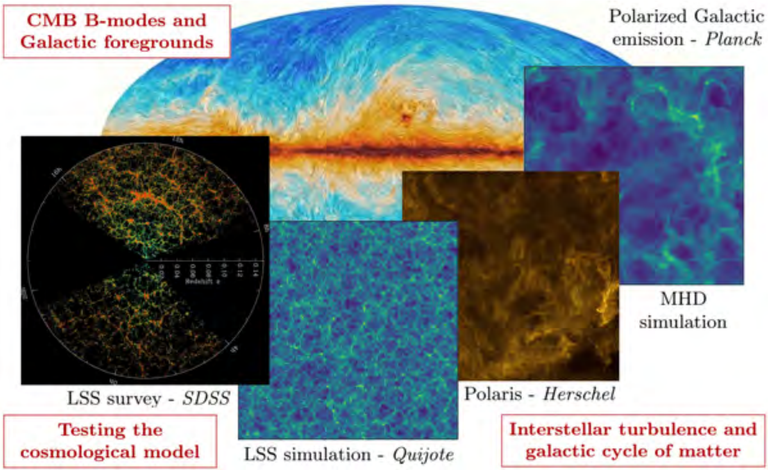
Generative model and component separation in limited data regime with Scattering Transform

Erwan Allys - ENS, Paris
(Physics laboratory and Center for data science)

"Artificial Intelligence for HPC@Exscale" Workshop,
Paris, October 3rd 2024

Outline

1. **Introduction**
2. **Scattering Transforms and generative models**
3. **Application to component separation**



→ Complex spatial/spectral processes ubiquitous in astrophysics

→ Sometimes no model and limited data regime...

Limited data regime in astrophysics

- **A limited amount of intricate observations**
  - A unique static multi-frequency sky
  - Mixture of non-stationary components
    $\rightarrow$ isolated processes can be very rare
    $\rightarrow$ Depends on angular scale/frequency

Limited data regime in astrophysics

- **A limited amount of intricate observations**
  - A unique static multi-frequency sky
  - Mixture of non-stationary components
    $\rightarrow$ isolated processes can be very rare
    $\rightarrow$ Depends on angular scale/frequency

- **Lack of models/training grounds**
  - Often no complete physical/numerical models
  - Simulations are very expensive
    $\rightarrow$ no or very limited training dataset

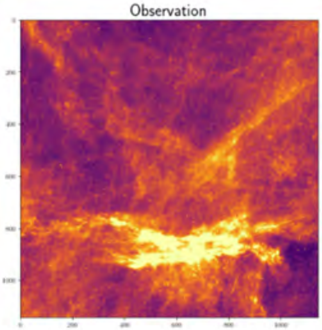
Limited data regime in astrophysics

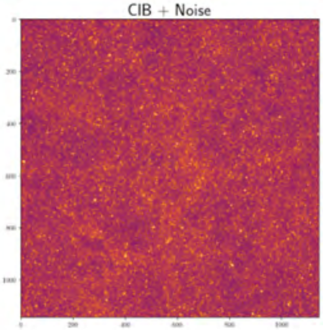
- **A limited amount of intricate observations**
  - A unique static multi-frequency sky
  - Mixture of non-stationary components
    → isolated processes can be very rare
    → Depends on angular scale/frequency

- **Lack of models/training grounds**
  - Often no complete physical/numerical models
  - Simulations are very expensive
    → no or very limited training dataset

→ Work mainly from obs. data and physical knowledge?
→ Rely on recent advances in data science?

Example I: CIB/Galactic dust emission

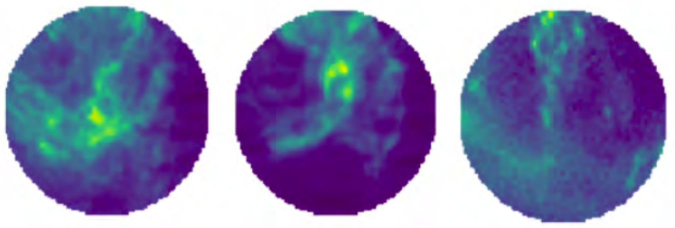




- **Galactic dust emission and Cosmic Infrared Background (CIB)**
  - Thermal dust emission in the interstellar medium
  - Cosmic background dominates a smaller scales
  - CIB isolated observation, no model for Galactic emission

→ Characterize Galactic dust emission on small scales?

