

Interpolation Rapide Sky to Sky pour l’Imagerie Radio Interférométrique

Nicolas MONNIER¹, François ORIEUX¹, Nicolas GAC¹, Cyril TASSE², Erwan RAFFIN³, David GUIBERT³

¹Laboratoire des Signaux et Systèmes, Univ. Paris-Saclay, CNRS, CentraleSupélec, 91 190 Gif-sur-Yvette, France

²GEPI, Observatoire de Paris, CNRS, Univ. Paris Diderot, 92 190 Meudon, France

³CEPP - Center for Excellence in Performance Programming, Atos Bull, 35 700 Rennes, France

nicolas.monnier@l2s.centralesupelec.fr, orieux@l2s.centralesupelec.fr
nicolas.gac@centralesupelec.fr, cyril.tasse@obspm.fr, erwan.raffin@atos.net

Résumé – La reconstruction d’images radio interférométriques nécessite le traitement de données dans l’espace de Fourier dont les coordonnées non-régulières empêchent l’utilisation de la FFT. La solution la plus commune repose sur des algorithmes d’interpolation, appelée *gridding*, coûteux en calculs, en particulier dans les algorithmes itératifs. Dans cet article, nous proposons une reformulation avantageuse, nommée *Sky to Sky*, permettant de réduire le coût calculatoire du *gridding* et son adjoint, le *degridding*, lorsqu’ils sont utilisés successivement. Nous étudions l’impact de l’interpolation vis-à-vis du coût de calcul et de l’erreur introduite. Finalement, nous utilisons notre algorithme dans une descente de gradient et montrons une accélération entre 1.2 et 16.4 sans approximation supplémentaire par rapport à l’état de l’art.

Abstract – Reconstruction of radio interferometric images requires the processing of data in Fourier space that do not have regular coordinates, preventing use of the Fast Fourier Transform. The most common solution is to rely on interpolation algorithms, called *gridding*, that are computationally expensive. In this paper, we propose an algorithmic reinterpretation, named *sky to sky*, to reduce the computation cost of the *gridding* operation and its adjoint, the *degridding*, when they are used successively. We analyze the impact of interpolation regarding the computation cost and the reconstruction error. We also illustrate this optimization on a full reconstruction with gradient descent and CLEAN algorithm. Finally, we obtain acceleration factors between 1.2 and 16.4 without additional approximation.

1 Introduction

L’imagerie radio interférométrique est une technique de reconstruction d’image du ciel à partir d’une observation radio [1]. Comme les données échantillonnent l’espace de Fourier de manière incomplète, la reconstruction peut être vue comme un problème de synthèse de Fourier connu comme un problème inverse mal posé [2]. Par ailleurs, les générations actuelles et futures de radio télescope, comme le VLA, LOFAR, ou SKA, sont caractérisées par une plus grande résolution que la génération précédente. Ces caractéristiques s’accompagnent d’un très grand nombre de données générées, posant un problème de très grande échelle.

On peut trouver dans la littérature de nombreuses techniques pour reconstruire une image du ciel basées sur le problème direct. La méthode dite CLEAN et ses variantes [3], implémentées dans des pipelines d’imagerie comme DDFacet ou WSClean [4], sont les plus communes. Nous pouvons également trouver des techniques basées sur le *compressed sensing* et sur l’optimisation convexe pour la déconvolution d’image [5].

En radio interférométrie, les mesures sont des coefficients de Fourier, et leurs coordonnées ne sont pas régulières. La transformée de Fourier discrète non-uniforme [6] est connue pour avoir un grand coût calculatoire en $O(n^2)$. De ce fait, une tech-

nique commune est d’interpoler les mesures sur des coordonnées régulières, opération appelée « *gridding* », pour utiliser la FFT qui est en $O(n \log n)$. À cause de la grande quantité de données à traiter, le calcul de ces opérateurs a besoin d’être réduit.

Dans la première partie, nous verrons le problème d’interférométrie radio et les étapes nécessaires de *gridding* et de son adjoint le *degridding*. Ensuite, nous présenterons l’optimisation algorithmique basée sur la succession des opérateurs, définie comme la méthode *Sky to Sky* (S2S). Enfin, nous présenterons des résultats expérimentaux sur le temps et la consommation mémoire de cette méthode, qui seront également illustrés par l’algorithme Cotton-Schwab CLEAN pour estimer un ciel simulé.

