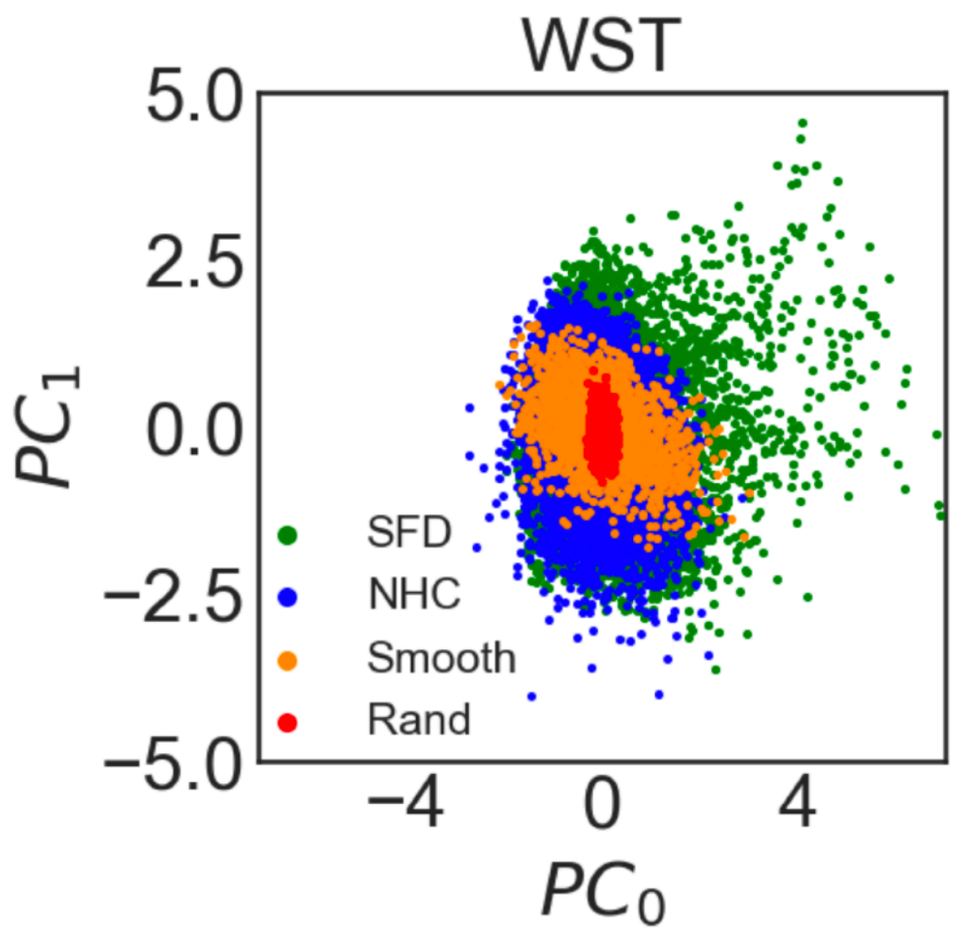
i.e., which dimensions in that vector space are useful?

PCA is not helpful

Consider 4 input image classes:

- SFD: Schlegel+ 1998 dust map
- NHC: same with random phase
- Smooth: smoothed white noise
- Rand: white noise

Principal Component Analysis (PCA) cannot separate these.



PCA identifies a linear basis containing the most variance.
That is *not* what we want.

Saydjari+ 2010.11963

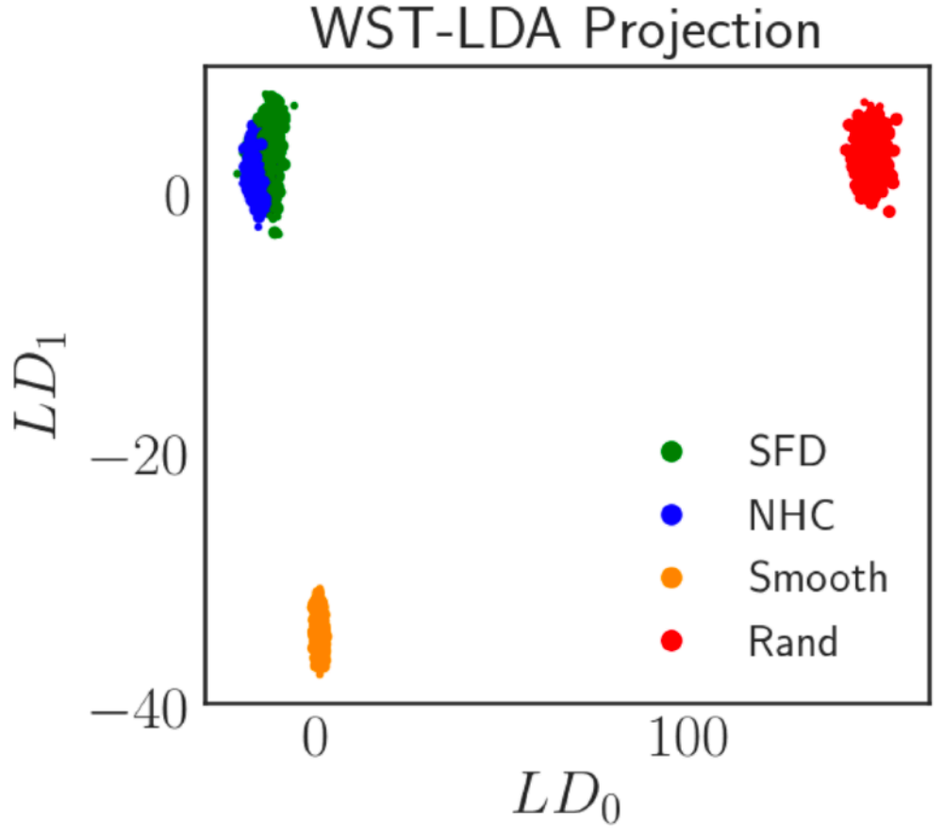
LDA works better

Linear Discriminant Analysis (LDA)

Finds a linear basis that maximizes separation of classes.

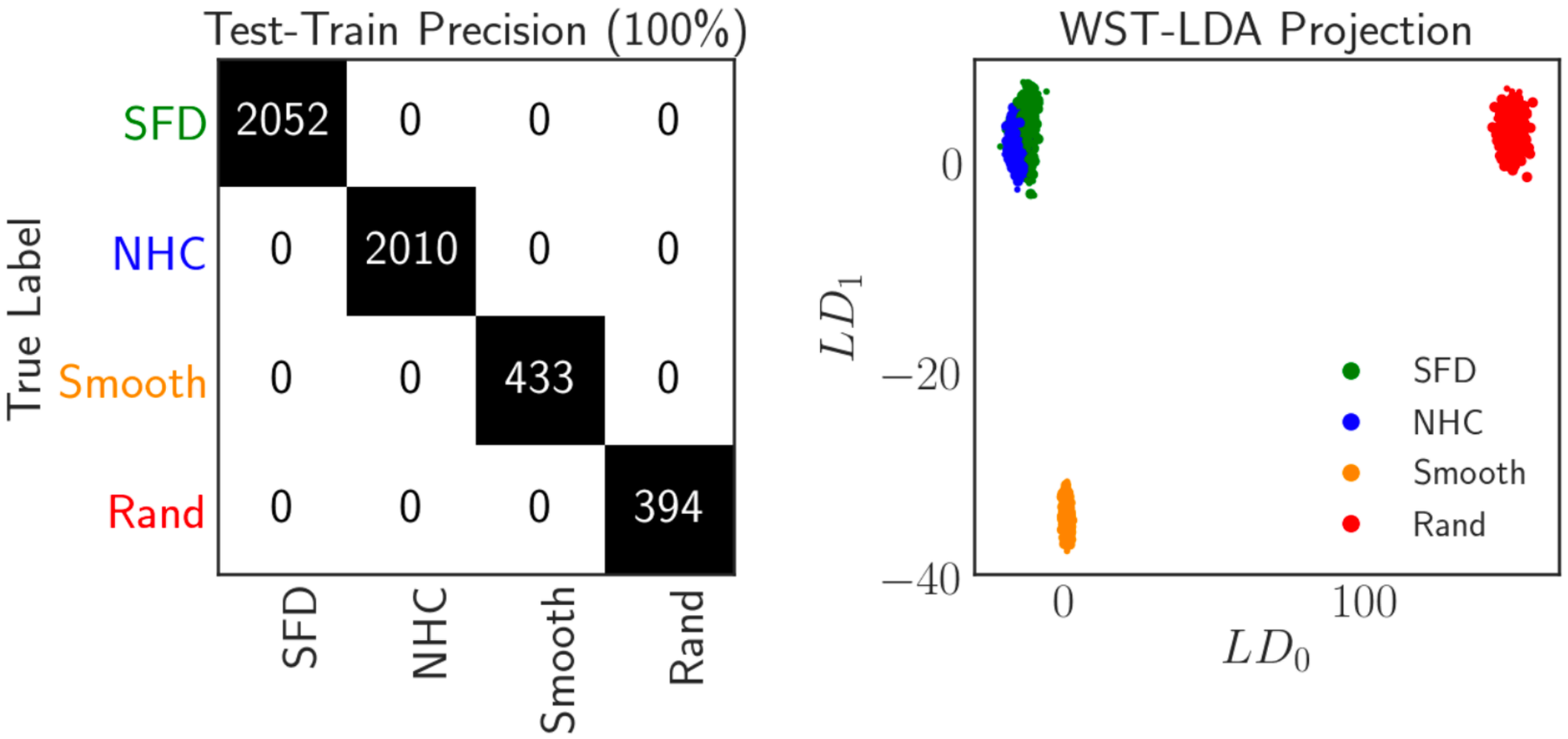
Supervised, not unsupervised like (PCA)

The 4 classes are easily separated.
(Remember there are more dimensions…)



Saydjari+ 2010.11963

WST-LDA



Saydjari+ 2010.11963

But that was too easy…

Let’s try a toy problem related to the interstellar medium.

