

# Imaging data from next-generation radio interferometric telescopes with compressive sensing

Jason McEwen

[www.jasonmcewen.org](http://www.jasonmcewen.org)

@jasonmcewen

*Mullard Space Science Laboratory (MSSL)  
University College London (UCL)*

*With thanks to my collaborators Laura Wolz, Filipe Abdalla, Rafael Carrillo & Yves Wiaux*

Applied and Computational Analysis Series,  
Department of Applied Mathematics and Theoretical Physics (DAMTP),  
University of Cambridge, October 2015

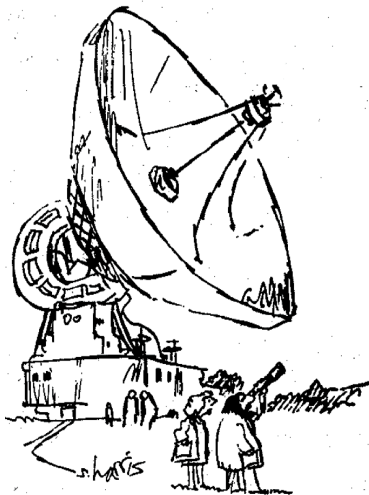








# Radio telescopes are big!



“Just checking.”



# Radio telescopes are big!



# Radio interferometric telescopes



# Next-generation of radio interferometry rapidly approaching

- **Square Kilometre Array (SKA)** construction scheduled to begin 2018.
- Many other pathfinder telescopes under construction, *e.g.* LOFAR, ASKAP, MeerKAT, MWA.
- Broad range of science goals.



**Figure:** Artist impression of SKA dishes. [Credit: SKA Organisation]



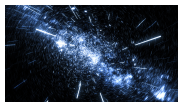


# Next-generation of radio interferometry rapidly approaching

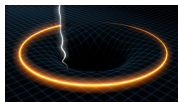
- **Square Kilometre Array (SKA)** construction scheduled to begin 2018.
- Many other pathfinder telescopes under construction, *e.g.* LOFAR, ASKAP, MeerKAT, MWA.
- Broad range of science goals.



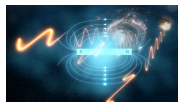
**Figure:** Artist impression of SKA dishes. [Credit: SKA Organisation]



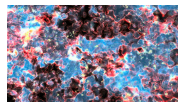
(a) Dark-energy



(b) GR



(c) Cosmic magnetism



(d) Epoch of reionization



(e) Exoplanets

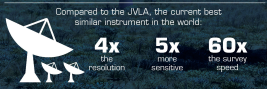
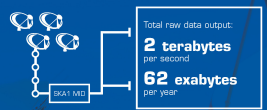
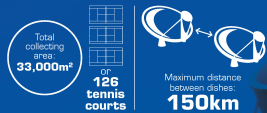
**Figure:** SKA science goals. [Credit: SKA Organisation]



# SKA sites

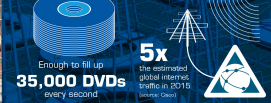
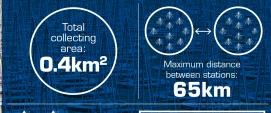
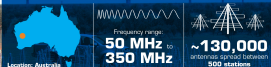
## SKA1 MID - the SKA's mid-frequency instrument

The Square Kilometre Array (SKA) will be the world's largest radio telescope, revolutionising our understanding of the Universe. The SKA will be built in two phases - SKA1 and SKA2 - starting in 2018, with SKA1 representing a fraction of the full SKA. SKA1 will include two instruments - SKA1 MID and SKA1 LOW - observing the Universe at different frequencies.



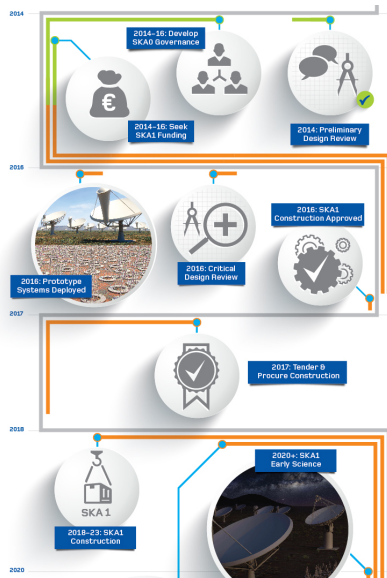
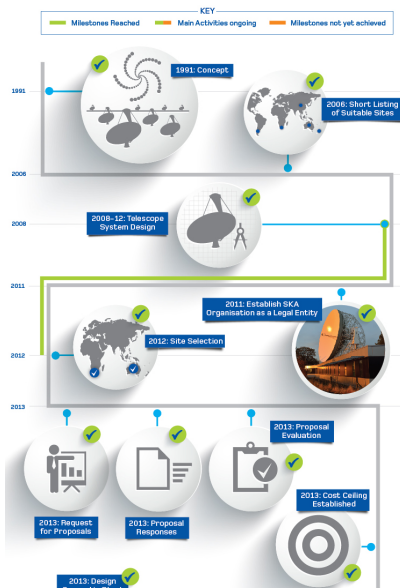
## SKA1 LOW - the SKA's low-frequency instrument

The Square Kilometre Array (SKA) will be the world's largest radio telescope, revolutionising our understanding of the Universe. The SKA will be built in two phases - SKA1 and SKA2 - starting in 2018, with SKA1 representing a fraction of the full SKA. SKA1 will include two instruments - SKA1 MID and SKA1 LOW - observing the Universe at different frequencies.




# SKA timeline

## High-level SKA Schedule








# The SKA poses a considerable big-data challenge

## Astronomical Data Deluge




### Square Kilometre Array

-  + A €1.5b billion global science project
-  + Astronomers and engineers from more than 70 institutes in 20 countries
-  + 3000 dishes, each 15m wide
-  + Using enough optical fibre to wrap twice around the Earth
-  + A combined collecting area of about one square kilometre

In excess of 1 Exabyte of raw data in a single day - more than the entire daily internet traffic


**@**

**Megadata**




IBM Information Intensive Framework


A prototype software architecture to manage the megadata generated by SKA



- + Automated data classification = faster with fewer errors
- + Guided search = easier access for scientists and non-scientists alike
- + Frees researchers to be more productive and creative



Enough raw data to fill over 15 million 64GB iPods every day



Top image: SPDO/Swinburne Astronomy Productions

# The SKA poses a considerable big-data challenge

## Astronomical Data Deluge



### Square Kilometre Array



+ A €1.5 billion global science project



+ Astronomers and engineers from more than 70 institutes in 20 countries



+ 3000 dishes, each 15m wide



+ Using enough optical fibre to wrap twice around the Earth

1,000,000 m<sup>2</sup>

+ A combined collecting area of about one square kilometre

Megadata



In excess of 1 Exabyte of raw data in a single day - more than the entire daily internet traffic



Enough raw data to fill over 15 million 64GB iPods every day



- + Automated data classification = faster with fewer errors
- + Guided search = easier access for scientists and non-scientists alike
- + Frees researchers to be more productive and creative

IBM  
Information  
Intensive  
Framework

A prototype software architecture to manage the megadata generated by SKA



# Outline

- 1 Compressive sensing and sparse regularisation
  - Introductory review
  - Analysis vs synthesis
  - Bayesian interpretations
- 2 Interferometric imaging with compressive sensing
  - Imaging
  - SARA algorithm
  - Continuous visibilities
- 3 Optimising telescope configurations
  - Spread spectrum effect
  - Simulations



# Outline

- 1 Compressive sensing and sparse regularisation
  - Introductory review
  - Analysis vs synthesis
  - Bayesian interpretations
- 2 Interferometric imaging with compressive sensing
  - Imaging
  - SARA algorithm
  - Continuous visibilities
- 3 Optimising telescope configurations
  - Spread spectrum effect
  - Simulations





# Compressive sensing

- Developed by Candes *et al.* 2006 and Donoho 2006 (and others).
- Although many underlying ideas around for a long time.
- Exploits the **sparsity** of natural signals.
- Very active area of research with many new developments.
- Acquisition versus imaging.



(a) Emmanuel Candes



(b) David Donoho



# Compressive sensing

- Developed by Candes *et al.* 2006 and Donoho 2006 (and others).
- Although many underlying ideas around for a long time.
- Exploits the **sparsity** of natural signals.
- Very active area of research with many new developments.
- **Acquisition** versus **imaging**.



(a) Emmanuel Candes



(b) David Donoho



# An introduction to compressive sensing

## Operator description

- Linear operator (linear algebra) representation of **signal decomposition**:

$$x(t) = \sum_i \alpha_i \Psi_i(t) \quad \rightarrow \quad \mathbf{x} = \sum_i \Psi_i \alpha_i = \begin{pmatrix} | \\ \Psi_0 \\ | \end{pmatrix} \alpha_0 + \begin{pmatrix} | \\ \Psi_1 \\ | \end{pmatrix} \alpha_1 + \dots \quad \rightarrow \quad \boxed{\mathbf{x} = \Psi \boldsymbol{\alpha}}$$

- Linear operator (linear algebra) representation of **measurement**:

$$y_i = \langle x, \Phi_j \rangle \quad \rightarrow \quad \mathbf{y} = \begin{pmatrix} - \Phi_0 - \\ - \Phi_1 - \\ \vdots \end{pmatrix} \mathbf{x} \quad \rightarrow \quad \boxed{\mathbf{y} = \Phi \mathbf{x}}$$

- Putting it together:

$$\boxed{\mathbf{y} = \Phi \mathbf{x} = \Phi \Psi \boldsymbol{\alpha}}$$



# An introduction to compressive sensing

## Operator description

- Linear operator (linear algebra) representation of **signal decomposition**:

$$x(t) = \sum_i \alpha_i \Psi_i(t) \quad \rightarrow \quad \mathbf{x} = \sum_i \Psi_i \alpha_i = \begin{pmatrix} | \\ \Psi_0 \\ | \end{pmatrix} \alpha_0 + \begin{pmatrix} | \\ \Psi_1 \\ | \end{pmatrix} \alpha_1 + \dots \quad \rightarrow \quad \boxed{\mathbf{x} = \Psi \boldsymbol{\alpha}}$$

- Linear operator (linear algebra) representation of **measurement**:

$$y_i = \langle x, \Phi_j \rangle \quad \rightarrow \quad \mathbf{y} = \begin{pmatrix} - \Phi_0 - \\ - \Phi_1 - \\ \vdots \end{pmatrix} \mathbf{x} \quad \rightarrow \quad \boxed{\mathbf{y} = \Phi \mathbf{x}}$$

- Putting it together:

$$\boxed{\mathbf{y} = \Phi \mathbf{x} = \Phi \Psi \boldsymbol{\alpha}}$$



# An introduction to compressive sensing

## Operator description

- Linear operator (linear algebra) representation of **signal decomposition**:

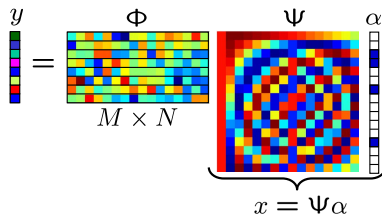
$$x(t) = \sum_i \alpha_i \Psi_i(t) \rightarrow x = \sum_i \Psi_i \alpha_i = \begin{pmatrix} | \\ \Psi_0 \\ | \end{pmatrix} \alpha_0 + \begin{pmatrix} | \\ \Psi_1 \\ | \end{pmatrix} \alpha_1 + \dots \rightarrow \boxed{x = \Psi \alpha}$$

- Linear operator (linear algebra) representation of **measurement**:

$$y_i = \langle x, \Phi_j \rangle \rightarrow y = \begin{pmatrix} - \Phi_0 - \\ - \Phi_1 - \\ \vdots \end{pmatrix} x \rightarrow \boxed{y = \Phi x}$$

- Putting it together:

$$\boxed{y = \Phi x = \Phi \Psi \alpha}$$



# An introduction to compressive sensing

## Promoting sparsity via $\ell_1$ minimisation

- Ill-posed inverse problem:

$$y = \Phi x + n = \Phi \Psi \alpha + n$$

- Solve by imposing a regularising prior that the signal to be recovered is sparse in  $\Psi$ , *i.e.* solve the following  $\ell_0$  optimisation problem:

$$\alpha^* = \arg \min_{\alpha} \|\alpha\|_0 \text{ such that } \|y - \Phi \Psi \alpha\|_2 \leq \epsilon$$

where the signal is synthesising by  $x^* = \Psi \alpha^*$ .