2 L’imagerie Interférométrique Radio

2.1 Problème direct

Un interféromètre radio est constitué d’un réseau d’antennes qui génère des mesures issues des émissions radio du ciel observé x . Chaque paire d’antennes est définie avec une ligne de base $b_i = (u, v, w)$, où u , v and w sont les coordonnées en longueur d’onde λ . Les mesures, aussi appelées visibilités,

sont définies par

$$v(u, v, w) = \iint \frac{x(l, m, n)}{n} e^{-2i\pi(ul+vm+w(n-1))} dl dm \quad (1)$$

où $x(l, m, n)$ est la distribution de la luminosité du ciel et l , m , and $n = \sqrt{1 - l^2 - m^2}$ sont les coordonnées en direction cosinus. Si le réseau d'antennes est coplanaire ($w \approx 0$) et le champ de vue est petit ($n \approx 1$), alors chaque visibilité est la transformée de Fourier spatiale 2D de la distribution du ciel. Ce résultat est connu par le théorème de van Cittert-Zernike [1], et Eq. (1) devient

$$v(u, v) = \iint x(l, m) e^{-2i\pi(ul+vm)} dl dm. \quad (2)$$

L'imagerie interférométrique radio a pour but de reconstruire le ciel à partir de la mesure incomplète des visibilités, faisant de ce problème un problème inverse mal posé.

Après discrétisation, le vrai ciel peut être représenté par une grille de $N \times N = P$ pixels tel que $\mathbf{x} \in \mathbb{R}^P$, et les visibilités comme un vecteur $\mathbf{v} \in \mathbb{C}^M$. La relation entre \mathbf{x} et \mathbf{v} est décrite par le problème direct discret

$$\mathbf{v} = \mathbf{S}\mathbf{F}\mathbf{x} + \mathbf{n} \quad (3)$$

où $\mathbf{F} \in \mathbb{C}^{P \times P}$ est la transformée de Fourier discrète, $\mathbf{S} \in \mathbb{C}^{M \times P}$ transforme P coefficients de Fourier en M visibilités, le degridding, et $\mathbf{n} \in \mathbb{C}^M$ du bruit. Enfin, l'image *dirty* $\mathbf{y} \in \mathbb{R}^P$ est définie comme la projection arrière des visibilités, où l'application de l'opérateur adjoint $\mathbf{y} = \mathbf{F}^\dagger \mathbf{S}^\dagger \mathbf{v}$, où \mathbf{F}^\dagger est la transformée de Fourier inverse, et $\mathbf{S}^\dagger \in \mathbb{C}^{P \times M}$ est l'opérateur de gridding, adjoint de \mathbf{S} .

2.2 Opérateur de gridding

La coordonnée continue $\mathbf{b}_i = (u_i, v_i)$ associée à la visibilité v_i est approximée à la position au plus proche voisin d'une grille fine aux pas (Δ_u, Δ_v) , tel que la nouvelle coordonnée est

$$\tilde{\mathbf{b}}_i = (n_i \Delta_u, m_i \Delta_v). \quad (4)$$

Si K est le facteur de suréchantillonnage, les pas FFT de la grille \mathbf{g} sont $(K\Delta_u, K\Delta_v)$. Dans ce cas, l'opérateur de gridding \mathbf{S}^\dagger , associé à la coordonnée \mathbf{b}_{kl} de la grille FFT, est défini par

$$\mathbf{g}(\mathbf{b}_{kl}) = \sum_{i=1}^M C^\dagger (n_k K \Delta_u - n_i \Delta_u, m_l K \Delta_v - m_i \Delta_v) v_i \quad (5)$$

où C^\dagger est un noyau. L'opération (5) n'est pas une convolution, car non stationnaire. Le choix et la taille de ce noyau pour sont des sujets bien connus dans la littérature [7]. Dans notre cas, nous avons utilisé une fonction de Kaiser-Bessel.