Classification of MHD sims

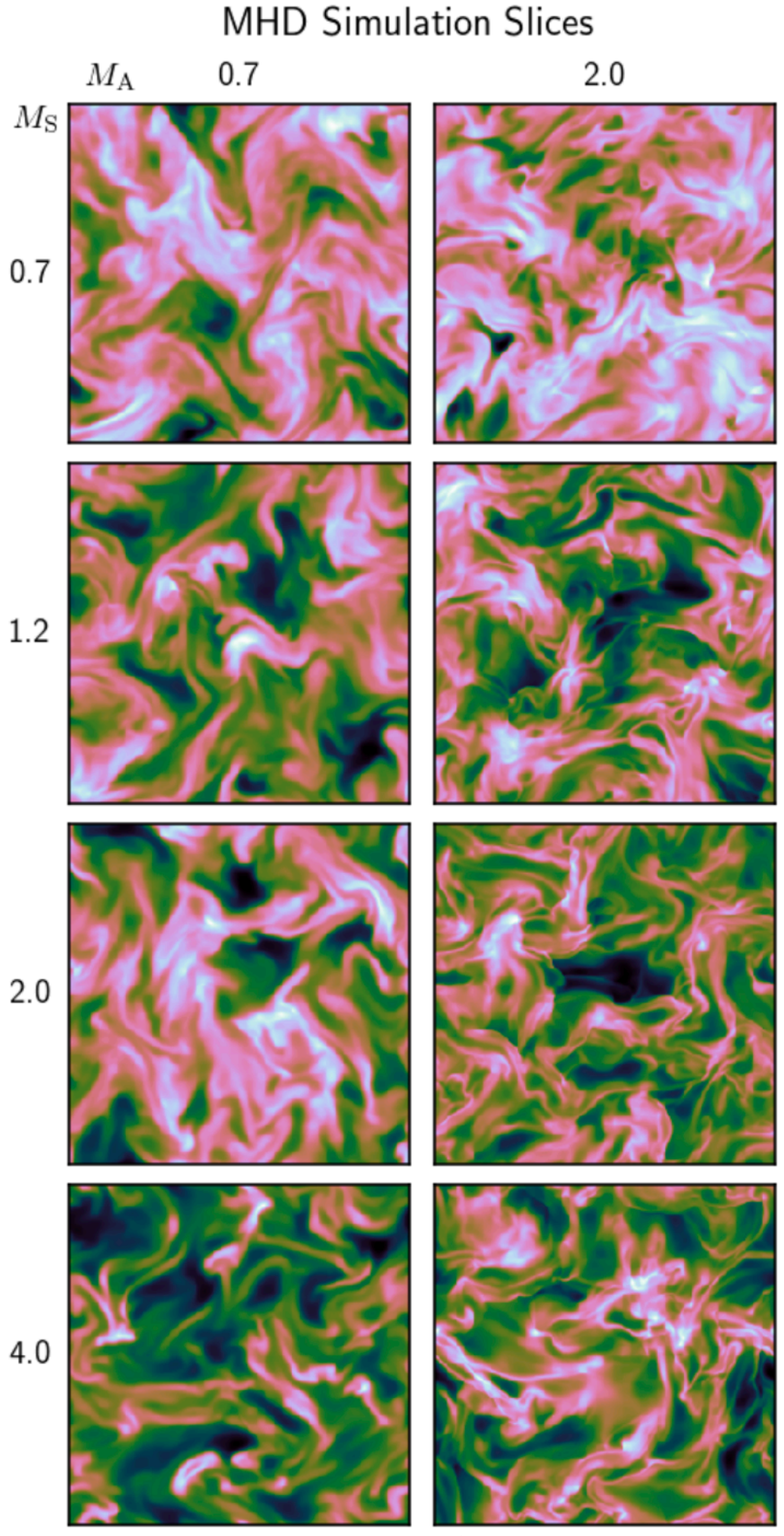
256x256 slices of MHD sims with varying sonic and Alfvénic Mach numbers provided by *CATS*.

Train on 7 time steps, test on 2.

Can WST-LDA tell these apart?

(We scale them to zero mean and unit variance before inputting)