- Recall norms given by:

$$\|\alpha\|_0 = \text{no. non-zero elements} \quad \|\alpha\|_1 = \sum_i |\alpha_i| \quad \|\alpha\|_2 = \left( \sum_i |\alpha_i|^2 \right)^{1/2}$$

- Solving this problem is **difficult** (combinatorial).
- Instead, solve the  $\ell_1$  optimisation problem (convex):

$$\alpha^* = \arg \min_{\alpha} \|\alpha\|_1 \text{ such that } \|y - \Phi \Psi \alpha\|_2 \leq \epsilon$$



# An introduction to compressive sensing

## Promoting sparsity via $\ell_1$ minimisation

- Ill-posed inverse problem:

$$y = \Phi x + n = \Phi \Psi \alpha + n$$

- Solve by imposing a regularising prior that the signal to be recovered is sparse in  $\Psi$ , *i.e.* solve the following  $\ell_0$  optimisation problem:

$$\alpha^* = \arg \min_{\alpha} \|\alpha\|_0 \text{ such that } \|y - \Phi \Psi \alpha\|_2 \leq \epsilon$$

where the signal is synthesising by  $x^* = \Psi \alpha^*$ .

- Recall norms given by:

$$\|\alpha\|_0 = \text{no. non-zero elements} \quad \|\alpha\|_1 = \sum_i |\alpha_i| \quad \|\alpha\|_2 = \left( \sum_i |\alpha_i|^2 \right)^{1/2}$$

- Solving this problem is **difficult** (combinatorial).
- Instead, solve the  $\ell_1$  optimisation problem (convex):

$$\alpha^* = \arg \min_{\alpha} \|\alpha\|_1 \text{ such that } \|y - \Phi \Psi \alpha\|_2 \leq \epsilon$$



# An introduction to compressive sensing

## Promoting sparsity via $\ell_1$ minimisation

- Ill-posed inverse problem:

$$y = \Phi x + n = \Phi \Psi \alpha + n$$

- Solve by imposing a regularising prior that the signal to be recovered is sparse in  $\Psi$ , *i.e.* solve the following  $\ell_0$  optimisation problem:

$$\alpha^* = \arg \min_{\alpha} \|\alpha\|_0 \text{ such that } \|y - \Phi \Psi \alpha\|_2 \leq \epsilon$$

where the signal is synthesising by  $x^* = \Psi \alpha^*$ .

- Recall norms given by:

$$\|\alpha\|_0 = \text{no. non-zero elements} \quad \|\alpha\|_1 = \sum_i |\alpha_i| \quad \|\alpha\|_2 = \left( \sum_i |\alpha_i|^2 \right)^{1/2}$$

- Solving this problem is **difficult** (combinatorial).
- Instead, solve the  $\ell_1$  optimisation problem (convex):

$$\alpha^* = \arg \min_{\alpha} \|\alpha\|_1 \text{ such that } \|y - \Phi \Psi \alpha\|_2 \leq \epsilon$$





# An introduction to compressive sensing

## Promoting sparsity via $\ell_1$ minimisation

- Ill-posed inverse problem:

$$y = \Phi x + n = \Phi \Psi \alpha + n$$

- Solve by imposing a regularising prior that the signal to be recovered is sparse in  $\Psi$ , *i.e.* solve the following  $\ell_0$  optimisation problem:

$$\alpha^* = \arg \min_{\alpha} \|\alpha\|_0 \text{ such that } \|y - \Phi \Psi \alpha\|_2 \leq \epsilon$$

where the signal is synthesising by  $x^* = \Psi \alpha^*$ .

- Recall norms given by:

$$\|\alpha\|_0 = \text{no. non-zero elements} \quad \|\alpha\|_1 = \sum_i |\alpha_i| \quad \|\alpha\|_2 = \left( \sum_i |\alpha_i|^2 \right)^{1/2}$$

- Solving this problem is **difficult** (combinatorial).
- Instead, solve the  $\ell_1$  optimisation problem (convex):

$$\alpha^* = \arg \min_{\alpha} \|\alpha\|_1 \text{ such that } \|y - \Phi \Psi \alpha\|_2 \leq \epsilon$$



# An introduction to compressive sensing

## Union of subspaces

- Space of sparse vectors given by the **union of subspaces** aligned with the coordinate axes.

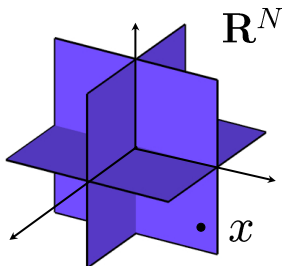


Figure: Space of the sparse vectors [Credit: Baraniuk]



# An introduction to compressive sensing

## Restricted isometry property (RIP)

- Solutions of  $\ell_0$  and  $\ell_1$  problems often the same.

- Restricted isometry property (RIP):

$$(1 - \delta_{2K})\|x_1 - x_2\|_2^2 \leq \|\Theta x_1 - \Theta x_2\|_2^2 \leq (1 + \delta_{2K})\|x_1 - x_2\|_2^2,$$

for  $K$ -sparse  $x_1$  and  $x_2$ , where  $\Theta = \Phi\Psi$ .

- Measurement must **preserve geometry** of sets of sparse vectors.

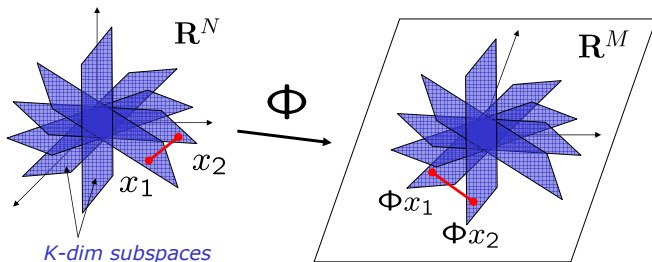


Figure: Measurement must preserve geometry of sets of sparse vectors. [Credit: Baraniuk]



# An introduction to compressive sensing

## Intuition

- Solutions of  $\ell_0$  and  $\ell_1$  problems often the same.
- Geometry of  $\ell_0$ ,  $\ell_2$  and  $\ell_1$  problems.

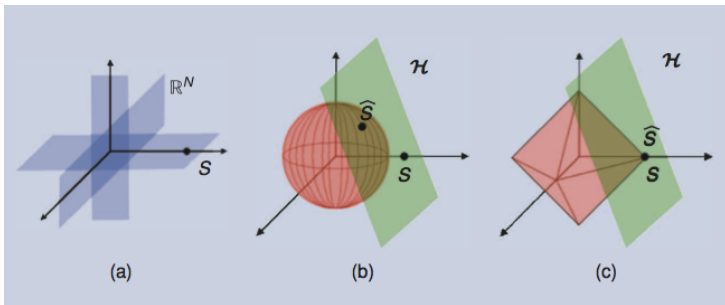


Figure: Geometry of (a)  $\ell_0$  (b)  $\ell_2$  and (c)  $\ell_1$  problems. [Credit: Baraniuk (2007)]



# An introduction to compressive sensing

## Sparsity and coherence

- In the absence of noise, compressed sensing is **exact!**
- Number of measurements required to achieve exact reconstruction is given by

$$M \geq c\mu^2 K \log N,$$

where  $K$  is the sparsity and  $N$  the dimensionality.

- The coherence between the measurement and sparsity basis is given by

$$\mu = \sqrt{N} \max_{i,j} |\langle \Psi_i, \Phi_j \rangle|.$$



# An introduction to compressive sensing

## Sparsity and coherence

- In the absence of noise, compressed sensing is **exact!**
- **Number of measurements** required to achieve exact reconstruction is given by

$$M \geq c\mu^2 K \log N ,$$

where  $K$  is the sparsity and  $N$  the dimensionality.

- The **coherence** between the measurement and sparsity basis is given by

$$\mu = \sqrt{N} \max_{i,j} |\langle \Psi_i, \Phi_j \rangle| .$$



# An introduction to compressive sensing

## Sparsity and coherence

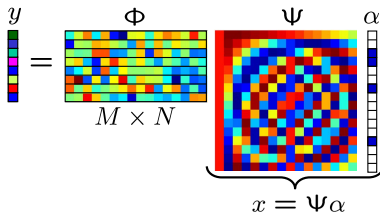
- In the absence of noise, compressed sensing is **exact!**
- Number of measurements** required to achieve exact reconstruction is given by

$$M \geq c\mu^2 K \log N,$$

where  $K$  is the sparsity and  $N$  the dimensionality.

- The **coherence** between the measurement and sparsity basis is given by

$$\mu = \sqrt{N} \max_{i,j} |\langle \Psi_i, \Phi_j \rangle|.$$



# Analysis vs synthesis

- Typically sparsity assumption is justified by analysing example signals in terms of atoms of the dictionary.
- But this is different to synthesising signals from atoms.
- Suggests an **analysis-based** framework (Elad *et al.* 2007, Nam *et al.* 2012):

$$x^* = \arg \min_x \|\Omega x\|_1 \text{ such that } \|y - \Phi x\|_2 \leq \epsilon.$$

analysis

- Contrast with **synthesis-based** approach:

$$x^* = \Psi \cdot \arg \min_{\alpha} \|\alpha\|_1 \text{ such that } \|y - \Phi \Psi \alpha\|_2 \leq \epsilon.$$

synthesis

- For **orthogonal bases**  $\Omega = \Psi^\dagger$  and the two approaches are **identical**.