2.3 Estimation du ciel

La reconstruction du ciel \mathbf{x} à partir de données incomplètes est un problème inverse mal posé. Par conséquent, l'estimation du ciel à reconstruire en imagerie interférométrique est généralement définie par

$$\hat{\mathbf{x}} = \arg \min_x \|\mathbf{v} - \mathbf{S}\mathbf{F}\mathbf{x}\|^2 + R(\mathbf{x}) \quad (6)$$

où $\|\mathbf{v} - \mathbf{S}\mathbf{F}\mathbf{x}\|^2$ est le terme de fidélité aux données et R un régularisateur [2]. Les algorithmes utilisés pour résoudre ce problème d'optimisation exigent de calculer de nombreux gradients du critère. Le coût en calcul du gradient du terme de fidélité de données

$$\nabla J(\mathbf{x}) = \mathbf{F}^\dagger \mathbf{S}^\dagger (\mathbf{v} - \mathbf{S}\mathbf{F}\mathbf{x}) \quad (7)$$

peut être très grand, car il nécessite le traitement des opérateurs de gridding et degridding. Nous proposons dans ce papier une nouvelle formulation pour réduire le coût en calcul d'une évaluation du gradient sans ajouter d'approximations.

3 Méthode Sky to Sky

3.1 Décomposition du gridding et degridding

L'opérateur de gridding \mathbf{S}^\dagger peut être décomposé en trois étapes : accumulation, convolution et sous-échantillonnage.

1. Accumulation \mathbf{A}^\dagger .

On définit \mathbf{g}' une grille fine de résolution (Δ_u, Δ_v) , de taille $KN \times KN = K^2P$, contenant les visibilités accumulées ayant des positions approchées identiques

$$\mathbf{g}'(\mathbf{b}_p) = \begin{cases} \sum_i v_i & \text{si } \tilde{\mathbf{b}}_i = \mathbf{b}_p, \\ 0 & \text{sinon,} \end{cases} \quad (8)$$

L'opérateur d'accumulation est donc défini par $\mathbf{A}^\dagger \in \{0, 1\}^{K^2P \times M}$ avec un 1 par colonne.

2. Convolution \mathbf{C}^\dagger .

Ensuite cette grille fine de visibilités accumulées est convoluée avec le noyau \mathbf{C}^\dagger

$$\tilde{\mathbf{g}}(\mathbf{b}_p) = \sum_i \sum_j C^\dagger(\mathbf{b}_p - \mathbf{b}_{ij}) \mathbf{g}'(\mathbf{b}_{ij}). \quad (9)$$

3. Sous-échantillonnage \mathbf{O}^\dagger .

La grille FFT est plus grossière d'un facteur K par rapport à la grille sur-échantillonnée. L'opérateur de sous-échantillonnage \mathbf{O}^\dagger est alors une matrice $P \times K^2P$, remplie d'un 1 par ligne et de 0 sinon.

Par conséquent, l'étape de gridding, illustrée par Fig. 1, peut être écrite comme

$$\mathbf{S}^\dagger = \mathbf{O}^\dagger \mathbf{C}^\dagger \mathbf{A}^\dagger. \quad (10)$$

L'opérateur de degridding \mathbf{S} est l'adjoint du gridding \mathbf{S}^\dagger . Ainsi, à partir de Eq. (10), nous pouvons définir la décomposition du degridding comme

$$\mathbf{S} = \mathbf{A}\mathbf{C}\mathbf{O} \quad (11)$$

où \mathbf{O} est un opérateur de sur-échantillonnage, \mathbf{C} est une convolution et \mathbf{A} est un opérateur d'échantillonnage. Cette décomposition est illustrée par Fig. 2.

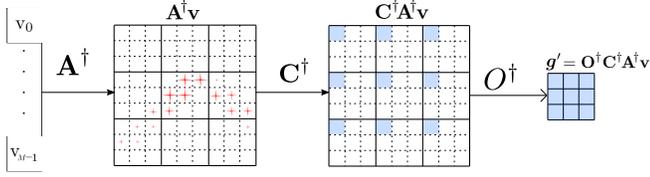


FIGURE 1 – Décomposition du gridding

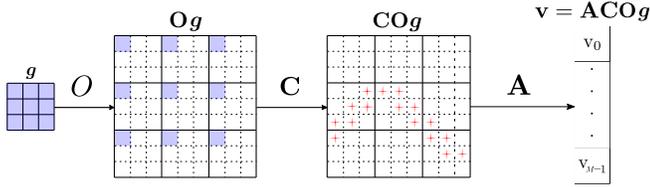


FIGURE 2 – Décomposition du degrading

3.2 Optimisation du calcul du gradient

La méthode standard pour calculer l'opérateur de gradient décrit par Eq. (7) peut être développée de telle sorte qu'elle devienne