Sims from CATS, Burkhart+ 2010.11227,
mhdturbulence.com



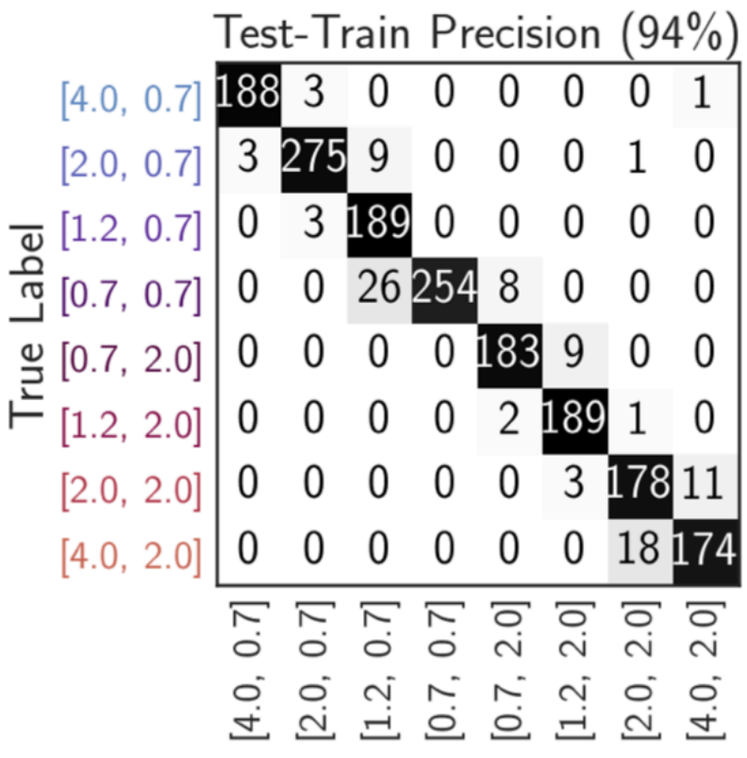
Saydjari+ 2010.11963

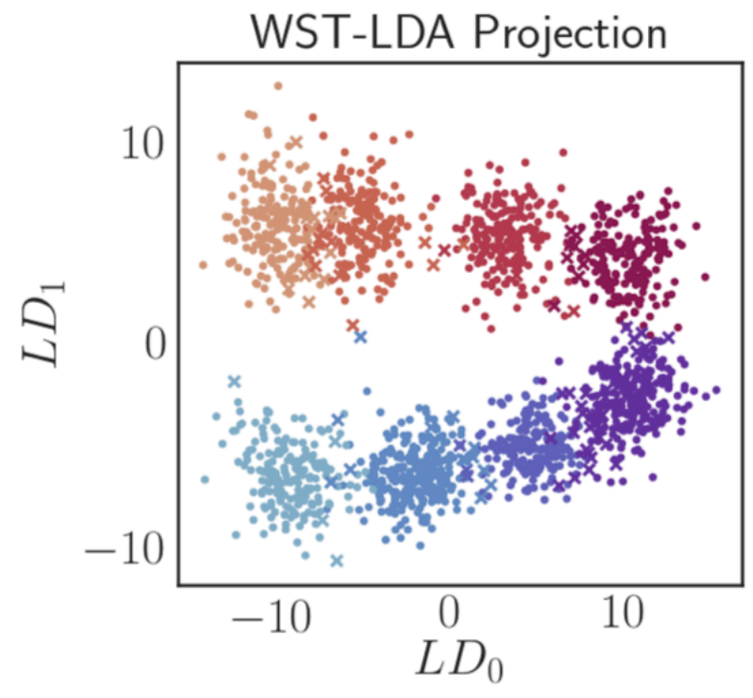
WST-LDA classifies MHD sims

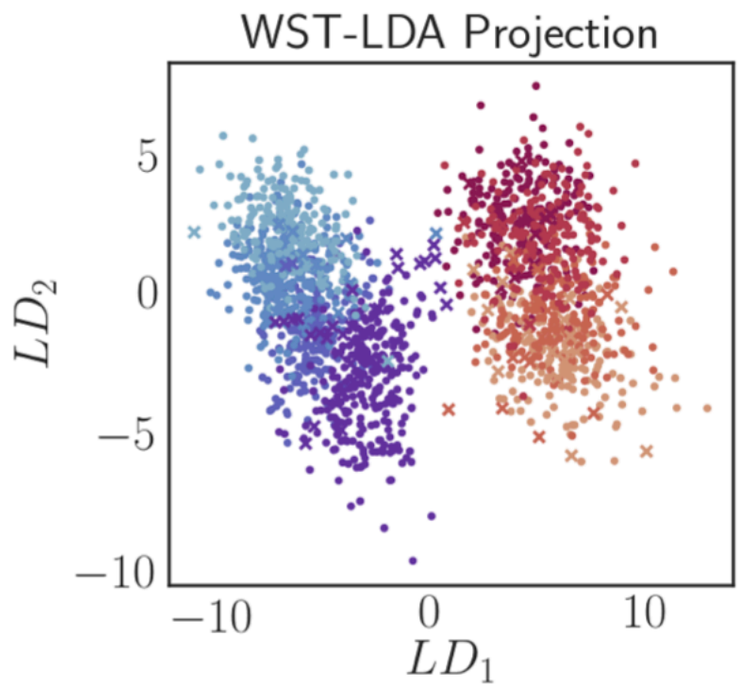
WST outputs 1827-D vectors.

The 8 classes make a helix in 3-D LDA space.

Test-train precision is 94%



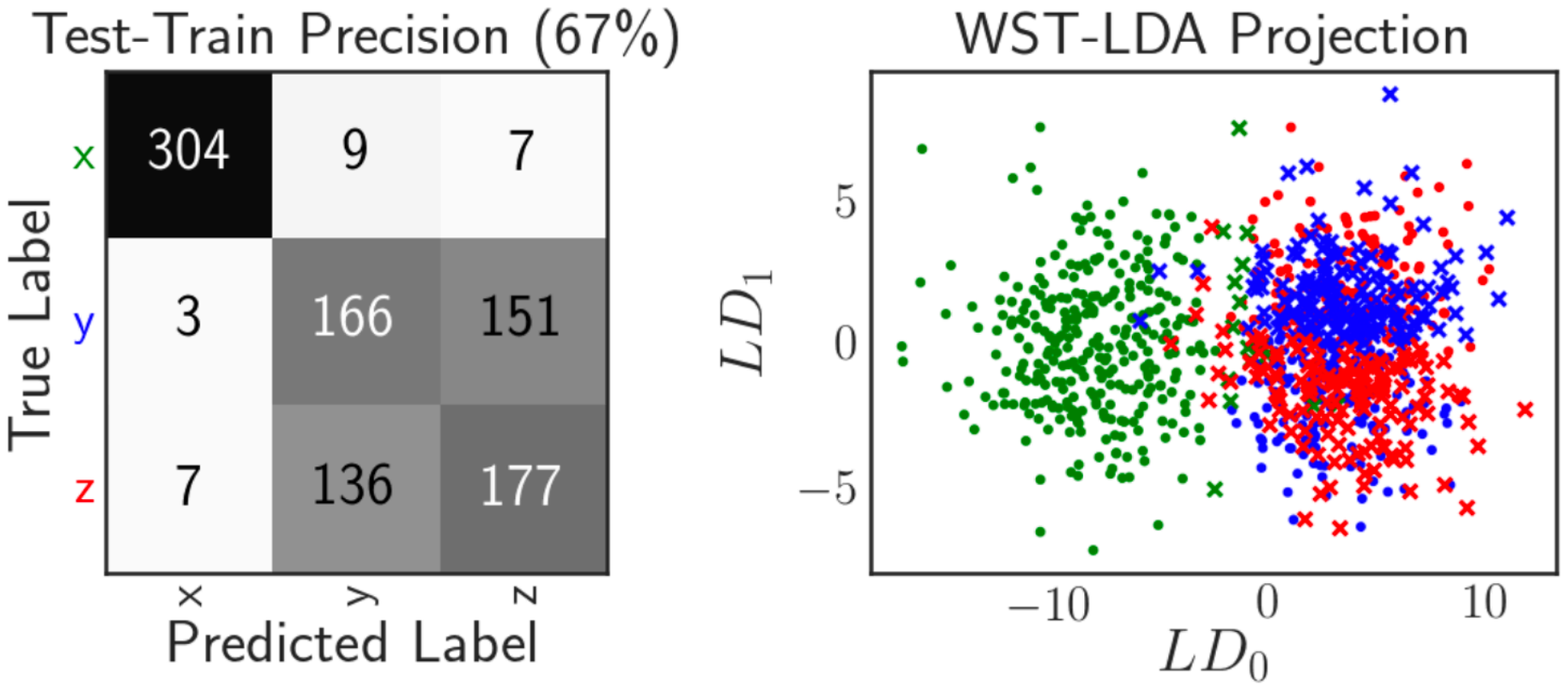




Saydjari+ 2010.11963

WST-LDA is sensitive to B-field direction

Magnetic field is in x direction.



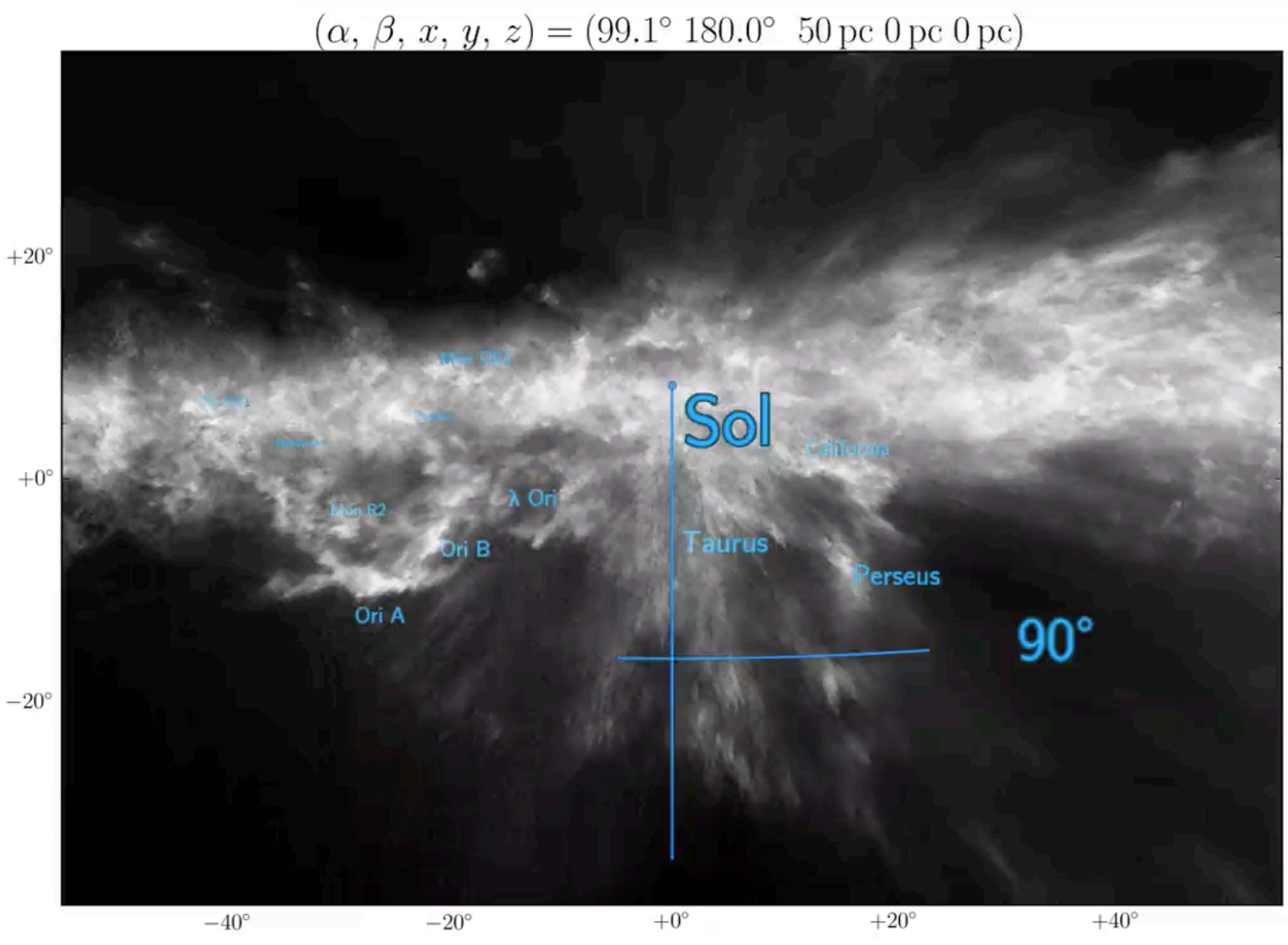
Saydjari+ 2010.11963

With WST-LDA in the toolbox

We might be able to

- read parameters off the ISM maps
- constrain 3-D dust inferences
- …?