Example II: WNM/CNM in HI observations



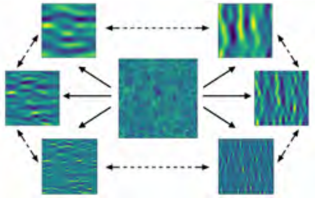
- **WNM/CNM component separation**
  - Warm (WNM) and Cold (CNM) Neutral Media
  - Two phases with different spectral/spatial properties
  - A few 1000s of unlabeled mixtures in HI (21cm) data

→ Learn phases structures directly from the data?

Outline

Scattering transform (ST) statistics

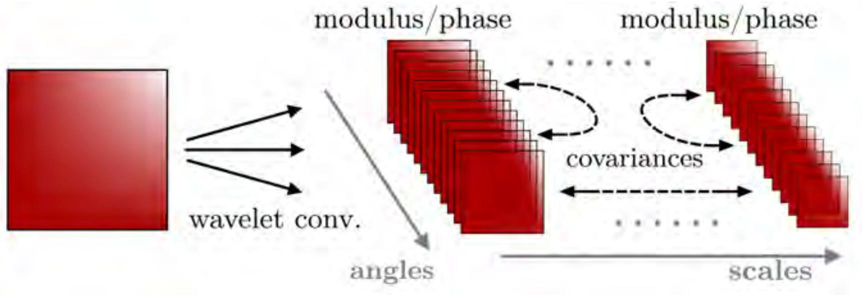
- **Scattering transform statistics** (Mallat+, 2010+)
  - Initially developed in data science
  - Inspired from neural networks
    → efficient characterization and reduced variance
  - Do not need any training stage
    → explicit mathematical form and interpretability



→ Wavelet filters separating the different scales
→ Coupling between scales with non-linearities

Scattering Transform (ST) statistics

- Computation of ST statistics *(EA+, 20)*



→ Shallow network with known filters/non-linearities
→ 1 coeff / pair/triplet of scales / type of interaction

Scattering Transform (ST) statistics

- **A family of statistics**
  - Different generations of statistics
    → Wavelet Scattering Transforms (WST) *(EA+, 19)*
    → Wavelet Phase Harmonics (WPH) *(EA+, 20)*
    → Scattering covariances/spectra *(Cheng+, 23)*
  - All share the same framework

Scattering Transform (ST) statistics

- **A family of statistics**
  - Different generations of statistics
    → Wavelet Scattering Transforms (WST) *(EA+, 19)*
    → Wavelet Phase Harmonics (WPH) *(EA+, 20)*
    → Scattering covariances/spectra *(Cheng+, 23)*
  - All share the same framework

- **Characterization and parameter inference**
  - Interstellar medium *(EA+ 19, Regaldo+20, Saydjari+, 20, Lei+, 22)*
  - Weak lensing *(Cheng+, 20, 21)*
  - Large scale structures *(EA+, 20, Eickenberg+, 22, Valogiannis+, 22a, 22b)*
  - 21cm epoch of reionization *(Greig+, 22, Hothi+, 23)*
  - ...

→ Very informative (sometimes on par with CNN!)
→ Wide range of applicability (generic, training-less)

Generative models from Scattering transforms

- **Generative model from ST statistics** *(Bruna, Mallat, 19)*
  - From the ST statistics $\Phi(s)$ of a map s
  - Maximum entropy microcanonical model
  - New approximate samples of a process from examples
  - Can quantitatively reproduce non-gaussian properties

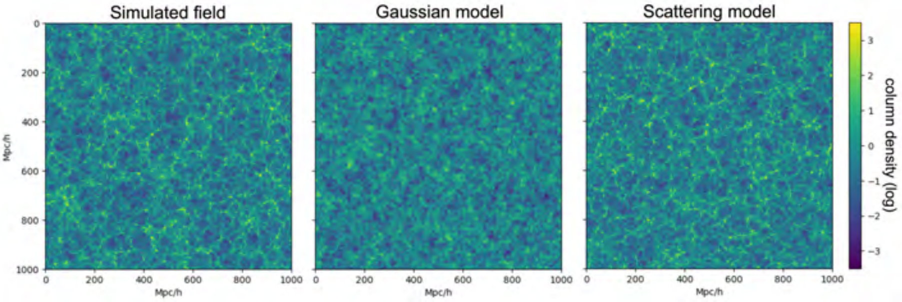
Generative models from Scattering transforms