# Analysis vs synthesis

- Typically sparsity assumption is justified by analysing example signals in terms of atoms of the dictionary.
- But this is different to synthesising signals from atoms.
- Suggests an **analysis-based** framework (Elad *et al.* 2007, Nam *et al.* 2012):

$$x^* = \arg \min_x \|\Omega x\|_1 \text{ such that } \|y - \Phi x\|_2 \leq \epsilon.$$

analysis

- Contrast with synthesis-based approach:

$$x^* = \Psi \cdot \arg \min_{\alpha} \|\alpha\|_1 \text{ such that } \|y - \Phi \Psi \alpha\|_2 \leq \epsilon.$$

synthesis

- For **orthogonal bases**  $\Omega = \Psi^\dagger$  and the two approaches are **identical**.



# Analysis vs synthesis

- Typically sparsity assumption is justified by analysing example signals in terms of atoms of the dictionary.
- But this is different to synthesising signals from atoms.
- Suggests an **analysis-based** framework (Elad *et al.* 2007, Nam *et al.* 2012):

$$\mathbf{x}^* = \arg \min_{\mathbf{x}} \|\Omega \mathbf{x}\|_1 \text{ such that } \|\mathbf{y} - \Phi \mathbf{x}\|_2 \leq \epsilon.$$

analysis

- Contrast with **synthesis-based** approach:

$$\mathbf{x}^* = \Psi \cdot \arg \min_{\boldsymbol{\alpha}} \|\boldsymbol{\alpha}\|_1 \text{ such that } \|\mathbf{y} - \Phi \Psi \boldsymbol{\alpha}\|_2 \leq \epsilon.$$

synthesis

- For **orthogonal bases**  $\Omega = \Psi^\dagger$  and the two approaches are **identical**.



# Analysis vs synthesis

- Typically sparsity assumption is justified by analysing example signals in terms of atoms of the dictionary.
- But this is different to synthesising signals from atoms.
- Suggests an **analysis-based** framework (Elad *et al.* 2007, Nam *et al.* 2012):

$$x^* = \arg \min_x \|\Omega x\|_1 \text{ such that } \|y - \Phi x\|_2 \leq \epsilon.$$

analysis

- Contrast with **synthesis-based** approach:

$$x^* = \Psi \cdot \arg \min_{\alpha} \|\alpha\|_1 \text{ such that } \|y - \Phi \Psi \alpha\|_2 \leq \epsilon.$$

synthesis

- For **orthogonal bases**  $\Omega = \Psi^\dagger$  and the two approaches are **identical**.



# Analysis vs synthesis

## Comparison

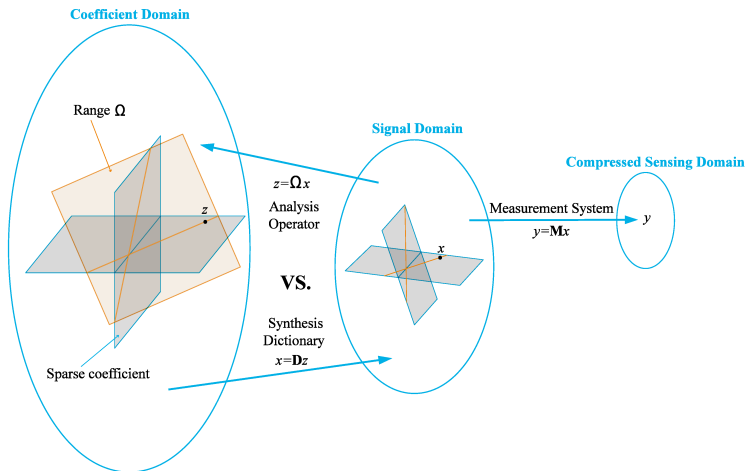


Figure: Analysis- and synthesis-based approaches [Credit: Nam *et al.* (2012)].



# Analysis vs synthesis

## Comparison

- **Synthesis-based approach is more general, while analysis-based approach more restrictive.**
- The more restrictive analysis-based approach may make it more robust to noise.
- The greater descriptive power of the synthesis-based approach may provide better signal representations (too descriptive?).



# Analysis vs synthesis

## Comparison

- **Synthesis-based** approach is **more general**, while **analysis-based** approach **more restrictive**.
- The more restrictive analysis-based approach may make it more robust to noise.
- The greater descriptive power of the synthesis-based approach may provide better signal representations (too descriptive?).



# Analysis vs synthesis

## Comparison

- **Synthesis-based** approach is **more general**, while **analysis-based** approach **more restrictive**.
- The more restrictive analysis-based approach may make it more robust to noise.
- The greater descriptive power of the synthesis-based approach may provide better signal representations (too descriptive?).



# Bayesian interpretations

## One Bayesian interpretation of the synthesis-based approach

- Consider the inverse problem:

$$\mathbf{y} = \Phi\Psi\boldsymbol{\alpha} + \mathbf{n} .$$

- Assume Gaussian noise, yielding the likelihood:

$$P(\mathbf{y} | \boldsymbol{\alpha}) \propto \exp\left(-\|\mathbf{y} - \Phi\Psi\boldsymbol{\alpha}\|_2^2 / (2\sigma^2)\right) .$$

- Consider the Laplacian prior:

$$P(\boldsymbol{\alpha}) \propto \exp\left(-\beta\|\boldsymbol{\alpha}\|_1\right) .$$

- The maximum *a-posteriori* (MAP) estimate (with  $\lambda = 2\beta\sigma^2$ ) is

$$\mathbf{x}_{\text{MAP-Synthesis}}^* = \Psi \cdot \arg \max_{\boldsymbol{\alpha}} P(\boldsymbol{\alpha} | \mathbf{y}) = \Psi \cdot \arg \min_{\boldsymbol{\alpha}} \|\mathbf{y} - \Phi\Psi\boldsymbol{\alpha}\|_2^2 + \lambda\|\boldsymbol{\alpha}\|_1 .$$

synthesis

- One possible Bayesian interpretation!
- Signal may be  $\ell_0$ -sparse, then solving  $\ell_1$  problem finds the correct  $\ell_0$ -sparse solution!





# Bayesian interpretations

## One Bayesian interpretation of the synthesis-based approach

- Consider the inverse problem:

$$\mathbf{y} = \Phi\Psi\boldsymbol{\alpha} + \mathbf{n} .$$

- Assume Gaussian noise, yielding the likelihood:

$$P(\mathbf{y} | \boldsymbol{\alpha}) \propto \exp\left(-\frac{\|\mathbf{y} - \Phi\Psi\boldsymbol{\alpha}\|_2^2}{2\sigma^2}\right) .$$

- Consider the Laplacian prior:

$$P(\boldsymbol{\alpha}) \propto \exp\left(-\beta\|\boldsymbol{\alpha}\|_1\right) .$$

- The **maximum *a-posteriori* (MAP) estimate** (with  $\lambda = 2\beta\sigma^2$ ) is

$$\mathbf{x}_{\text{MAP-Synthesis}}^* = \Psi \cdot \arg \max_{\boldsymbol{\alpha}} P(\boldsymbol{\alpha} | \mathbf{y}) = \Psi \cdot \arg \min_{\boldsymbol{\alpha}} \|\mathbf{y} - \Phi\Psi\boldsymbol{\alpha}\|_2^2 + \lambda\|\boldsymbol{\alpha}\|_1 .$$

synthesis

- One possible Bayesian interpretation!
- Signal may be  $\ell_0$ -sparse, then solving  $\ell_1$  problem finds the correct  $\ell_0$ -sparse solution!



# Bayesian interpretations

## One Bayesian interpretation of the synthesis-based approach

- Consider the inverse problem:

$$\mathbf{y} = \Phi\Psi\boldsymbol{\alpha} + \mathbf{n}.$$

- Assume Gaussian noise, yielding the likelihood:

$$P(\mathbf{y} | \boldsymbol{\alpha}) \propto \exp\left(-\|\mathbf{y} - \Phi\Psi\boldsymbol{\alpha}\|_2^2 / (2\sigma^2)\right).$$

- Consider the Laplacian prior:

$$P(\boldsymbol{\alpha}) \propto \exp\left(-\beta\|\boldsymbol{\alpha}\|_1\right).$$

- The **maximum *a-posteriori* (MAP) estimate** (with  $\lambda = 2\beta\sigma^2$ ) is

$$\mathbf{x}_{\text{MAP-Synthesis}}^* = \Psi \cdot \arg \max_{\boldsymbol{\alpha}} P(\boldsymbol{\alpha} | \mathbf{y}) = \Psi \cdot \arg \min_{\boldsymbol{\alpha}} \|\mathbf{y} - \Phi\Psi\boldsymbol{\alpha}\|_2^2 + \lambda\|\boldsymbol{\alpha}\|_1.$$

synthesis

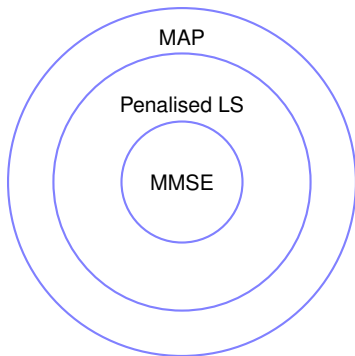
- One** possible Bayesian interpretation!
- Signal may be  $\ell_0$ -sparse**, then solving  $\ell_1$  problem finds the correct  $\ell_0$ -sparse solution!



# Bayesian interpretations

## Other Bayesian interpretations of the synthesis-based approach

- Other Bayesian interpretations are also possible (Gribonval 2011).
- Minimum mean square error (MMSE) estimators
  - synthesis-based estimators with appropriate penalty function, *i.e.* penalised least-squares (LS)
  - MAP estimators



# Bayesian interpretations

## One Bayesian interpretation of the analysis-based approach

- For the analysis-based approach, the MAP estimate is then

$$\mathbf{x}_{\text{MAP-Analysis}}^* = \arg \max_{\mathbf{x}} P(\mathbf{x} | \mathbf{y}) = \arg \min_{\mathbf{x}} \|\mathbf{y} - \Phi \mathbf{x}\|_2^2 + \lambda \|\Omega \mathbf{x}\|_1 .$$

analysis

- Identical to the synthesis-based approach if  $\Omega = \Psi^\dagger$ .
- But for **redundant dictionaries**, the analysis-based MAP estimate is

$$\mathbf{x}_{\text{MAP-Analysis}}^* = \Omega^\dagger \cdot \arg \min_{\gamma \in \text{column space } \Omega} \|\mathbf{y} - \Phi \Omega^\dagger \gamma\|_2^2 + \lambda \|\gamma\|_1 .$$

analysis

- Analysis-based approach **more restrictive** than synthesis-based.
- Similar ideas promoted by Maisinger & Hobson (2004) in a Bayesian framework for wavelet MEM (maximum entropy method).