3-D dust: We can derive the 3-D distribution of dust in the Milky Way using the colors and brightness of ~ 10^9 stars.



Gregory Green (see argonaut.skymaps.info)

An expressive prior

In order to make a 3-D dust map, it is useful to express our prior on the structure of dust.

A Gaussian Process prior is sometimes assumed.
This may not be good enough.

WST allows a more expressive prior that take account of the non-Gaussianity of the dust.

This should lead to better 3-D dust maps in the future.
(Stay tuned!)

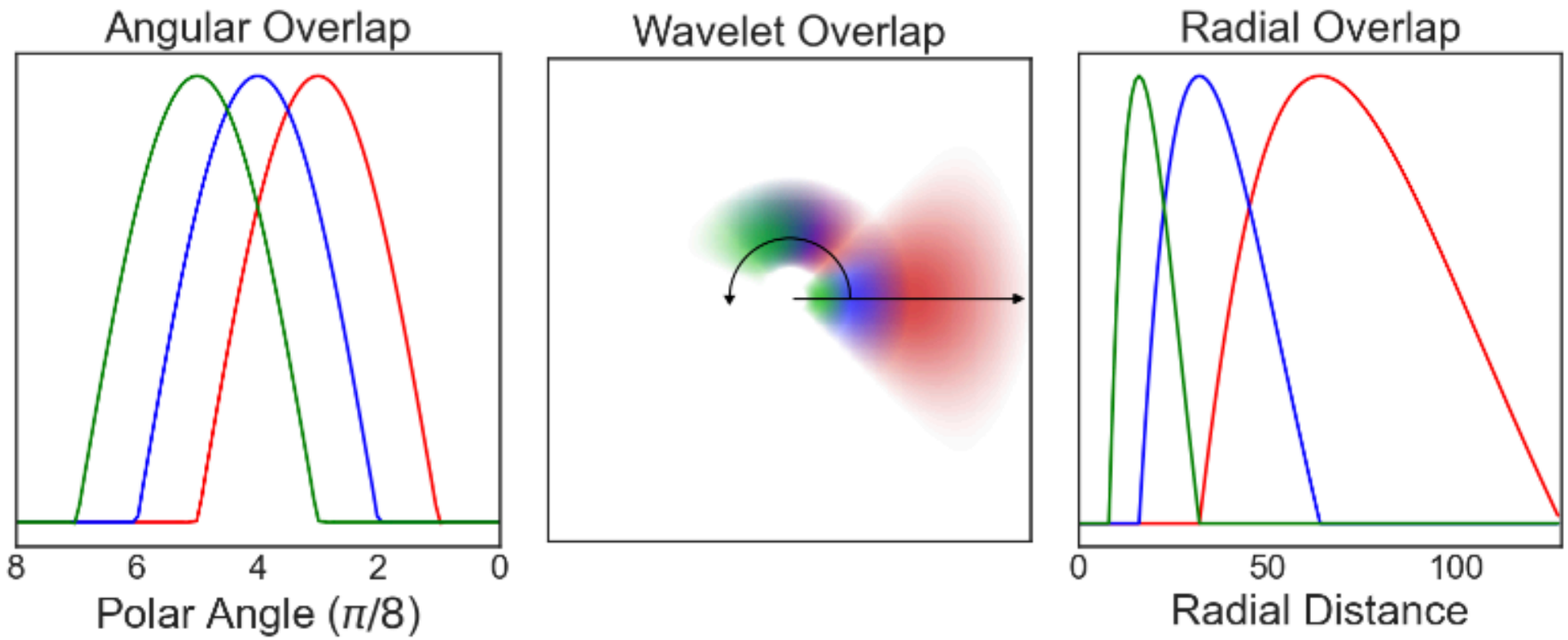
Current work

A new filter bank
- compact in Fourier domain
- exactly conserves “energy” energy=sum (values squared)
- exactly translation invariant
- approximately rotation invariant (~1E-3 in energy)
- implementation is fast.

Testing rotational invariance on MNIST

Triglets

A new filter bank: *triglets*
- cosine window in angle and log(r)

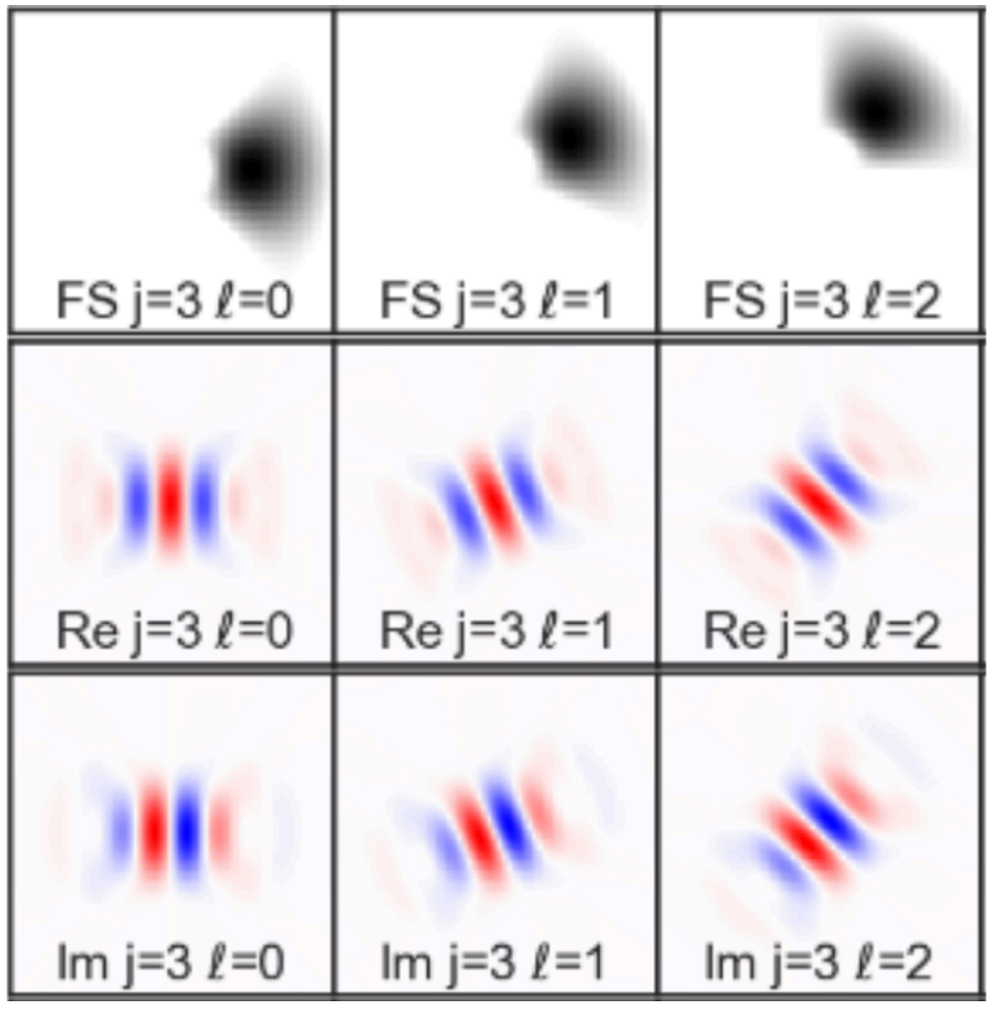


Saydjari & Finkbeiner (2021)

Triglets

A new filter bank: *triglets*
- cosine window in angle and log(r)
- compact support
- (go to zero at origin)
- sum of squares = 1.

As with all wavelets, one must be careful at the smallest and largest scales.