$$\begin{aligned} \nabla J(x) &= \mathbf{F}^\dagger \mathbf{S}^\dagger \mathbf{v} - \mathbf{F}^\dagger \mathbf{S}^\dagger \mathbf{S} \mathbf{F} x \\ &= \mathbf{y} - \mathbf{F}^\dagger \mathbf{S}^\dagger \mathbf{S} \mathbf{F} x. \end{aligned} \quad (12)$$

Cette décomposition, mentionnée par la Fast Holographic Deconvolution (FHD) [8], sera appelée la méthode Sky to Sky (S2S). On peut noter que $\mathbf{F}^\dagger \mathbf{S}^\dagger \mathbf{v}$ est l'image *dirty* \mathbf{y} , qui n'est calculée qu'une seule fois. La succession des opérateurs devient

$$\mathbf{S}^\dagger \mathbf{S} = \mathbf{O}^\dagger \mathbf{C}^\dagger \mathbf{A}^\dagger \mathbf{A} \mathbf{C} \mathbf{O} \quad (13)$$

avec $\mathbf{A}^\dagger \mathbf{A} = \mathbf{A}^*$, et $\mathbf{A}^* \in \mathbb{N}^{K^2 P \times K^2 P}$ une matrice diagonale, tel que $\mathbf{a} = \text{diag}(\mathbf{A}^*)$. La valeur de chaque élément de \mathbf{a} correspond au nombre de visibilité ayant leurs coordonnées approchées à la même position. Cette décomposition est illustrée par Fig. 3.

En posant M' le nombre d'éléments non nuls dans \mathbf{a} , les cas pratiques montrent que $M \gg M'$ en fonction du facteur de sur-échantillonnage K . Par conséquent, la méthode S2S évite de traiter le M -vecteur des visibilité brutes, mais un M' -vecteur plus petit. De plus, contrairement à FHD [8] qui construit et stocke $\mathbf{S}^\dagger \mathbf{S} \in \mathbb{C}^{P \times P}$ sous forme de matrice, nous n'avons besoin que de construire le vecteur *sparse* \mathbf{a} . Ainsi, pour la coordonnée \mathbf{b}_k , l'équation $\mathbf{g}' = \mathbf{S}^\dagger \mathbf{S} \mathbf{g}$ s'écrit

$$\mathbf{g}'(\mathbf{b}_k) = \sum_{i'=1}^{M'} \mathbf{C}^\dagger \left(\mathbf{b}_k - \mathbf{b}_{i'} \right) a_{i'} \sum_{i,j} \mathbf{C} \left(\mathbf{b}_{i'} - \mathbf{b}_{ij} \right) \mathbf{g}(\mathbf{b}_{ij}). \quad (14)$$

Pour un même facteur de sur-échantillonnage, le gradient calculé par la méthode S2S Eq. (12) et la méthode STD Eq. (7) est le même. De plus, le coût de calcul du gradient par la méthode S2S est donné par

$$O(2M' C_{supp}^2 + M' + 2P \log P + P) \quad (15)$$

alors que le coût de calcul de l'approche standard est de

$$O(2M C_{supp}^2 + 2P \log P + M). \quad (16)$$

Par conséquent, pour le même K , la méthode présentée est plus efficace et sans erreurs supplémentaires lorsque