- **Generative model from ST statistics** *(Bruna, Mallat, 19)*
  - From the ST statistics $\Phi(s)$ of a map s
  - Maximum entropy microcanonical model
  - New approximate samples of a process from examples
  - Can quantitatively reproduce non-gaussian properties

- **Practical implementation**
  - Constraints $\Phi(s)$ from a (set of) data s
  - Sampled with a gradient-descent algorithm
    $\rightarrow$ from a white noise realization
    $\rightarrow$ optimizing $\tilde{s}$ such that $\Phi(\tilde{s}) \simeq \Phi(s)$

Generative models from Scattering transforms

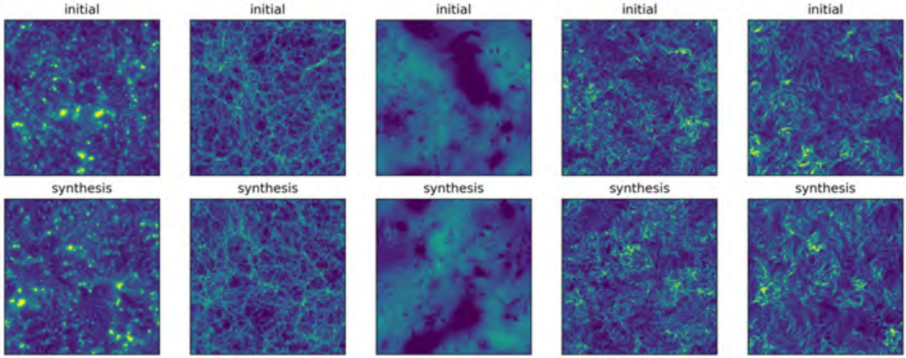
- **Quantitative validation of syntheses** *(EA+, 20)*
  - Large scale structures density field, Wavelets Phase Harmonics



→ **Usual (NG) statistics very well reproduced (up to 1-10 %)**

Generative models from Scattering transforms

- **Syntheses from a single image** *(Cheng+, 24)*
  - Scattering spectra + physical dimensionality reduction



→ **Realistic NG models from a few hundreds coefficients!**
→ **Well adapted for a large number of physical fields**

Generative models from Scattering transforms

- **ST generative models**
  - Low-dimensional: few 100s to low 1000s
  - For regular "physical" fields
  - Extended to various fields

    → Higher dimensions (3D+)

    → On the sphere ***(Mousset+, 24)***

    → Multi-channel ***(Régaldo+, 22)***

Generative models from Scattering transforms

- **ST generative models**
  - Low-dimensional: few 100s to low 1000s
  - For regular "physical" fields
  - Extended to various fields
    - → Higher dimensions (3D+)
    - → On the sphere *(Mousset+, 24)*
    - → Multi-channel *(Régaldo+, 22)*

- **Needs related to NumPEx**
  - Versatile and scalable ST computations:
    - → Adaptable Wavelet transforms
    - → Fourier Transform (for modulus)
  - Sampling algorithms (other than gradient descent ?)
    - → Efficiently tracing the gradients
    - → Incl. mean-field approach? *(Häggbom+, 24)*

Outline

1. Introduction
2. Scattering Transforms and generative models
3. **Application to component separation**

Statistical component separation

- **Framework of the problem**
  - We observe a mixture $d = s + c$
    $\rightarrow$ d data, s signal of interest, c contamination
  - Use knowledge of c to recover s
  - No model for s...

Statistical component separation

- **Framework of the problem**
  - We observe a mixture $d = s + c$
    $\rightarrow$ d data, s signal of interest, c contamination
  - Use knowledge of c to recover s
  - No model for s...

- **Using ST generative models**
  - Assume S is described by p_μ ST models
    $\rightarrow$ Parametrized by μ such that $E_{s \sim p_\mu}[\phi(x)] = \mu$
  - Estimate (part of) the $p(\mu|d)$ posterior

$\rightarrow$ Statistical component separation
$\rightarrow$ Allows to get a posterior model of s

Statistical component separation

- **Bayesian approach of the problem**
  - Parameter space: μ
  - Data space: $\phi(d)$
  - Forward model: $f(s) = d$ converted in ST space

$$\mu \xrightarrow{p_\mu(s)} s \xrightarrow{f(s)=s+c} d \xrightarrow{ST} \phi(d)$$

Statistical component separation

- **Bayesian approach of the problem**
  - Parameter space: μ
  - Data space: $\phi(d)$
  - Forward model: $f(s) = d$ converted in ST space

$$\mu \xrightarrow{p_\mu(s)} s \xrightarrow{f(s)=s+c} d \xrightarrow{ST} \phi(d)$$