Saydjari & Finkbeiner (2021)

Triglets

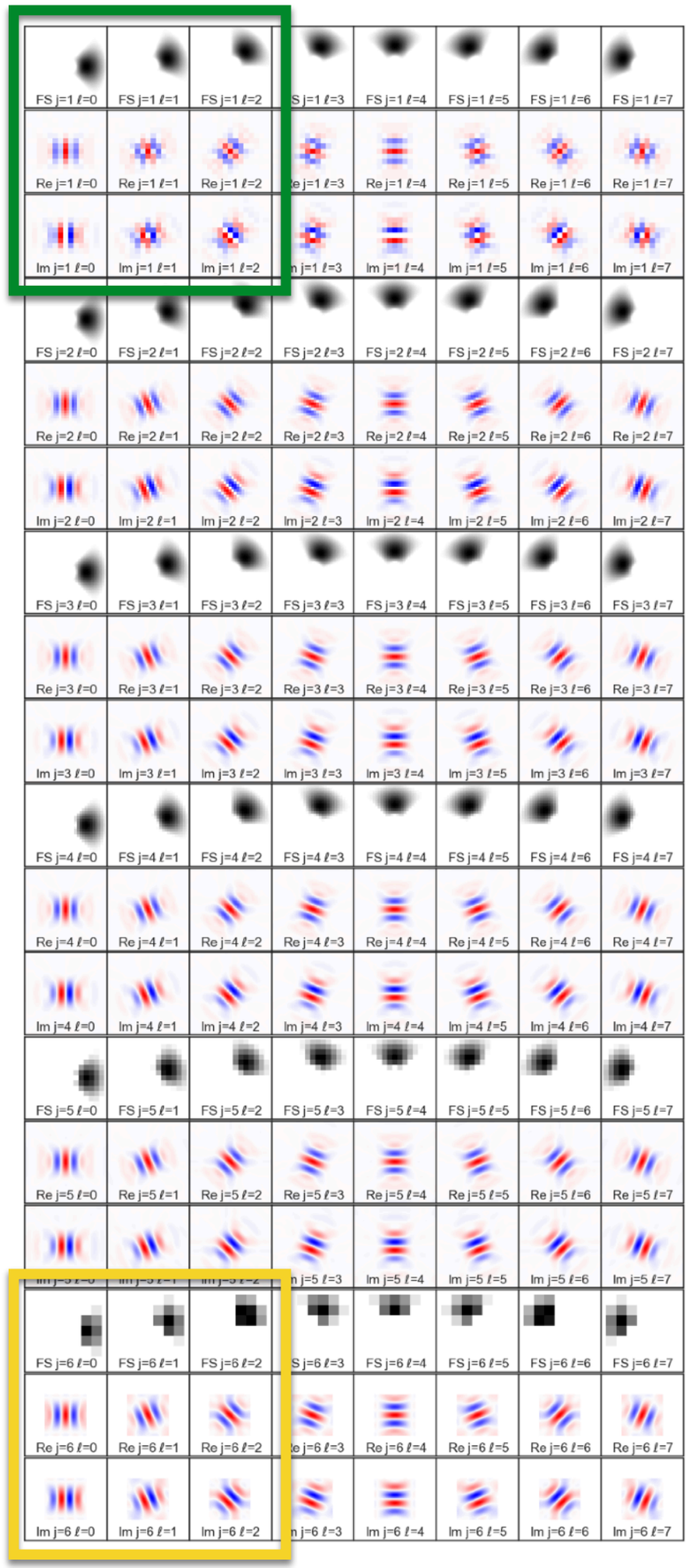
A new filter bank: *triglets*

- cosine window in angle and log(r)
- compact support
- (go to zero at origin)
- sum of squares = 1.

- (Non-directional wavelets at smallest and largest scales not shown)

As with all wavelets, one must be careful at the smallest and largest scales.

Saydjari & Finkbeiner (2021)

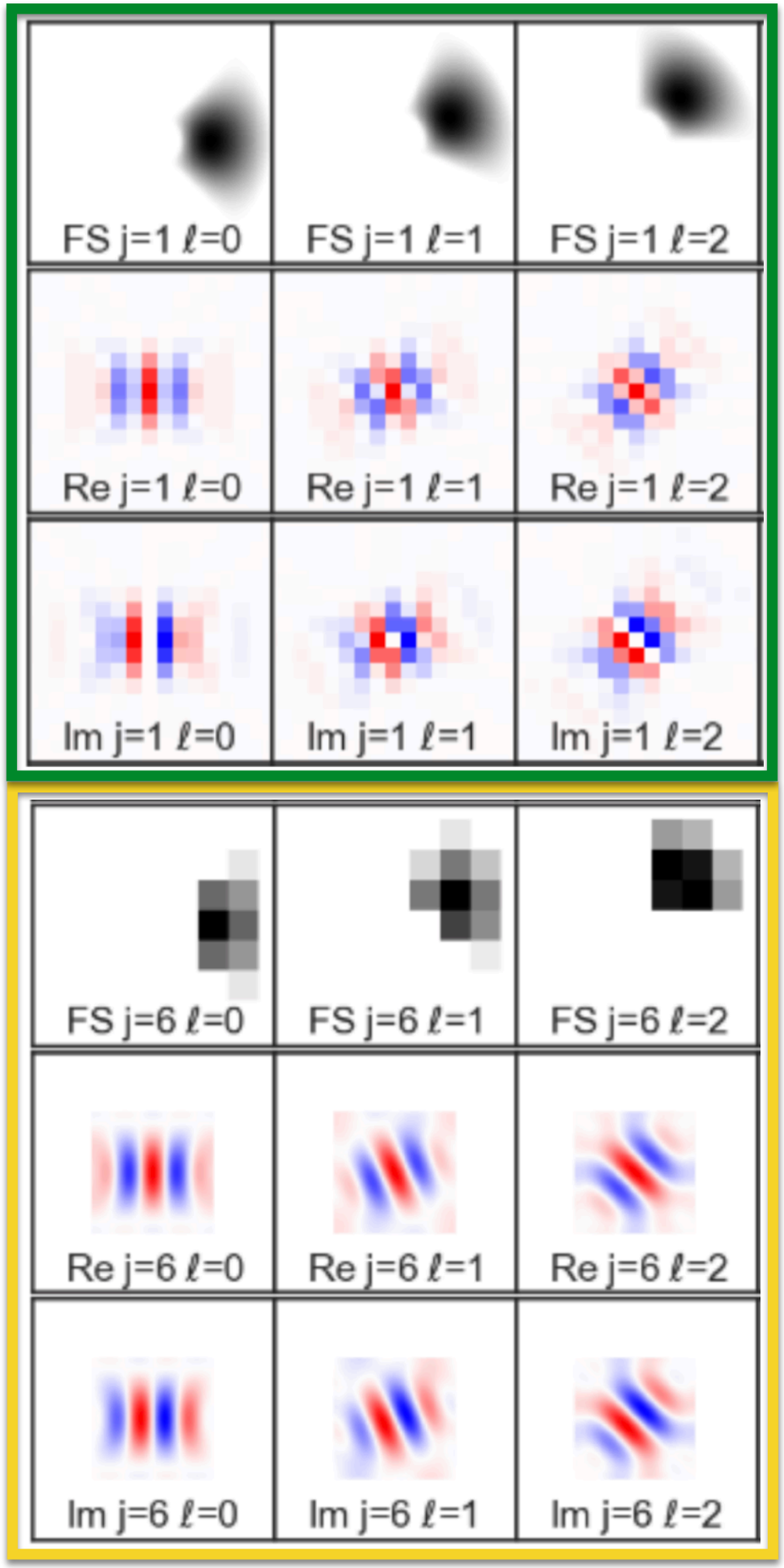


Adequate sampling -> invariance

Small wavelets (j=1) are coarsely sampled in the real domain.

Big wavelets (j=6) are coarsely sampled in the Fourier domain.

Either compromises rotation invariance.



Saydjari & Finkbeiner (2021)

"ISO" coefficients

We can take linear combinations of coefficients that are (nearly) rotation invariant.