$$M' < \frac{2M C_{supp}^2 + M - P}{2C_{supp}^2 + 1}. \quad (17)$$

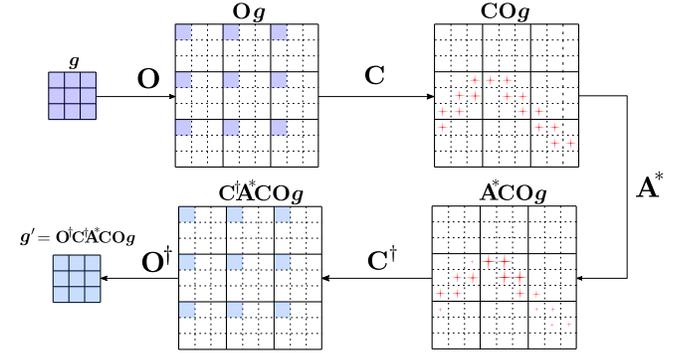


FIGURE 3 – Décomposition du degrading-gridding

4 Expérimentations et Résultats

Nous montrons l'avantage d'utiliser la méthode S2S pour améliorer la vitesse d'une évaluation de gradient et l'impact sur l'utilisation de la mémoire en fonction du facteur de sur-échantillonnage K . Contrairement à FHD [8], nous n'avons pas besoin de stocker la matrice.

Les visibilité simulées ont été générées en suivant Eq. (2) pour une observation de 8 heures et 64 canaux de fréquence du VLA-D avec un temps de 1 seconde entre deux échantillons et une seule polarisation. Toutes les expériences ont été réalisées sur un système BullSequana X400 basé sur un processeur Intel(R) Xeon(R) Platinum 8358, exécutant une première implémentation d'un code parallèle Cython.

4.1 Résultats du calcul de gradient

4.1.1 Erreur liée au facteur de sur-échantillonnage

Les visibilité approximées des coordonnées $\tilde{\mathbf{b}}_i$ dépendent de K . Comme le montre [9], un facteur de sur-échantillonnage faible implique une erreur plus élevée. Les Fig. 4 et 5 montrent l'erreur maximale

$$\text{MaxAE}(K) = \max(|\nabla(\mathbf{x}_{K=63}) - \nabla(\mathbf{x}_{K=j})|) \quad (18)$$

et l'erreur absolue moyenne entre les gradients

$$\text{MAE}(K) = \frac{1}{P} \sum_{i=0}^P |\nabla(\mathbf{x}_{i,K=63}) - \nabla(\mathbf{x}_{i,K=j})|. \quad (19)$$

Comme prévu, l'erreur augmente lorsque le facteur de sur-échantillonnage diminue. De plus, il est courant d'utiliser une technique de moyennage pour réduire la quantité de données à traiter [1] entraînant une perte d'information et des artefacts. Comme

prévu, l'erreur est plus élevée, mais se rapproche de la courbe des données brutes pour de faibles facteurs de sur-échantillonnage.

4.1.2 Coût en temps de calcul

Le temps de calcul pour les données brutes et moyennées est indiqué dans le tableau 1. Pour tout facteur de sur-échantillonnage et tout ensemble de données, la méthode proposée produit le même résultat que la méthode STD avec un temps de calcul plus rapide. Pour $K = 5$, la méthode S2S calcule $\nabla J(\mathbf{x})$ en utilisant *des données brutes sans moyennage* mais avec un facteur d'accélération de 16.3.

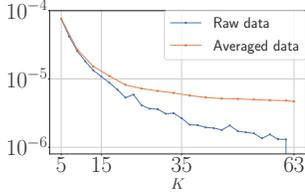


FIGURE 4 – MAE.

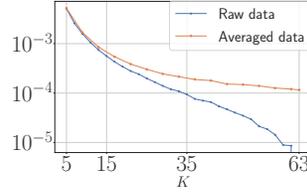


FIGURE 5 – MaxAE.

K	STD	S2S	STD moy.	S2S moy.
5	242.9	14.8 ($\times 16.4$)	33.7	13.9 ($\times 2.4$)
15	242.9	24.1 ($\times 10.0$)	33.7	22.1 ($\times 1.6$)
35	242.9	43.9 ($\times 5.6$)	33.7	27.0 ($\times 1.3$)
63	242.9	71.8 ($\times 3.4$)	33.7	28.5 ($\times 1.2$)

TABLE 1 – Temps en secondes pour évaluer le gradient.