- **Estimating the posterior**
  - Forward $f(\mu) = \phi(d)$ samplable in ST space
  - Broad $p(\mu)$ prior can be defined around $\phi(d)$

$\Rightarrow$ Well posed Bayesian problem

Statistical component separation

- **Bayesian approach of the problem**
  - Parameter space: μ
  - Data space: $\phi(d)$
  - Forward model: $f(s) = d$ converted in ST space

$$\mu \xrightarrow{p_\mu(s)} s \xrightarrow{f(s)=s+c} d \xrightarrow{ST} \phi(d)$$

- **Estimating the posterior**
  - Forward $f(\mu) = \phi(d)$ samplable in ST space
  - Broad $p(\mu)$ prior can be defined around $\phi(d)$

$\Rightarrow$ Well posed Bayesian problem

$\rightarrow$ Sample/data space dim $\sim$ 500 without structure
$\rightarrow$ Forward is still expensive (a few seconds on a GPU)

Solution with a gradient descent approach

- **Generative model from available sample**
  - Estimate $\phi(s)$ from sample s
  - Generate maps $\tilde{s}$ such that

$$\Phi(\tilde{s}) \simeq \Phi(s)$$

  - Sampled with gradient descent from white noise

Solution with a gradient descent approach

- **Generative model from available sample**
  - Estimate $\phi(s)$ from sample s
  - Generate maps $\tilde{s}$ such that

$$\Phi(\tilde{s}) \simeq \Phi(s)$$

  - Sampled with gradient descent from white noise

- **Indirect observation with know contamination**
  - $d = s_0 + c_0$, assume we have $p(c)$
  - Find map(s) $\tilde{s}$ such that

$$\langle \Phi(\tilde{s} + c_i) \rangle_i \simeq \Phi(d)$$

  - Gradient descent from d (for instance)

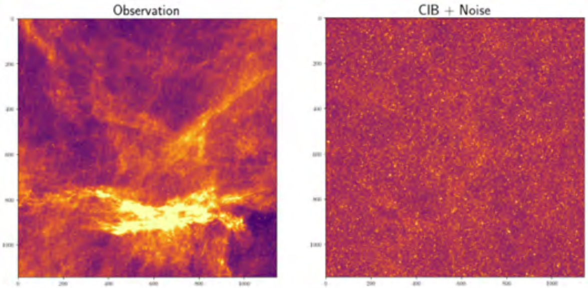
Solution with a gradient descent approach

- **Generative model from available sample**
  - Estimate $\phi(s)$ from sample s
  - Generate maps $\tilde{s}$ such that
$$\Phi(\tilde{s}) \simeq \Phi(s)$$
  - Sampled with gradient descent from white noise

- **Indirect observation with know contamination**
  - $d = s_0 + c_0$, assume we have $p(c)$
  - Find map(s) $\tilde{s}$ such that
$$\langle \Phi(\tilde{s} + c_i) \rangle_i \simeq \Phi(d)$$
  - Gradient descent from d (for instance)

→ Framework for component separation
→ Can include various other statistical constraints

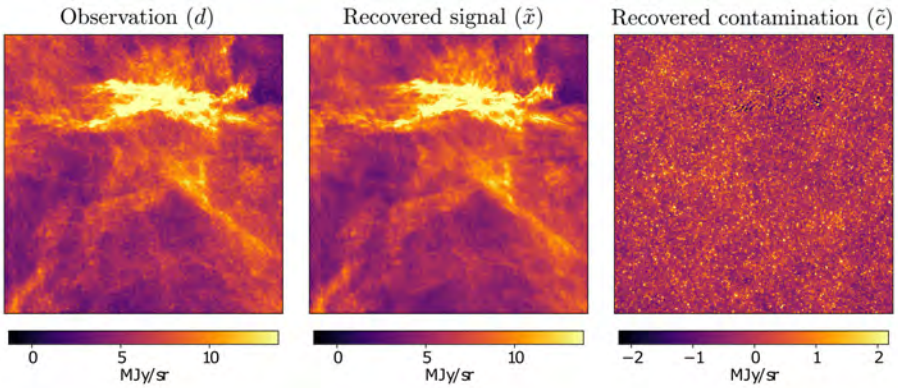
Solution with a gradient descent approach