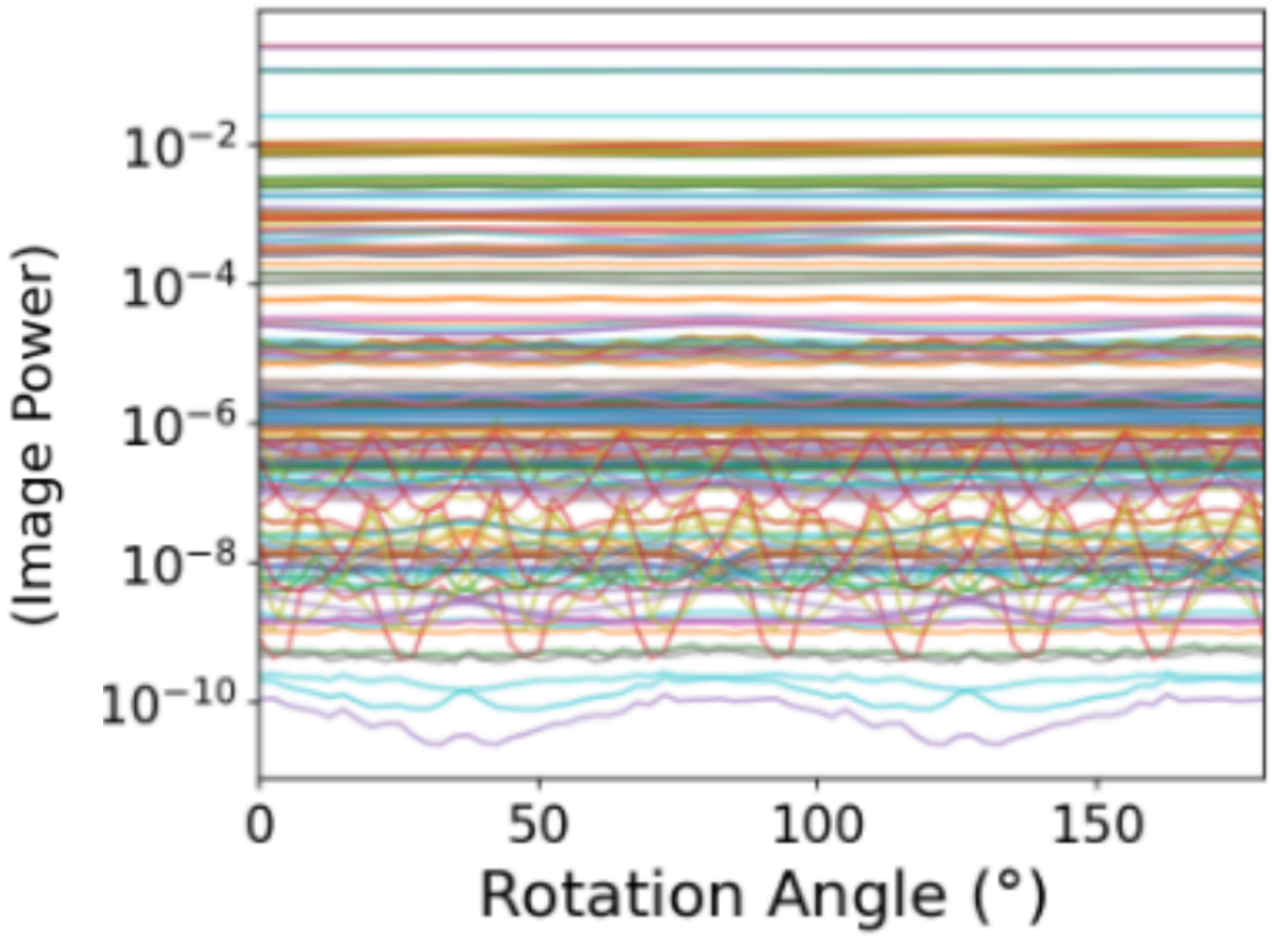
$$S_1(j_1, l_1), \quad S_2(j_1, j_2, l_1, l_2)$$

Become

$$S_1(j_1), S_2(j_1, j_2, \Delta l)$$

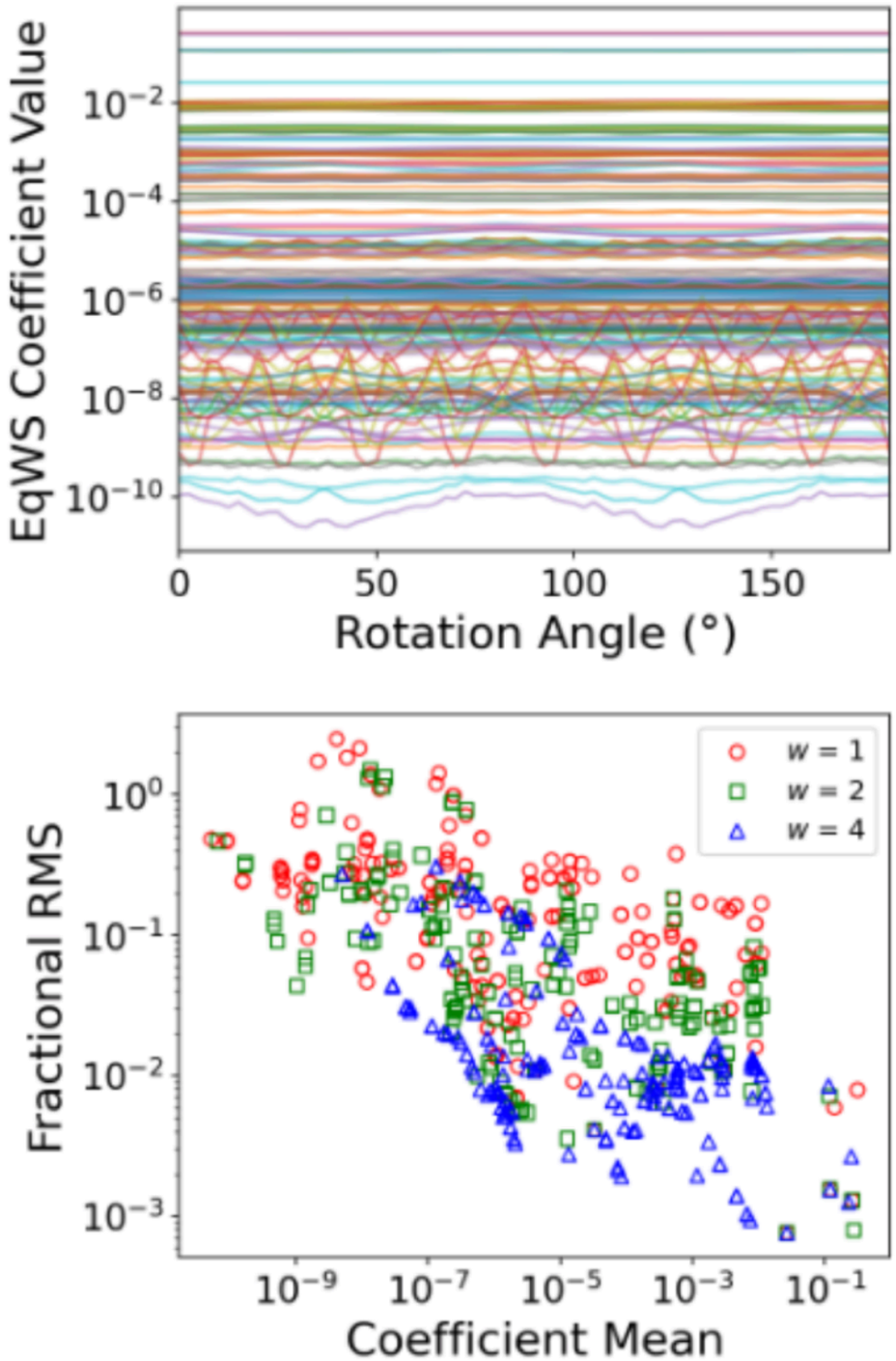
Spinning rod test

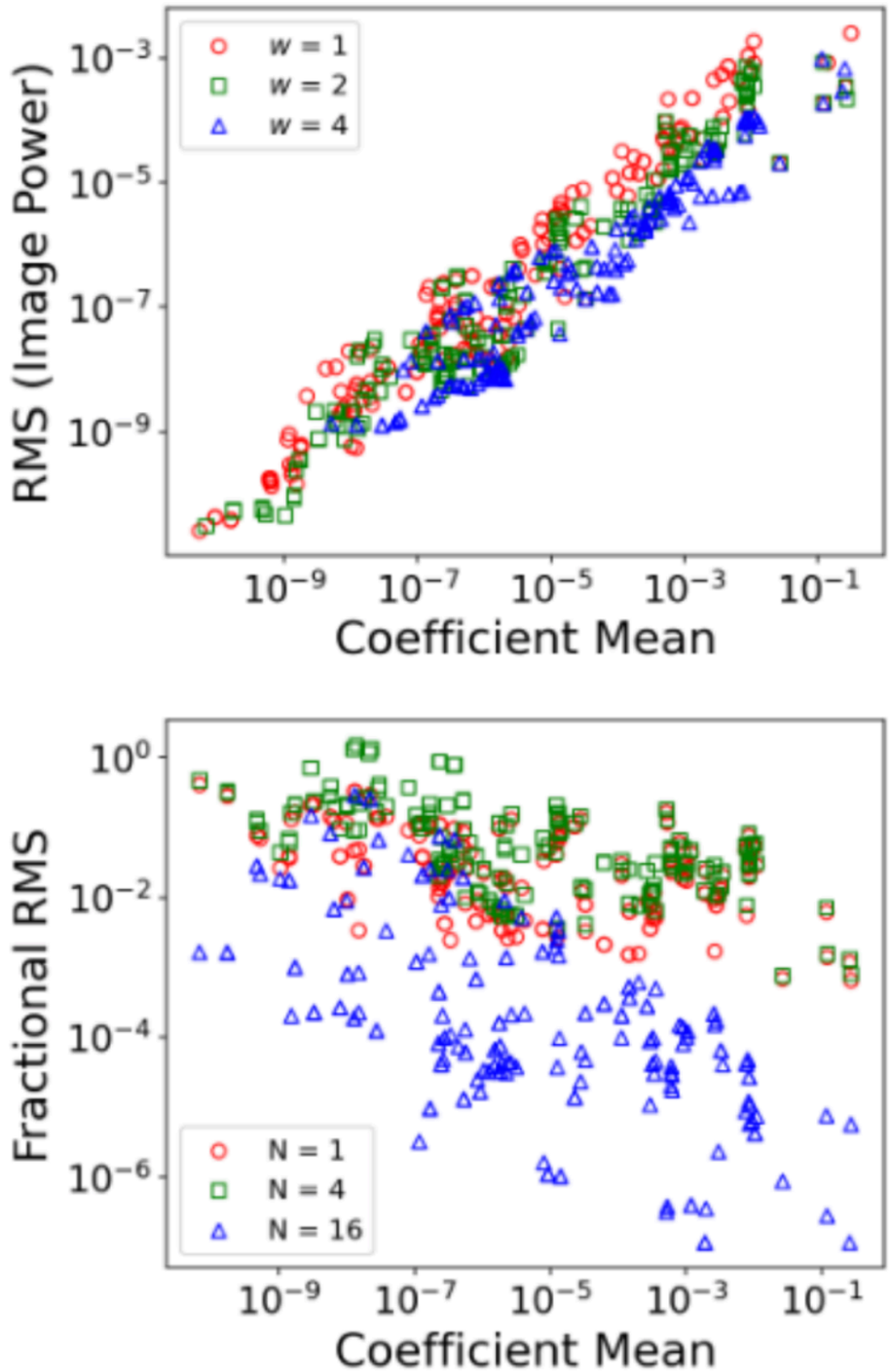
Coefficients with most of the image power are nearly invariant. Coefficients with < ppm power are unstable.