# Bayesian interpretations

## One Bayesian interpretation of the analysis-based approach

- For the analysis-based approach, the MAP estimate is then

$$\mathbf{x}_{\text{MAP-Analysis}}^* = \arg \max_{\mathbf{x}} P(\mathbf{x} | \mathbf{y}) = \arg \min_{\mathbf{x}} \|\mathbf{y} - \Phi \mathbf{x}\|_2^2 + \lambda \|\Omega \mathbf{x}\|_1 .$$

analysis

- Identical to the synthesis-based approach if  $\Omega = \Psi^\dagger$ .
- But for **redundant dictionaries**, the analysis-based MAP estimate is

$$\mathbf{x}_{\text{MAP-Analysis}}^* = \Omega^\dagger \cdot \arg \min_{\gamma \in \text{column space } \Omega} \|\mathbf{y} - \Phi \Omega^\dagger \gamma\|_2^2 + \lambda \|\gamma\|_1 .$$

analysis

- Analysis-based approach **more restrictive** than synthesis-based.
- Similar ideas promoted by Maisinger & Hobson (2004) in a Bayesian framework for wavelet MEM (maximum entropy method).



# Bayesian interpretations

## One Bayesian interpretation of the analysis-based approach

- For the analysis-based approach, the MAP estimate is then

$$\mathbf{x}_{\text{MAP-Analysis}}^* = \arg \max_{\mathbf{x}} P(\mathbf{x} | \mathbf{y}) = \arg \min_{\mathbf{x}} \|\mathbf{y} - \Phi \mathbf{x}\|_2^2 + \lambda \|\Omega \mathbf{x}\|_1 .$$

analysis

- Identical to the synthesis-based approach if  $\Omega = \Psi^\dagger$ .
- But for **redundant dictionaries**, the analysis-based MAP estimate is

$$\mathbf{x}_{\text{MAP-Analysis}}^* = \Omega^\dagger \cdot \arg \min_{\gamma \in \text{column space } \Omega} \|\mathbf{y} - \Phi \Omega^\dagger \gamma\|_2^2 + \lambda \|\gamma\|_1 .$$

analysis

- Analysis-based approach **more restrictive** than synthesis-based.
- Similar ideas promoted by Maisinger & Hobson (2004) in a Bayesian framework for wavelet MEM (maximum entropy method).



# Bayesian interpretations

## One Bayesian interpretation of the analysis-based approach

- For the analysis-based approach, the MAP estimate is then

$$\mathbf{x}_{\text{MAP-Analysis}}^* = \arg \max_{\mathbf{x}} P(\mathbf{x} | \mathbf{y}) = \arg \min_{\mathbf{x}} \|\mathbf{y} - \Phi \mathbf{x}\|_2^2 + \lambda \|\Omega \mathbf{x}\|_1 .$$

analysis

- Identical to the synthesis-based approach if  $\Omega = \Psi^\dagger$ .
- But for **redundant dictionaries**, the analysis-based MAP estimate is

$$\mathbf{x}_{\text{MAP-Analysis}}^* = \Omega^\dagger \cdot \arg \min_{\gamma \in \text{column space } \Omega} \|\mathbf{y} - \Phi \Omega^\dagger \gamma\|_2^2 + \lambda \|\gamma\|_1 .$$

analysis

- Analysis-based approach **more restrictive** than synthesis-based.
- Similar ideas promoted by Maisinger & Hobson (2004) in a Bayesian framework for wavelet MEM (maximum entropy method).



# Outline

- 1 Compressive sensing and sparse regularisation
  - Introductory review
  - Analysis vs synthesis
  - Bayesian interpretations
- 2 Interferometric imaging with compressive sensing
  - Imaging
  - SARA algorithm
  - Continuous visibilities
- 3 Optimising telescope configurations
  - Spread spectrum effect
  - Simulations

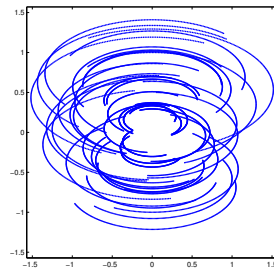




# Radio interferometric telescopes acquire “Fourier” measurements



“Fourier”  
Measurements



# Radio interferometric inverse problem

- The **complex visibility** measured by an interferometer is given by

$$y(\mathbf{u}, w) = \int_{D^2} A(\mathbf{l}) x(\mathbf{l}) C(\|\mathbf{l}\|_2) e^{-i2\pi\mathbf{u}\cdot\mathbf{l}} \frac{d^2\mathbf{l}}{n(\mathbf{l})},$$

visibilities

where the  $w$ -modulation  $C(\|\mathbf{l}\|_2)$  is given by

$$C(\|\mathbf{l}\|_2) \equiv e^{i2\pi w(1-\sqrt{1-\|\mathbf{l}\|^2})}.$$

$w$ -modulation

- Various assumptions are often made regarding the size of the **field-of-view**:

- Small-field with

$$\|\mathbf{l}\|^2 w \ll 1$$

 $\Rightarrow$ 

$$C(\|\mathbf{l}\|) \approx 1$$

- Small-field with

$$\|\mathbf{l}\|^2 w \ll 1$$

 $\Rightarrow$ 

$$C(\|\mathbf{l}\|) = e^{i2\pi w \|\mathbf{l}\|^2}$$

- Wide-field

 $\Rightarrow$ 

$$C(\|\mathbf{l}\|) = e^{i2\pi w(1-\sqrt{1-\|\mathbf{l}\|^2})}$$



# Radio interferometric inverse problem

- The **complex visibility** measured by an interferometer is given by

$$y(\mathbf{u}, \mathbf{w}) = \int_{D^2} A(\mathbf{l}) x(\mathbf{l}) C(\|\mathbf{l}\|_2) e^{-i2\pi\mathbf{u}\cdot\mathbf{l}} \frac{d^2\mathbf{l}}{n(\mathbf{l})},$$

visibilities

where the  $w$ -modulation  $C(\|\mathbf{l}\|_2)$  is given by

$$C(\|\mathbf{l}\|_2) \equiv e^{i2\pi w(1-\sqrt{1-\|\mathbf{l}\|^2})}.$$

$w$ -modulation

- Various assumptions are often made regarding the size of the **field-of-view**:

- Small-field with

$$\|\mathbf{l}\|^2 w \ll 1$$

$\Rightarrow$

$$C(\|\mathbf{l}\|_2) \simeq 1$$

- Small-field with

$$\|\mathbf{l}\|^4 w \ll 1$$

$\Rightarrow$

$$C(\|\mathbf{l}\|_2) \simeq e^{i\pi w \|\mathbf{l}\|^2}$$

- Wide-field

$\Rightarrow$

$$C(\|\mathbf{l}\|_2) = e^{i2\pi w(1-\sqrt{1-\|\mathbf{l}\|^2})}$$



# Radio interferometric inverse problem

- Consider the ill-posed inverse problem of radio interferometric imaging:

$$y = \Phi x + n,$$

where  $y$  are the measured visibilities,  $\Phi$  is the linear measurement operator,  $x$  is the underlying image and  $n$  is instrumental noise.

- Measurement operator  $\Phi = MFCA$  may incorporate:
  - primary beam  $A$  of the telescope;
  - w-modulation modulation  $C$ ;
  - Fourier transform  $F$ ;
  - masking  $M$  which encodes the incomplete measurements taken by the interferometer.



# Radio interferometric inverse problem

- Consider the ill-posed inverse problem of radio interferometric imaging:

$$y = \Phi x + n,$$

where  $y$  are the measured visibilities,  $\Phi$  is the linear measurement operator,  $x$  is the underlying image and  $n$  is instrumental noise.

- Measurement operator  $\Phi = MFCA$  may incorporate:
  - primary beam  $A$  of the telescope;
  - $w$ -modulation modulation  $C$ ;
  - Fourier transform  $F$ ;
  - masking  $M$  which encodes the incomplete measurements taken by the interferometer.



# Radio interferometric inverse problem

- Consider the ill-posed inverse problem of radio interferometric imaging:

$$y = \Phi x + n,$$

where  $y$  are the measured visibilities,  $\Phi$  is the linear measurement operator,  $x$  is the underlying image and  $n$  is instrumental noise.

- Measurement operator  $\Phi = MFCA$  may incorporate:
  - primary beam  $A$  of the telescope;
  - $w$ -modulation modulation  $C$ ;
  - Fourier transform  $F$ ;
  - masking  $M$  which encodes the incomplete measurements taken by the interferometer.

Interferometric imaging: recover an image from noisy and incomplete Fourier measurements.



# Interferometric imaging with compressed sensing

- Solve the interferometric imaging problem

$$y = \Phi x + n \quad \text{with} \quad \Phi = \mathbf{MFC A} ,$$

by applying a **prior on sparsity** of the signal in a **sparsifying dictionary**  $\Psi$ .

- Basis Pursuit (BP) denoising problem

$$\alpha^* = \arg \min_{\alpha} \|\alpha\|_1 \quad \text{such that} \quad \|y - \Phi \Psi \alpha\|_2 \leq \epsilon ,$$

BPDN

where the image is synthesised by  $x^* = \Psi \alpha^*$ .

- Total Variation (TV) denoising problem

$$x^* = \arg \min_x \|x\|_{\text{TV}} \quad \text{such that} \quad \|y - \Phi x\|_2 \leq \epsilon .$$

TVDN



# Interferometric imaging with compressed sensing

- Solve the interferometric imaging problem

$$y = \Phi x + n \quad \text{with} \quad \Phi = \mathbf{MFC A} ,$$

by applying a **prior on sparsity** of the signal in a **sparsifying dictionary**  $\Psi$ .

- Basis Pursuit (BP) denoising problem

$$\alpha^* = \arg \min_{\alpha} \|\alpha\|_1 \quad \text{such that} \quad \|y - \Phi \Psi \alpha\|_2 \leq \epsilon ,$$

BP/DN

where the image is synthesised by  $x^* = \Psi \alpha^*$ .

- Total Variation (TV) denoising problem

$$x^* = \arg \min_x \|x\|_{\text{TV}} \quad \text{such that} \quad \|y - \Phi x\|_2 \leq \epsilon .$$

TV/DN





# Interferometric imaging with compressed sensing

- Solve the interferometric imaging problem

$$y = \Phi x + n \quad \text{with} \quad \Phi = \mathbf{MFC A} \quad ,$$

by applying a **prior on sparsity** of the signal in a **sparsifying dictionary**  $\Psi$ .

- Basis Pursuit (BP) denoising problem

$$\alpha^* = \arg \min_{\alpha} \|\alpha\|_1 \quad \text{such that} \quad \|y - \Phi \Psi \alpha\|_2 \leq \epsilon, \quad \text{BPDN}$$

where the image is synthesised by  $x^* = \Psi \alpha^*$ .

- Total Variation (TV) denoising problem

$$x^* = \arg \min_x \|x\|_{\text{TV}} \quad \text{such that} \quad \|y - \Phi x\|_2 \leq \epsilon. \quad \text{TVDN}$$



# SARA for radio interferometric imaging

## Algorithm

- Sparsity averaging reweighted analysis (**SARA**) for RI imaging (Carrillo, McEwen & Wiaux 2012)
- Consider a dictionary composed of a concatenation of orthonormal bases, i.e.

$$\Psi = \frac{1}{\sqrt{q}} [\Psi_1, \Psi_2, \dots, \Psi_q],$$

thus  $\Psi \in \mathbb{R}^{N \times D}$  with  $D = qN$ .