- **Dust emission/Cosmic Infrared Background (Auclair+, 24)**
  - $d = s + c$, s thermal dust emission, c CIB
  - CIB model from separate observation (cosmological $\Rightarrow$ homogeneous)
  - Two constraints, with $\{c_i\}_i$ from ST model

$$\left\langle \Phi(\tilde{s} + c_i) \right\rangle_i \simeq \Phi(d), \qquad \Phi(\tilde{c}) = \Phi(c)$$

- Recovered components (Auclair+, 24)



→ Statistical component separation solely from obs. data!
→ Thermal dust is recovered at an unprecedented resolution

Solution with a gradient descent approach

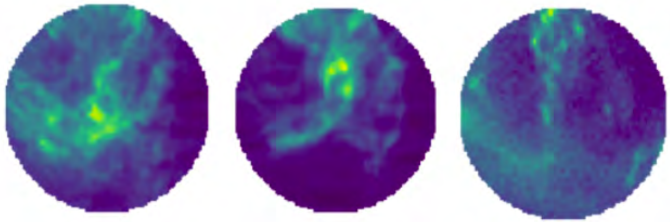
- **Statistical component separation with ST**
  - Efficient and versatile framework
    $\rightarrow$ **(Régaldo+, 21, Delouis+22, Siahkoohi+23)**
  - Allow joint handling of various constraints
    $\rightarrow$ 10s of losses in recent works
    $\rightarrow$ Incl. cross-statistics with ancillary data
  - Still need to get a proper mathematical framework

Solution with a gradient descent approach

- **Statistical component separation with ST**
  - Efficient and versatile framework
    → **(Régaldo+, 21, Delouis+22, Siahkoohi+23)**
  - Allow joint handling of various constraints
    → 10s of losses in recent works
    → Incl. cross-statistics with ancillary data
  - Still need to get a proper mathematical framework

- **Needs related to NumPEx**
  - High dimensional optimization
    → Incl. high number of constraints
    → Exploration of "good" solutions
  - High-dimensional Bayesian inversion
    → Sampling spaces of dim. a few 100s
    → Incl. adaptive learning

Unsupervised separation of HI data

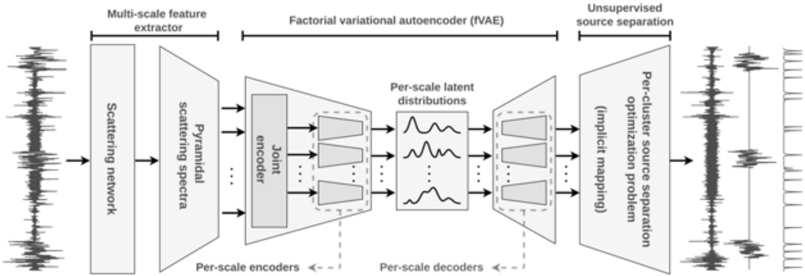


- **Recovering phase structures from the data**
  - A lot of HI observations of CNM/WNM/noise mixture
  - No specific knowledge on each mixture
    → No isolated component observed

→ Directly learn components ST models from the data?

Unsupervised separation of HI data

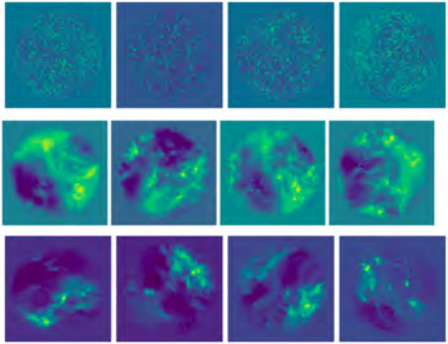
- **Variational Auto-Encoder (VAE) in ST space** (Siahkoohi+, 23a,b)



→ Unsupervised learning of components in ST space

Unsupervised separation of HI data

- **Application to HI data (Lei+, work in progress)**
  - Unsupervised identification of 3 components



→ **WNM/CNM/noise seem quite well modeled!**
→ **Interfacing ST models with other ML algorithms**

Conclusion

- **Scattering Transforms**
  → Non-Gaussian statistics inspired from neural network
  → Efficient low-dim modeling of physical processes
- **New tools for (astro-)physics and beyond**
  → ST models as basis for various inv. problems/comp. sep.
  → First proofs of concepts obtained in component separation
  → Ability to work with a very limited amount of data!
- **Help from NumPEx would be precious**
  → Not always easy to interface with standard algorithms
  → Scale to larger problems from proofs of concept
  → Need to develop HPC expertise as a whole

Thanks for your attention!