Saydjari & Finkbeiner (2021)

Spinning rod test





Saydjari & Finkbeiner (2021)

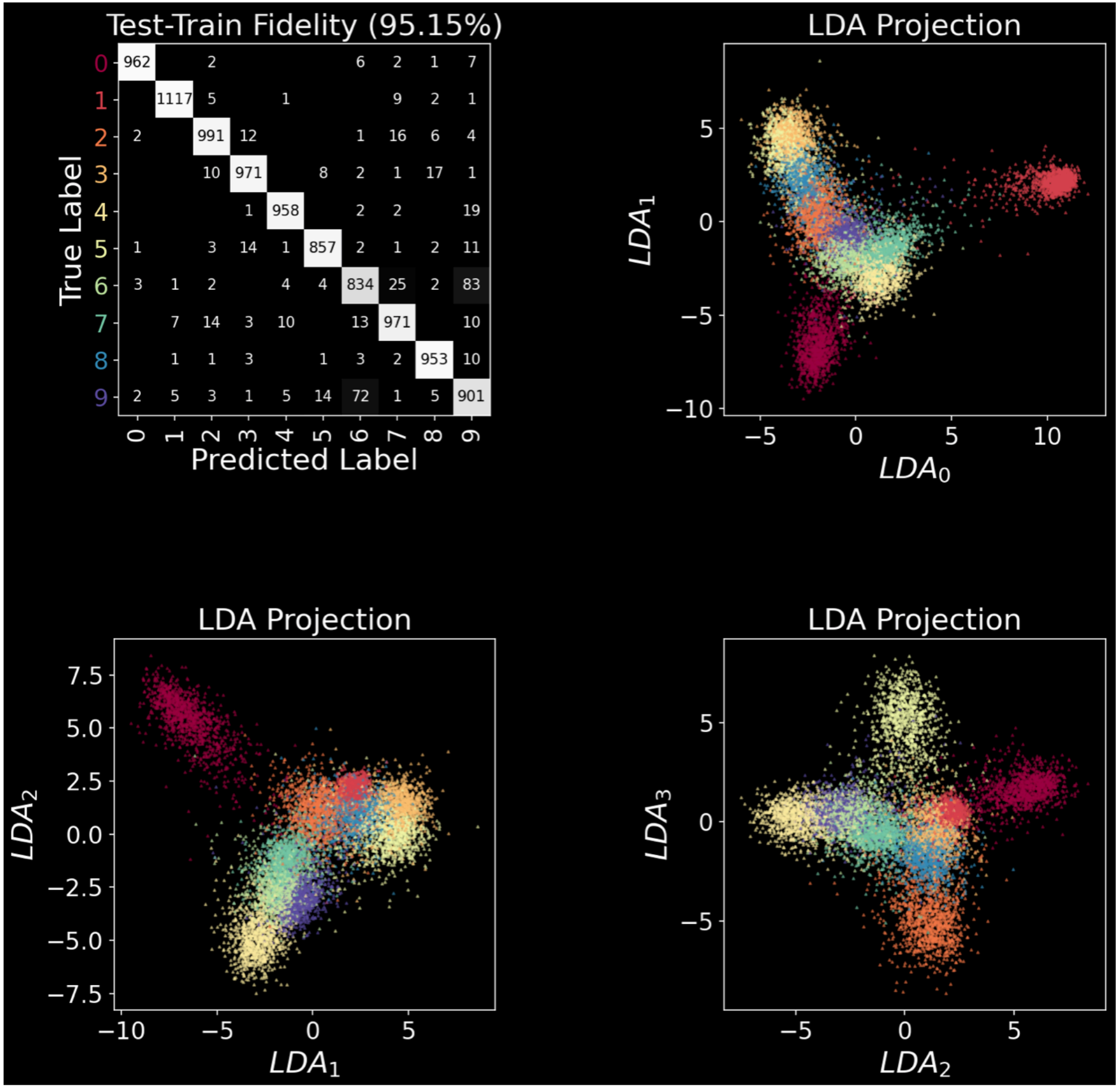
Toy problem: let's apply this to rotated MNIST

http://yann.lecun.com/exdb/mnist/

MNIST handwritten digits (Oriented)

$S_1(j_1, l_1)$,

$S_2(j_1, j_2, l_1, l_2)$



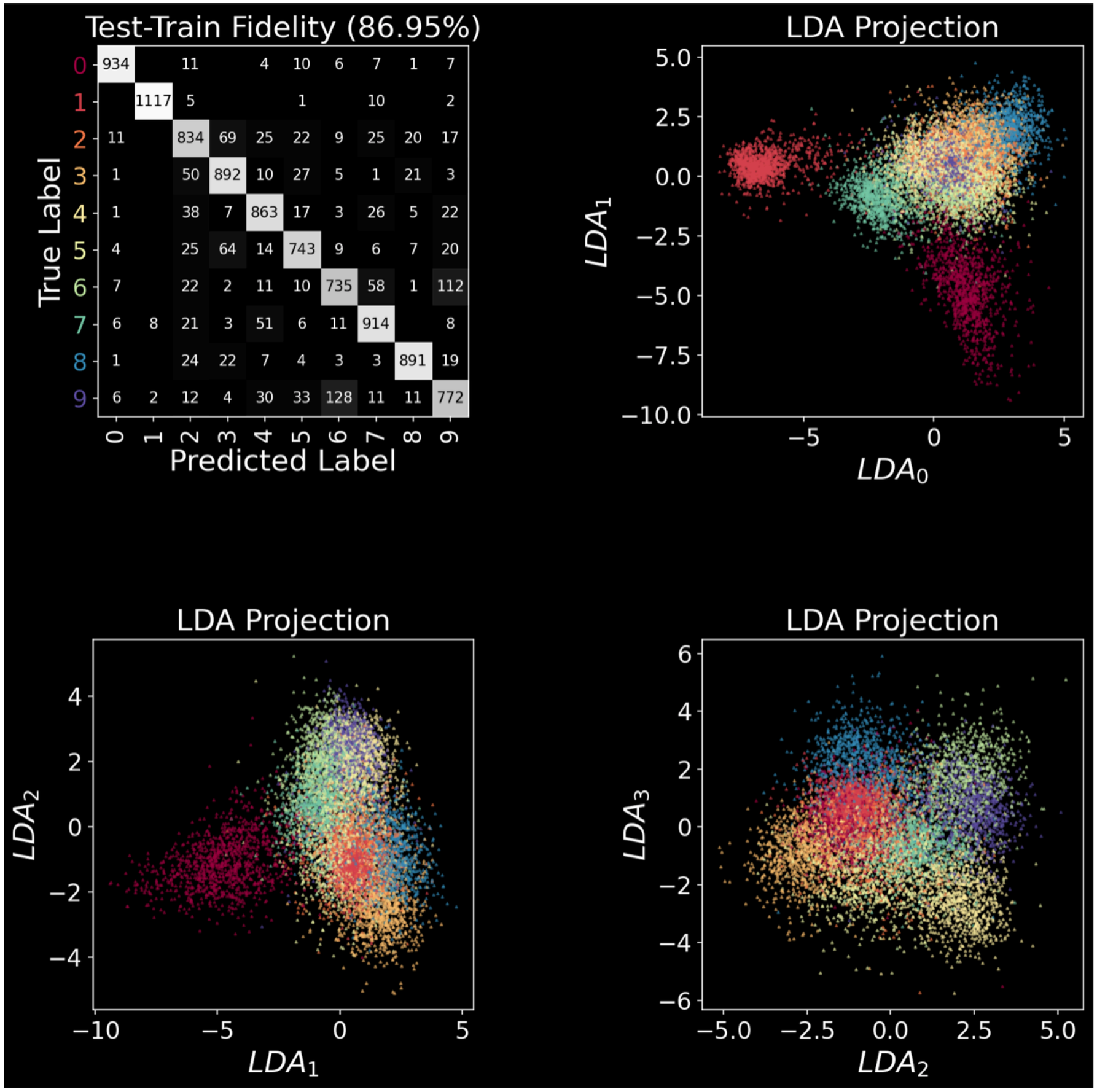
(Preliminary results)

Saydjari & Finkbeiner (2021)

MNIST handwritten digits (Random orientation)

ISO

$S_1(j_1), S_2(j_1, j_2, \Delta l)$



(Preliminary results)

Saydjari & Finkbeiner (2021)

Not as good as a spherical CNN, But *far* simpler!