- We consider the following bases: Dirac (i.e. pixel basis); Haar wavelets (promotes gradient sparsity); Daubechies wavelet bases two to eight.  
 $\Rightarrow$  concatenation of 9 bases
- Promote average sparsity by solving the reweighted  $\ell_1$  analysis problem:

$$\min_{\bar{x} \in \mathbb{R}^N} \|W\Psi^T \bar{x}\|_1 \quad \text{subject to} \quad \|\mathbf{y} - \Phi \bar{x}\|_2 \leq \epsilon \quad \text{and} \quad \bar{x} \geq 0,$$

SARA

where  $W \in \mathbb{R}^{D \times D}$  is a diagonal matrix with positive weights.

- Solve a sequence of reweighted  $\ell_1$  problems using the solution of the previous problem as the inverse weights  $\rightarrow$  approximate the  $\ell_0$  problem.



# SARA for radio interferometric imaging

## Algorithm

- Sparsity averaging reweighted analysis (**SARA**) for RI imaging (Carrillo, McEwen & Wiaux 2012)
- Consider a dictionary composed of a **concatenation of orthonormal bases**, i.e.

$$\Psi = \frac{1}{\sqrt{q}} [\Psi_1, \Psi_2, \dots, \Psi_q],$$

thus  $\Psi \in \mathbb{R}^{N \times D}$  with  $D = qN$ .

- We consider the following bases: **Dirac** (i.e. pixel basis); **Haar wavelets** (promotes gradient sparsity); **Daubechies wavelet bases two to eight**.  
 $\Rightarrow$  concatenation of 9 bases
- Promote average sparsity by solving the **reweighted  $\ell_1$  analysis problem**:

$$\min_{\bar{x} \in \mathbb{R}^N} \|W\Psi^T \bar{x}\|_1 \quad \text{subject to} \quad \|\mathbf{y} - \Phi \bar{x}\|_2 \leq \epsilon \quad \text{and} \quad \bar{x} \geq 0,$$

SARA

where  $W \in \mathbb{R}^{D \times D}$  is a diagonal matrix with positive weights.

- Solve a sequence of reweighted  $\ell_1$  problems using the solution of the previous problem as the inverse weights  $\rightarrow$  approximate the  $\ell_0$  problem.



# SARA for radio interferometric imaging

## Algorithm

- Sparsity averaging reweighted analysis (**SARA**) for RI imaging (Carrillo, McEwen & Wiaux 2012)
- Consider a dictionary composed of a **concatenation of orthonormal bases**, i.e.

$$\Psi = \frac{1}{\sqrt{q}} [\Psi_1, \Psi_2, \dots, \Psi_q],$$

thus  $\Psi \in \mathbb{R}^{N \times D}$  with  $D = qN$ .

- We consider the following bases: **Dirac** (i.e. pixel basis); **Haar wavelets** (promotes gradient sparsity); **Daubechies wavelet bases two to eight**.  
 $\Rightarrow$  concatenation of 9 bases
- Promote average sparsity by solving the **reweighted  $\ell_1$  analysis problem**:

$$\min_{\bar{x} \in \mathbb{R}^N} \|W\Psi^T \bar{x}\|_1 \quad \text{subject to} \quad \|y - \Phi \bar{x}\|_2 \leq \epsilon \quad \text{and} \quad \bar{x} \geq 0,$$

SARA

where  $W \in \mathbb{R}^{D \times D}$  is a diagonal matrix with positive weights.

- Solve a sequence of reweighted  $\ell_1$  problems using the solution of the previous problem as the inverse weights  $\rightarrow$  approximate the  $\ell_0$  problem.



# SARA for radio interferometric imaging

## Algorithm

- Sparsity averaging reweighted analysis (**SARA**) for RI imaging (Carrillo, McEwen & Wiaux 2012)
- Consider a dictionary composed of a **concatenation of orthonormal bases**, i.e.

$$\Psi = \frac{1}{\sqrt{q}} [\Psi_1, \Psi_2, \dots, \Psi_q],$$

thus  $\Psi \in \mathbb{R}^{N \times D}$  with  $D = qN$ .

- We consider the following bases: **Dirac** (i.e. pixel basis); **Haar wavelets** (promotes gradient sparsity); **Daubechies wavelet bases two to eight**.  
 $\Rightarrow$  concatenation of 9 bases
- Promote average sparsity by solving the **reweighted  $\ell_1$  analysis problem**:

$$\min_{\bar{x} \in \mathbb{R}^N} \|W\Psi^T \bar{x}\|_1 \quad \text{subject to} \quad \|y - \Phi \bar{x}\|_2 \leq \epsilon \quad \text{and} \quad \bar{x} \geq 0,$$

SARA

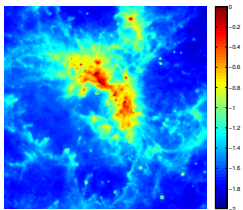
where  $W \in \mathbb{R}^{D \times D}$  is a diagonal matrix with positive weights.

- Solve a sequence of reweighted  $\ell_1$  problems using the solution of the previous problem as the inverse weights  $\rightarrow$  **approximate the  $\ell_0$  problem**.



# SARA for radio interferometric imaging

## Results on simulations

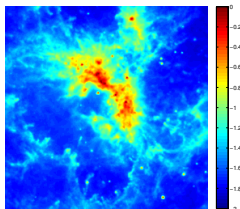


(a) Original

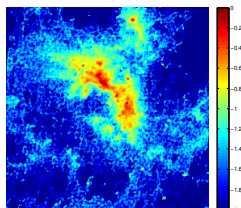


# SARA for radio interferometric imaging

## Results on simulations



(a) Original

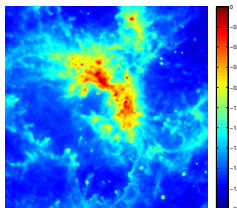


(b) "CLEAN" (SNR=16.67 dB)

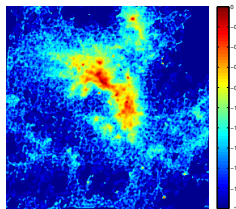


# SARA for radio interferometric imaging

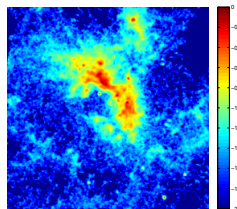
## Results on simulations



(a) Original



(b) "CLEAN" (SNR=16.67 dB)



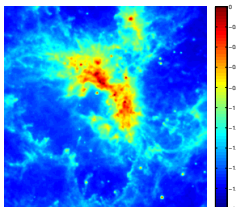
(c) "MS-CLEAN" (SNR=17.87 dB)



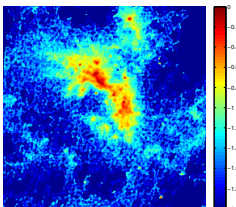


# SARA for radio interferometric imaging

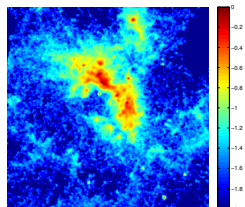
## Results on simulations



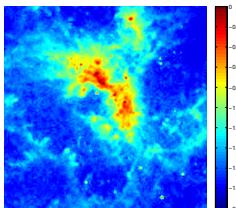
(a) Original



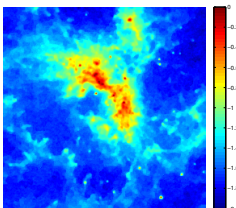
(b) "CLEAN" (SNR=16.67 dB)



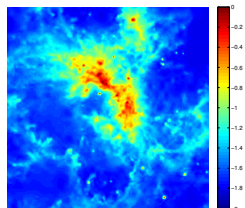
(c) "MS-CLEAN" (SNR=17.87 dB)



(d) BPDb8 (SNR=24.53 dB)



(e) TV (SNR=26.47 dB)



(f) SARA (SNR=29.08 dB)



# SARA for radio interferometric imaging

## Results on simulations

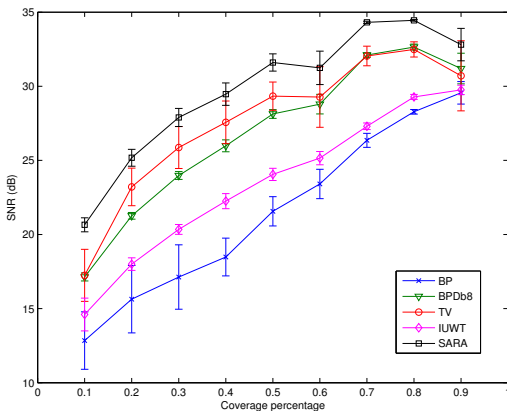


Figure: Reconstruction fidelity vs visibility coverage.



# SARA for natural imaging



(a) Original



(b) Daubechies 8



(c) SARA

**Figure:** Cameraman reconstruction from 30% of Fourier measurements.



# SARA for natural imaging

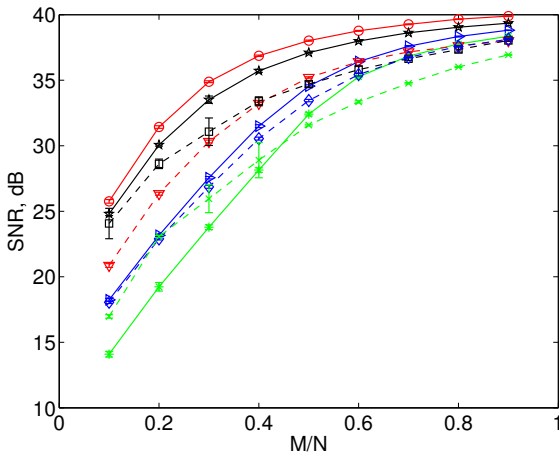


Figure: Reconstruction fidelity vs measurement ratio for Cameraman.



# SARA for natural images

## Results on simulations



(a) Original

(b) Daubechies 8

(c) SARA

**Figure:** Lena reconstruction from 30% of Fourier measurements.



# SARA for natural images

## Results on simulations

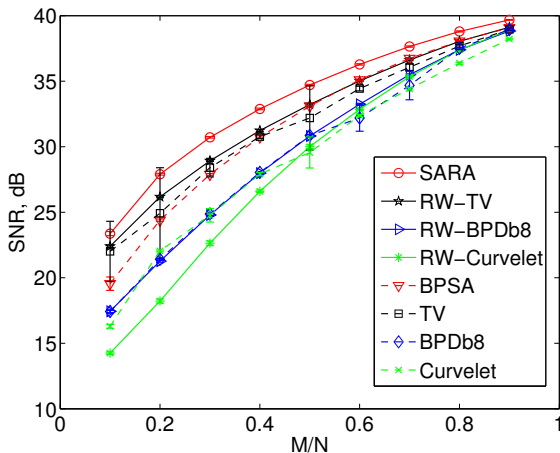


Figure: Reconstruction fidelity vs measurement ratio for Lena.