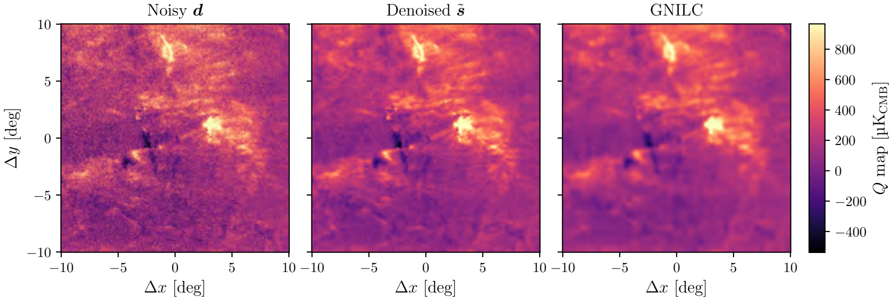
Separation of polarized dust and noise

- **Application to dust polarized emission and noise (Regaldo+, 21)**
  - $d = s + c$ *Planck* polarization data at 353GHz
  - s polarized dust emission, c inhomogeneous noise
  - 300 noise realizations c_i from Planck team
  - Optimization done from d to keep largest scales

$$\langle \Phi(\tilde{s} + c_i) \rangle_i \simeq \Phi(d)$$

Separation of polarized dust and noise

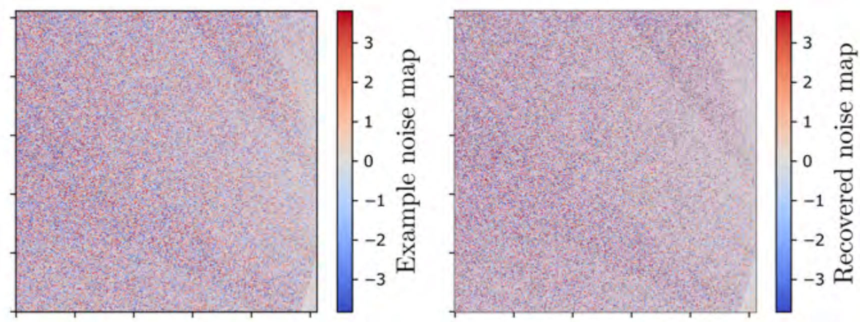
- **Application to Chameleon-Musca region (Régaldo+ 21)**



→ **Transition btw. deterministic and statistical**
→ **Conceptual validation of the method**

Separation of polarized dust and noise

- Recovered contamination (Régaldo+ 21)

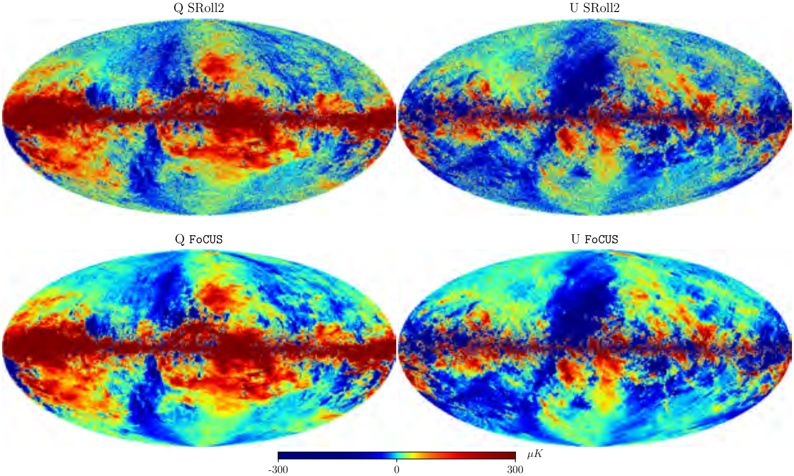


→ Statistical separation of components
→ Residual structures could also be constrained

Separation of polarized dust and noise

- **Refinements of this work on the whole sky (Delouis+, 22)**
  - Introduce additional constraints
    → 3 constraints including cross-statistics
  - Educated normalization of each constraint
    → constraints normalized by variance over $\{c_i\}$
  - Introduce local constraints for non-stationary
    → 4 selected regions for Galactic heterogeneity

- **Full sky results at 353GHz** (Delouis+, 22)



→ Deterministic up to SNR≃0.1, statistical up to SNR≃0.01
→ Efficient and versatile framework for statistical comp. separation