ACCURACY OF EQWS+LDA ON BENCHMARK DATASETS

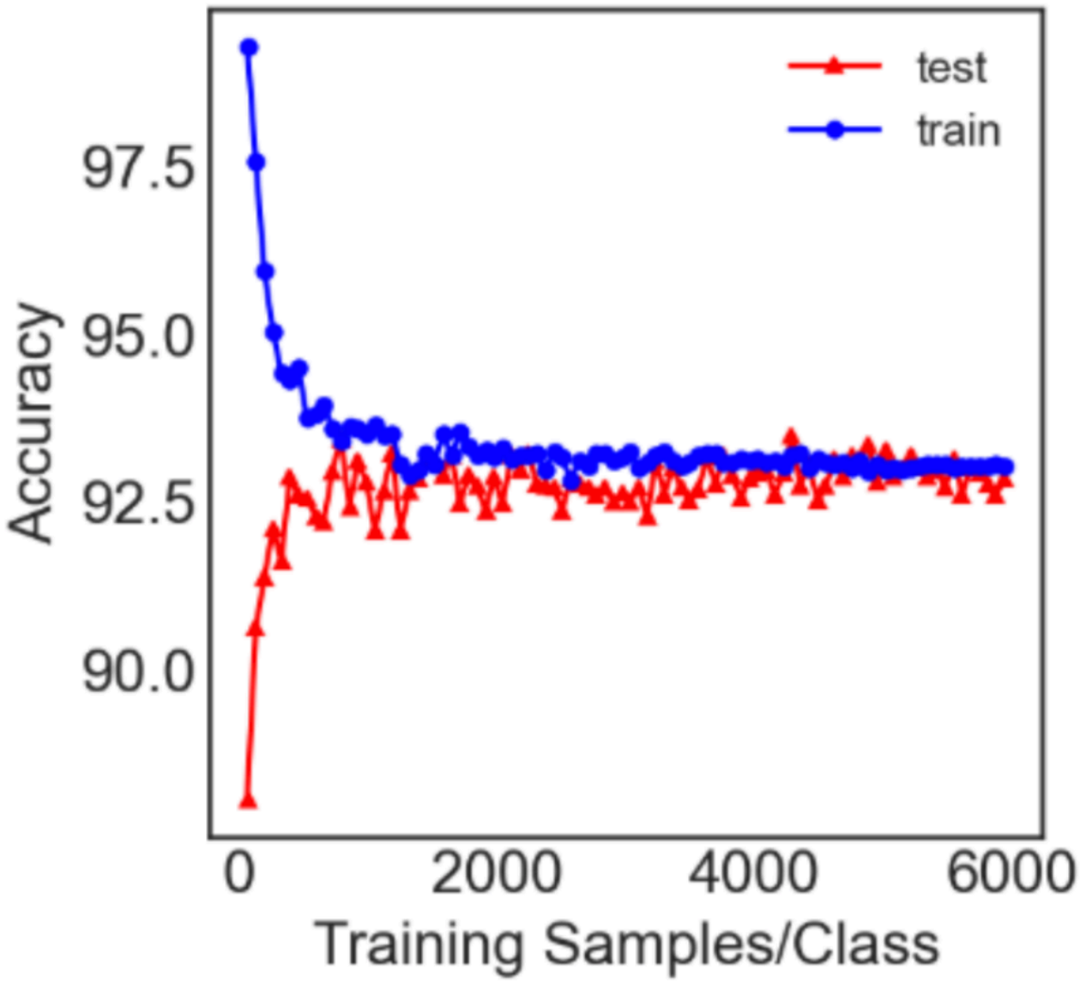
	NR/NR	R/R	NR/R
EqWS+LDA REG	96.23	92.3(1)	50.2(5)
EqWS+LDA ISO	93.5	92.10(5)	87.7(2)
EqWS+LDA REG (3 train angle)	94.15	92.3(1)	88.4(2)
EqWS+LDA ISO (3 train angle)	**91.95**	**92.10(5)**	**92.12(8)**
Cohen et al. [53]	95.59	94.62	93.40
Kondor et al. [54]	96.4	96.6	96.0
Esteves et al. [46]	99.37(5)	99.37(1)	99.08(12)
Planar CNN [46]	99.07	81.07(63)	17.23(71)

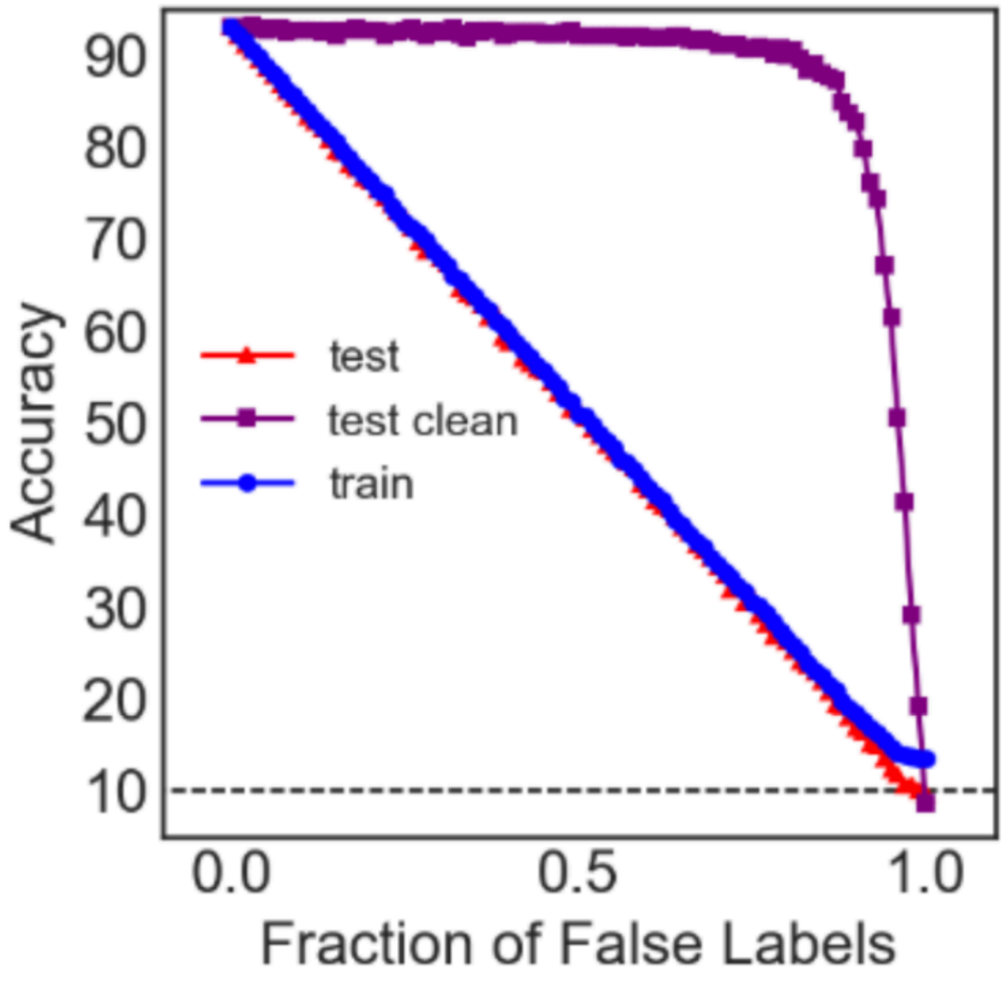
[46] C. Esteves, A. Makadia, and K. Daniilidis, "Spin-weighted spherical cnns," *arXiv preprint arXiv:2006.10731*, 2020.

[53] T. S. Cohen, M. Geiger, J. Köhler, and M. Welling, "Spherical cnns," *arXiv preprint arXiv:1801.10130*, 2018.

[54] R. Kondor, Z. Lin, and S. Trivedi, "Clebsch-gordan nets: a fully fourier space spherical convolutional neural network," *arXiv preprint arXiv:1806.09231*, 2018.

A simpler approach may require less training data.
Or lower quality training data.





Saydjari & Finkbeiner (2021)

Summary

Patterns of non-Gaussianity generally appear in physical systems of interest, yet the Gaussian Process assumption is usually made.

CNNs are highly sensitive to such patterns, but may require more training data than we have.

The wavelet scattering transform (WST) or our equivariant version (EqWS) provide sensitivity to non-Gaussianity, ~ invariance under translation and rotation, and *may* exceed CNN performance in the data-starved limit.