# Supporting continuous visibilities

## Algorithm

- Ideally we would like to model the **continuous Fourier transform operator**

$$\Phi = \mathbf{F}^c .$$

- But this is **impracticably slow!**
- Incorporated gridding into our CS interferometric imaging framework (Carrillo *et al.* 2013).
- Model with measurement operator

$$\Phi = \mathbf{GFDZ} ,$$

where we incorporate:

- convolutional **gridding operator**  $\mathbf{G}$ ;
- fast Fourier transform  $\mathbf{F}$ ;
- normalisation operator  $\mathbf{D}$  to undo the convolution gridding;
- zero-padding operator  $\mathbf{Z}$  to upsample the discrete visibility space.



# Supporting continuous visibilities

## Algorithm

- Ideally we would like to model the continuous Fourier transform operator

$$\Phi = \mathbf{F}^c .$$

- But this is impracticably slow!
- Incorporated gridding into our CS interferometric imaging framework (Carrillo *et al.* 2013).
- Model with measurement operator

$$\Phi = \mathbf{GFDZ} ,$$

where we incorporate:

- convolutional gridding operator  $\mathbf{G}$ ;
- fast Fourier transform  $\mathbf{F}$ ;
- normalisation operator  $\mathbf{D}$  to undo the convolution gridding;
- zero-padding operator  $\mathbf{Z}$  to upsample the discrete visibility space.





# Supporting continuous visibilities

## Algorithm

- Ideally we would like to model the **continuous Fourier transform operator**

$$\Phi = \mathbf{F}^c .$$

- But this is **impracticably slow!**
- Incorporated gridding into our CS interferometric imaging framework (Carrillo *et al.* 2013).
- Model with measurement operator

$$\Phi = \mathbf{GFDZ} ,$$

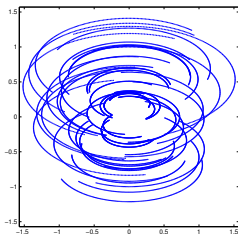
where we incorporate:

- convolutional **gridding operator G**;
- **fast Fourier transform F**;
- **normalisation operator D** to undo the convolution gridding;
- **zero-padding operator Z** to upsample the discrete visibility space.

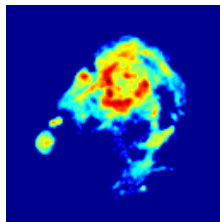


# Supporting continuous visibilities

## Results on simulations



(a) Coverage



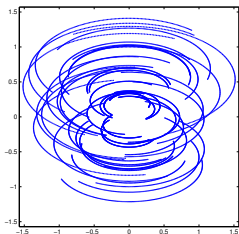
(b) M31 (ground truth)

Figure: Reconstructed images from continuous visibilities.

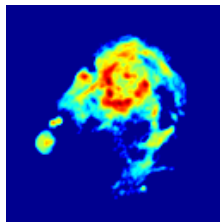


# Supporting continuous visibilities

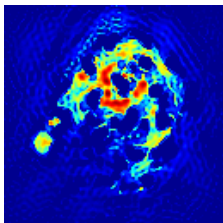
## Results on simulations



(a) Coverage



(b) M31 (ground truth)



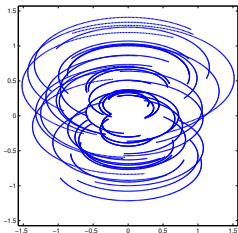
(c) "CLEAN" (SNR= 8.2dB)

Figure: Reconstructed images from continuous visibilities.

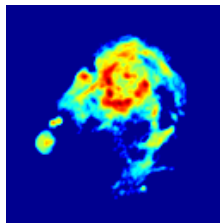


# Supporting continuous visibilities

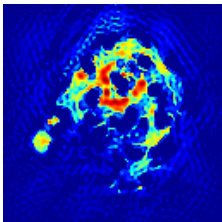
## Results on simulations



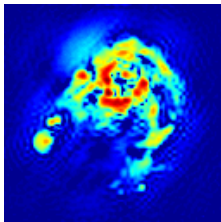
(a) Coverage



(b) M31 (ground truth)



(c) "CLEAN" (SNR= 8.2dB)



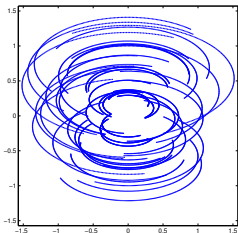
(d) "MS-CLEAN" (SNR= 11.1dB)

Figure: Reconstructed images from continuous visibilities.

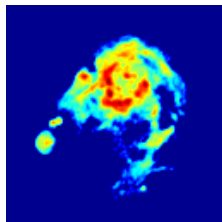


# Supporting continuous visibilities

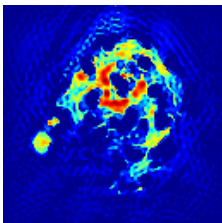
## Results on simulations



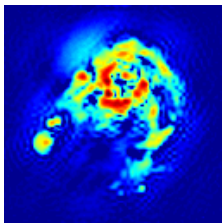
(a) Coverage



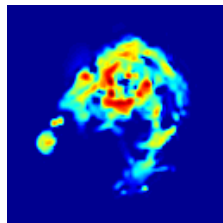
(b) M31 (ground truth)



(c) "CLEAN" (SNR= 8.2dB)



(d) "MS-CLEAN" (SNR= 11.1dB)



(e) SARA (SNR= 13.4dB)

Figure: Reconstructed images from continuous visibilities.



# Outline

- 1 Compressive sensing and sparse regularisation
  - Introductory review
  - Analysis vs synthesis
  - Bayesian interpretations
- 2 Interferometric imaging with compressive sensing
  - Imaging
  - SARA algorithm
  - Continuous visibilities
- 3 Optimising telescope configurations
  - Spread spectrum effect
  - Simulations



# Optimising telescope configurations

## Spread spectrum effect

- Use theory of compressive sensing to optimise telescope configurations.
- Non-coplanar baselines and wide fields  $\rightarrow w$ -modulation  $\rightarrow$  spread spectrum effect  $\rightarrow$  improves reconstruction quality (first considered by Wiaux *et al.* 2009b).
- The  $w$ -modulation operator  $\mathbf{C}$  has elements defined by

$$C(l, m) \equiv e^{i2\pi w(1 - \sqrt{1 - l^2 - m^2})} \simeq e^{i\pi w \|l\|^2} \quad \text{for } \|l\|^4 w \ll 1,$$

giving rise to to a linear chirp.



# Optimising telescope configurations

## Spread spectrum effect

- Use theory of compressive sensing to optimise telescope configurations.
- Non-coplanar baselines and wide fields  $\rightarrow$   $w$ -modulation  $\rightarrow$  spread spectrum effect  $\rightarrow$  improves reconstruction quality (first considered by Wiaux *et al.* 2009b).
- The  $w$ -modulation operator  $\mathbb{C}$  has elements defined by

$$C(l, m) \equiv e^{i2\pi w(1 - \sqrt{1 - l^2 - m^2})} \simeq e^{i\pi w \|l\|^2} \quad \text{for } \|l\|^4 w \ll 1,$$

giving rise to to a linear chirp.





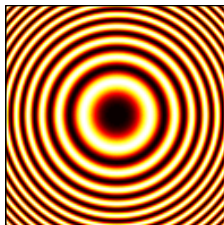
# Optimising telescope configurations

## Spread spectrum effect

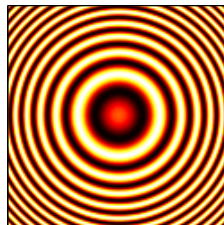
- Use theory of compressive sensing to optimise telescope configurations.
- Non-coplanar baselines and wide fields  $\rightarrow w$ -modulation  $\rightarrow$  spread spectrum effect  $\rightarrow$  improves reconstruction quality (first considered by Wiaux *et al.* 2009b).
- The  $w$ -modulation operator  $\mathbf{C}$  has elements defined by

$$C(l, m) \equiv e^{i2\pi w(1 - \sqrt{1 - l^2 - m^2})} \simeq e^{i\pi w \|l\|^2} \quad \text{for } \|l\|^4 w \ll 1,$$

giving rise to to a linear chirp.



(a) Real part



(b) Imaginary part

Figure: Chirp modulation.



# Recap compressive sensing preliminaries

## Sparsity and coherence

- What drives the quality of compressive sensing reconstruction?
- Number of measurements required to achieve exact reconstruction is given by

$$M \geq c\mu^2 K \log N,$$

where  $K$  is the sparsity and  $N$  the dimensionality.

- The coherence between the measurement and sparsity basis is given by

$$\mu = \sqrt{N} \max_{i,j} |\langle \Psi_i, \Phi_j \rangle|.$$



# Recap compressive sensing preliminaries

## Sparsity and coherence

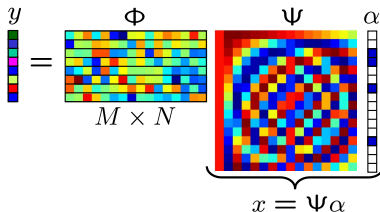
- What drives the quality of compressive sensing reconstruction?
- **Number of measurements** required to achieve exact reconstruction is given by

$$M \geq c\mu^2 K \log N,$$

where  $K$  is the sparsity and  $N$  the dimensionality.

- The **coherence** between the measurement and sparsity basis is given by

$$\mu = \sqrt{N} \max_{i,j} |\langle \Psi_i, \Phi_j \rangle|.$$



# Spread spectrum effect

## Overview

### *Spread spectrum effect in a nutshell*

- 1 Radio interferometers take (essentially) **Fourier measurements**.
- 2 Recall, the coherence is the maximum inner product between measurement vectors and sparsifying atoms.
- 3 Thus, **coherence** is (essentially) the **maximum of the Fourier coefficients** of the atoms of the sparsifying dictionary.
- 4  **$w$ -modulation spreads the spectrum** of the atoms of the sparsifying dictionary, reducing the maximum Fourier coefficient.
- 5 Spreading the spectrum **reduces coherence**, thus **improving reconstruction fidelity**.

- Consistent with findings of Carozzi et al. (2013) from information theoretic approach.
- Studied for constant  $w$  (for simplicity) by Wiaux *et al.* (2009b).
- Studied for varying  $w$  (with realistic images and various sparse representations) by Wolz *et al.* (2013).



# Spread spectrum effect

## Overview

### *Spread spectrum effect in a nutshell*

- 1 Radio interferometers take (essentially) **Fourier measurements**.
- 2 Recall, the coherence is the maximum inner product between measurement vectors and sparsifying atoms.
- 3 Thus, **coherence** is (essentially) the **maximum of the Fourier coefficients** of the atoms of the sparsifying dictionary.
- 4  **$w$ -modulation spreads the spectrum** of the atoms of the sparsifying dictionary, reducing the maximum Fourier coefficient.
- 5 Spreading the spectrum **reduces coherence**, thus **improving reconstruction fidelity**.

- Consistent with findings of Carozzi et al. (2013) from information theoretic approach.
- Studied for constant  $w$  (for simplicity) by Wiaux *et al.* (2009b).
- Studied for varying  $w$  (with realistic images and various sparse representations) by Wolz *et al.* (2013).



# Spread spectrum effect

## Overview

### *Spread spectrum effect in a nutshell*

- 1 Radio interferometers take (essentially) **Fourier measurements**.
- 2 Recall, the coherence is the maximum inner product between measurement vectors and sparsifying atoms.
- 3 Thus, **coherence** is (essentially) the **maximum of the Fourier coefficients** of the atoms of the sparsifying dictionary.
- 4  **$w$ -modulation spreads the spectrum** of the atoms of the sparsifying dictionary, reducing the maximum Fourier coefficient.
- 5 Spreading the spectrum **reduces coherence**, thus **improving reconstruction fidelity**.

- Consistent with findings of Carozzi et al. (2013) from information theoretic approach.
- Studied for constant  $w$  (for simplicity) by Wiaux *et al.* (2009b).
- Studied for varying  $w$  (with realistic images and various sparse representations) by Wolz *et al.* (2013).



# Spread spectrum effect

## Overview

### *Spread spectrum effect in a nutshell*

- 1 Radio interferometers take (essentially) **Fourier measurements**.
- 2 Recall, the coherence is the maximum inner product between measurement vectors and sparsifying atoms.
- 3 Thus, **coherence** is (essentially) the **maximum of the Fourier coefficients** of the atoms of the sparsifying dictionary.
- 4  **$w$ -modulation spreads the spectrum** of the atoms of the sparsifying dictionary, reducing the maximum Fourier coefficient.
- 5 Spreading the spectrum **reduces coherence**, thus **improving reconstruction fidelity**.

- Consistent with findings of Carozzi et al. (2013) from information theoretic approach.
- Studied for constant  $w$  (for simplicity) by Wiaux *et al.* (2009b).
- Studied for varying  $w$  (with realistic images and various sparse representations) by Wolz *et al.* (2013).



# Spread spectrum effect

## Overview

### *Spread spectrum effect in a nutshell*

- 1 Radio interferometers take (essentially) **Fourier measurements**.
- 2 Recall, the coherence is the maximum inner product between measurement vectors and sparsifying atoms.
- 3 Thus, **coherence** is (essentially) the **maximum of the Fourier coefficients** of the atoms of the sparsifying dictionary.
- 4  **$w$ -modulation spreads the spectrum** of the atoms of the sparsifying dictionary, reducing the maximum Fourier coefficient.
- 5 Spreading the spectrum **reduces coherence**, thus **improving reconstruction fidelity**.

- Consistent with findings of Carozzi et al. (2013) from information theoretic approach.
- Studied for constant  $w$  (for simplicity) by Wiaux *et al.* (2009b).
- Studied for varying  $w$  (with realistic images and various sparse representations) by Wolz *et al.* (2013).





# Spread spectrum effect

## Overview

### *Spread spectrum effect in a nutshell*

- 1 Radio interferometers take (essentially) **Fourier measurements**.
- 2 Recall, the coherence is the maximum inner product between measurement vectors and sparsifying atoms.
- 3 Thus, **coherence** is (essentially) the **maximum of the Fourier coefficients** of the atoms of the sparsifying dictionary.
- 4  **$w$ -modulation spreads the spectrum** of the atoms of the sparsifying dictionary, reducing the maximum Fourier coefficient.
- 5 Spreading the spectrum **reduces coherence**, thus **improving reconstruction fidelity**.

- Consistent with findings of Carozzi et al. (2013) from information theoretic approach.
- Studied for constant  $w$  (for simplicity) by Wiaux *et al.* (2009b).
- Studied for varying  $w$  (with realistic images and various sparse representations) by Wolz *et al.* (2013).



# Spread spectrum effect

## Overview

### *Spread spectrum effect in a nutshell*

- 1 Radio interferometers take (essentially) **Fourier measurements**.
- 2 Recall, the coherence is the maximum inner product between measurement vectors and sparsifying atoms.
- 3 Thus, **coherence** is (essentially) the **maximum of the Fourier coefficients** of the atoms of the sparsifying dictionary.
- 4  **$w$ -modulation spreads the spectrum** of the atoms of the sparsifying dictionary, reducing the maximum Fourier coefficient.
- 5 Spreading the spectrum **reduces coherence**, thus **improving reconstruction fidelity**.

- Consistent with findings of Carozzi et al. (2013) from information theoretic approach.
- Studied for constant  $w$  (for simplicity) by Wiaux *et al.* (2009b).
- Studied for varying  $w$  (with realistic images and various sparse representations) by Wolz *et al.* (2013).



# Spread spectrum effect

## Overview

### *Spread spectrum effect in a nutshell*

- 1 Radio interferometers take (essentially) **Fourier measurements**.
- 2 Recall, the coherence is the maximum inner product between measurement vectors and sparsifying atoms.
- 3 Thus, **coherence** is (essentially) the **maximum of the Fourier coefficients** of the atoms of the sparsifying dictionary.
- 4  **$w$ -modulation spreads the spectrum** of the atoms of the sparsifying dictionary, reducing the maximum Fourier coefficient.
- 5 Spreading the spectrum **reduces coherence**, thus **improving reconstruction fidelity**.

- Consistent with findings of Carozzi et al. (2013) from information theoretic approach.
- Studied for **constant  $w$**  (for simplicity) by Wiaux *et al.* (2009b).
- Studied for **varying  $w$**  (with realistic images and various sparse representations) by Wolz *et al.* (2013).



# Spread spectrum effect

## Sparse $w$ -projection algorithm

- Apply the  $w$ -projection algorithm (Cornwell *et al.* 2008) to shift the  $w$ -modulation through the Fourier transform:

$$\Phi = \mathbf{MFC}\mathbf{A} \Rightarrow \Phi = \hat{\mathbf{C}}\mathbf{F}\mathbf{A} .$$

- Naively, expressing the application of the  $w$ -modulation in this manner is computationally less efficient than the original formulation but it has **two important advantages**.
- Different  $w$  for each  $(u, v)$ , while still exploiting FFT.
- Many of the elements of  $\hat{\mathbf{C}}$  will be close to zero.
- Support of  $w$ -modulation in Fourier space determined dynamically.



# Spread spectrum effect

## Sparse $w$ -projection algorithm

- Apply the  $w$ -projection algorithm (Cornwell *et al.* 2008) to shift the  $w$ -modulation through the Fourier transform:

$$\Phi = \mathbf{MFC}\mathbf{A} \Rightarrow \Phi = \hat{\mathbf{C}}\mathbf{F}\mathbf{A}$$

- Naively, expressing the application of the  $w$ -modulation in this manner is computationally less efficient than the original formulation but it has **two important advantages**.
  - Different  $w$  for each  $(u, v)$ , while still exploiting FFT.
  - Many of the elements of  $\hat{\mathbf{C}}$  will be close to zero.
  - Support of  $w$ -modulation in Fourier space determined dynamically.



# Spread spectrum effect

## Sparse $w$ -projection algorithm

- Apply the  $w$ -projection algorithm (Cornwell *et al.* 2008) to shift the  $w$ -modulation through the Fourier transform:

$$\Phi = \mathbf{MFC}\mathbf{A} \quad \Rightarrow \quad \Phi = \hat{\mathbf{C}}\mathbf{F}\mathbf{A} .$$

- Naively, expressing the application of the  $w$ -modulation in this manner is computationally less efficient than the original formulation but it has **two important advantages**.
- Different  $w$  for each  $(u, v)$ , while still exploiting FFT.
- Many of the elements of  $\hat{\mathbf{C}}$  will be close to zero.
- Support of  $w$ -modulation in Fourier space determined dynamically.



# Spread spectrum effect

## Sparse $w$ -projection algorithm

- Apply the  $w$ -projection algorithm (Cornwell *et al.* 2008) to shift the  $w$ -modulation through the Fourier transform:

$$\Phi = \mathbf{MFC}\mathbf{A} \quad \Rightarrow \quad \Phi = \hat{\mathbf{C}}\mathbf{F}\mathbf{A} .$$

- Naively, expressing the application of the  $w$ -modulation in this manner is computationally less efficient than the original formulation but it has **two important advantages**.
- Different  $w$  for each  $(u, v)$ , while still exploiting FFT.
- Many of the elements of  $\hat{\mathbf{C}}$  will be close to zero.
- Support of  $w$ -modulation in Fourier space determined dynamically.



# Spread spectrum effect

## Sparse $w$ -projection algorithm

- Apply the  $w$ -projection algorithm (Cornwell *et al.* 2008) to shift the  $w$ -modulation through the Fourier transform:

$$\Phi = \mathbf{MFC}\mathbf{A} \Rightarrow \Phi = \hat{\mathbf{C}}\mathbf{F}\mathbf{A} .$$

- Naively, expressing the application of the  $w$ -modulation in this manner is computationally less efficient than the original formulation but it has **two important advantages**.
- Different  $w$  for each  $(u, v)$ , while still exploiting FFT.
- Many of the elements of  $\hat{\mathbf{C}}$  will be close to zero.
- Support** of  $w$ -modulation in Fourier space **determined dynamically**.





# Spread spectrum effect for varying $w$

## Results on simulations

- Perform simulations to assess the effectiveness of the spread spectrum effect in the presence of **varying  $w$** .
- Consider idealised simulations with uniformly random visibility sampling.

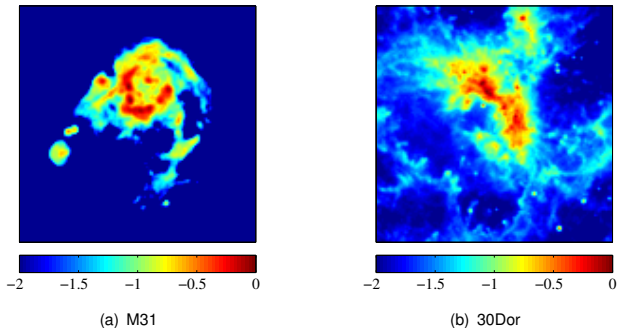
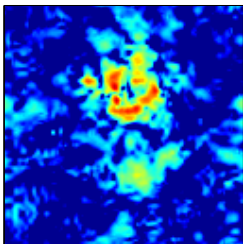


Figure: Ground truth images in logarithmic scale.



# Spread spectrum effect for varying $w$

## Results on simulations



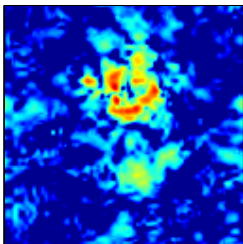
(a)  $w_d = 0 \rightarrow \text{SNR} = 5\text{dB}$

Figure: Reconstructed images of M31 for 10% coverage.

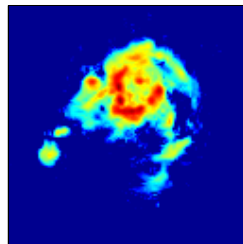


# Spread spectrum effect for varying $w$

## Results on simulations



(a)  $w_d = 0 \rightarrow \text{SNR} = 5\text{dB}$



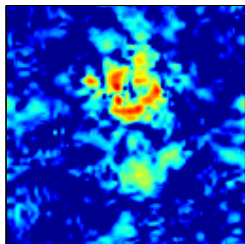
(c)  $w_d = 1 \rightarrow \text{SNR} = 19\text{dB}$

Figure: Reconstructed images of M31 for 10% coverage.

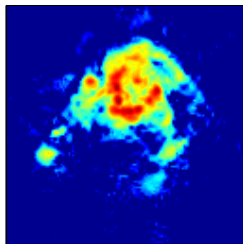


# Spread spectrum effect for varying $w$

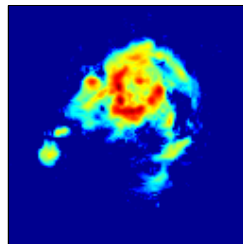
Results on simulations



(a)  $w_d = 0 \rightarrow \text{SNR} = 5\text{dB}$



(b)  $w_d \sim \mathcal{U}(0, 1) \rightarrow \text{SNR} = 16\text{dB}$



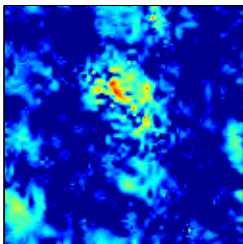
(c)  $w_d = 1 \rightarrow \text{SNR} = 19\text{dB}$

Figure: Reconstructed images of M31 for 10% coverage.



# Spread spectrum effect for varying $w$

## Results on simulations



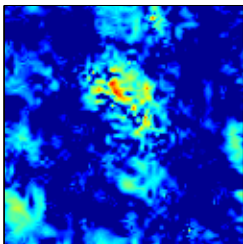
(a)  $w_d = 0 \rightarrow \text{SNR} = 2\text{dB}$

Figure: Reconstructed images of 30Dor for 10% coverage.

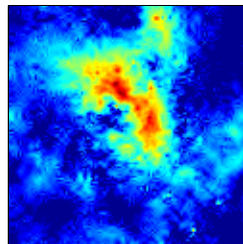


# Spread spectrum effect for varying $w$

Results on simulations



(a)  $w_d = 0 \rightarrow \text{SNR} = 2\text{dB}$



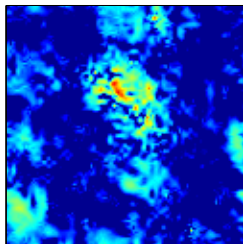
(c)  $w_d = 1 \rightarrow \text{SNR} = 15\text{dB}$

Figure: Reconstructed images of 30Dor for 10% coverage.

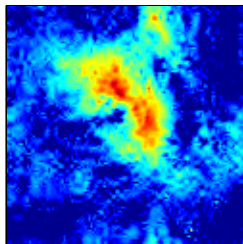


# Spread spectrum effect for varying $w$

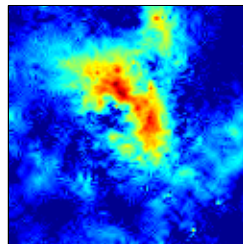
Results on simulations



(a)  $w_d = 0 \rightarrow \text{SNR} = 2\text{dB}$



(b)  $w_d \sim \mathcal{U}(0, 1) \rightarrow \text{SNR} = 12\text{dB}$



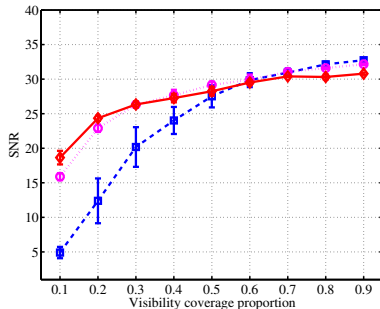
(c)  $w_d = 1 \rightarrow \text{SNR} = 15\text{dB}$

Figure: Reconstructed images of 30Dor for 10% coverage.



# Spread spectrum effect for varying $w$

## Results on simulations



(a) Daubechies 8 (Db8) wavelets

Figure: Reconstruction fidelity for M31.

Improvement in reconstruction fidelity due to the spread spectrum effect for varying  $w$  is almost as large as the case of constant maximum  $w$ .

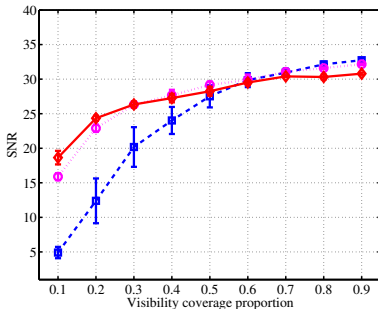
- As expected, for the case where coherence is already optimal, there is little improvement.





# Spread spectrum effect for varying $w$

## Results on simulations



(a) Daubechies 8 (Db8) wavelets

Figure: Reconstruction fidelity for M31.

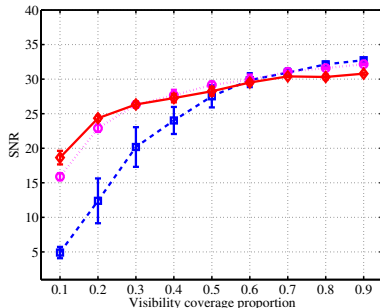
Improvement in reconstruction fidelity due to the spread spectrum effect for varying  $w$  is almost as large as the case of constant maximum  $w$ .

- As expected, for the case where coherence is already optimal, there is little improvement.

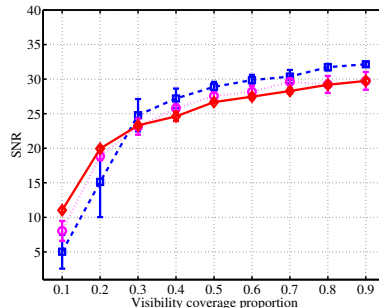


# Spread spectrum effect for varying $w$

## Results on simulations



(a) Daubechies 8 (Db8) wavelets



(b) Dirac basis

Figure: Reconstruction fidelity for M31.

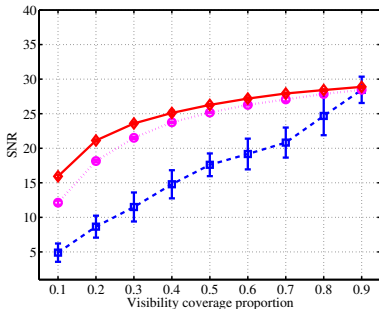
Improvement in reconstruction fidelity due to the spread spectrum effect for varying  $w$  is almost as large as the case of constant maximum  $w$ .

- As expected, for the case where coherence is already optimal, there is little improvement.

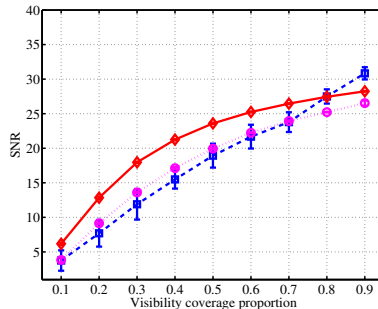


# Spread spectrum effect for varying $w$

## Results on simulations



(a) Daubechies 8 (Db8) wavelets



(b) Dirac basis

Figure: Reconstruction fidelity for 30Dor.

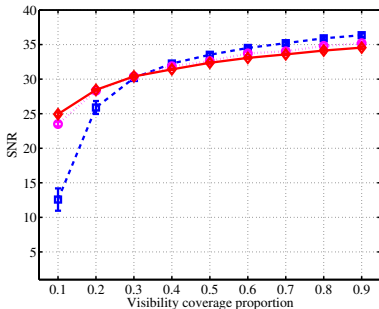
Improvement in reconstruction fidelity due to the spread spectrum effect for varying  $w$  is almost as large as the case of constant maximum  $w$ .

- As expected, for the case where coherence is already optimal, there is little improvement.

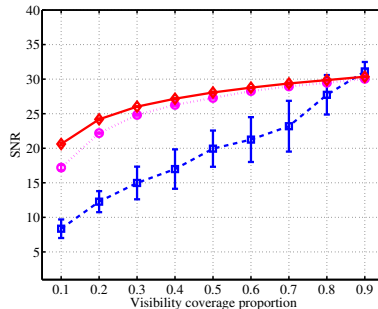


# Spread spectrum effect for varying $w$

## Results on simulations



(a) M31



(b) 30 Dor

Figure: Reconstruction fidelity using SARA.

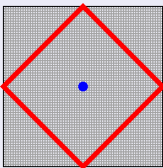
Improvement in reconstruction fidelity due to the spread spectrum effect for varying  $w$  is almost as large as the case of constant maximum  $w$ .

- As expected, for the case where coherence is already optimal, there is little improvement.



## Public codes

## SOPT code

<http://basp-group.github.io/sopt/>*Sparse OPTimisation*

Carrillo, McEwen, Wiaux

SOPT is an open-source code that provides functionality to perform sparse optimisation using state-of-the-art convex optimisation algorithms.

## PURIFY code

<http://basp-group.github.io/purify/>*Next-generation radio interferometric imaging*

Carrillo, McEwen, Wiaux

PURIFY is an open-source code that provides functionality to perform radio interferometric imaging, leveraging recent developments in the field of compressive sensing and convex optimisation.



## Conclusions & outlook

- Effectiveness of compressive sensing for radio interferometric imaging demonstrated.
- Theory of compressive sensing can be used to optimise telescope configuration.
- Exploit state-of-the-art convex optimisation algorithms that support parallelisation.

Apply to observations made by real interferometric telescopes.

Develop fast convex optimisation algorithms that are parallelised and distributed to scale to big-data.



## Conclusions & outlook

- Effectiveness of compressive sensing for radio interferometric imaging demonstrated.
- Theory of compressive sensing can be used to optimise telescope configuration.
- Exploit state-of-the-art convex optimisation algorithms that support parallelisation.

Apply to observations made by real interferometric telescopes.

Develop fast convex optimisation algorithms that are parallelised and distributed to scale to big-data.



## Conclusions & outlook

- Effectiveness of compressive sensing for radio interferometric imaging demonstrated.
- Theory of compressive sensing can be used to optimise telescope configuration.
- Exploit state-of-the-art convex optimisation algorithms that support parallelisation.

Apply to observations made by real interferometric telescopes.

Develop fast convex optimisation algorithms that are parallelised and distributed to scale to big-data.





## Conclusions & outlook

- Effectiveness of compressive sensing for radio interferometric imaging demonstrated.
- Theory of compressive sensing can be used to optimise telescope configuration.
- Exploit state-of-the-art convex optimisation algorithms that support parallelisation.

Apply to observations made by real interferometric telescopes.

Develop fast convex optimisation algorithms that are parallelised and distributed to scale to big-data.