4.1.3 Coût en mémoire

La méthode STD nécessite de garder en mémoire les visibilités. La méthode S2S doit garder en mémoire les images *dirty* et le vecteur \mathbf{a} . Comme $M \gg P$, la méthode proposée permet un coût mémoire plus faible, voir Tab 2. Le coût mémoire de la méthode proposée est légèrement supérieur à celui de la méthode STD avec moyennage qui introduit des erreurs.

K	STD	S2S	STD moy.	S2S moy.
5	1.94	0.46	0.20	0.45
15	1.94	0.59	0.20	0.55
35	1.94	0.85	0.20	0.63
63	1.94	1.20	0.20	0.65

TABLE 2 – Coût mémoire pour évaluer le gradient en GB

4.2 Résultats de la méthode d'estimation du ciel

L'estimation du ciel simulé, composé de 50 sources ponctuelles d'amplitude entre 2×10^{-2} et 1, a été réalisée à l'aide de l'algorithme CLEAN [10] de Cotton-Schwab. Cet algorithme itératif est divisé en deux parties. La première partie, appelée cycle majeur, calcule le gradient du terme de fidélité des données $\nabla J(\mathbf{x}^{(k)})$. La seconde partie, appelée cycle mineur, approche le problème comme une déconvolution. La Fig. 6(a) montre un zoom de l'image *dirty*. Nous avons effectué deux cycles majeurs et 2500 cycles mineurs. Toutes les sources ponctuelles

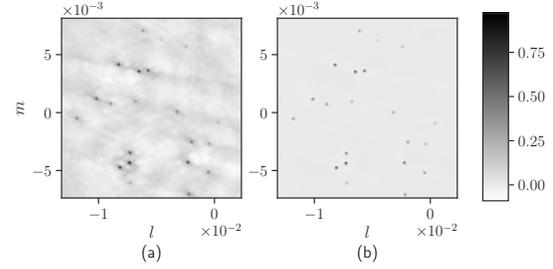


FIGURE 6 – Zoom : (a) *Dirty* image (b) Reconstruction.

initiales sont identifiées dans l'image restaurée comme illustré Fig. 6(b).

Le tableau. 3 montre le temps d'exécution de l'algorithme pour chaque cas. Encore une fois, nous obtenons des facteurs d'accélération avec une accélération maximale de 3.44.

K	STD	S2S	STD moy.	S2S moy.
5	641.2	186.8 ($\times 3.4$)	117.3	77.8 ($\times 1.5$)
15	641.2	201.1 ($\times 3.2$)	117.3	94.0 ($\times 1.2$)
35	641.2	240.6 ($\times 2.7$)	117.3	103.8 ($\times 1.1$)
63	641.2	296.4 ($\times 2.1$)	117.3	106.8 ($\times 1.1$)

TABLE 3 – Temps total en secondes pour CLEAN

5 Conclusion

Nous avons montré une décomposition des opérateurs de gridding et de degriding utilisés dans les problèmes d'imagerie par radio interférométrie. Nous avons proposé une méthode de Sky to Sky (S2S) pour réduire le coût de calcul sans erreur supplémentaire. La méthode S2S permet de réduire le coût mémoire tout en ayant un facteur d'accélération jusqu'à 16.4 par rapport à la méthode standard. La reconstruction d'un ciel en condition réaliste a montré un facteur d'accélération jusqu'à 3.4. Enfin, nos expériences ont montré que la méthode S2S est toujours plus efficace que la méthode standard pour les mêmes paramètres de grille.

Références

- [1] A. R. Thompson, J. M. Moran, and G. W. Swenson, *Interferometry and synthesis in radio astronomy*, 2nd ed. Wiley, 2001.
- [2] E. Thiébaud and al., "Principles of image reconstruction in optical interferometry : tutorial," *Journal of the Optical Society of America A*, vol. 34, no. 6, 2017.
- [3] J. Högbom, "Aperture synthesis with a non-regular distribution of interferometer baselines," *Astronomy and Astrophysics*, vol. Vol. 15, 1974.
- [4] C. Tasse and al., "Faceting for direction-dependent spectral deconvolution," *Astronomy and Astrophysics*, vol. 611, p. A87, 2018.
- [5] R. Carrillo and al., "Purify : a new approach to radio-interferometric imaging," *Monthly Notices of the Royal Astronomical Society*, vol. 439, no. 4, 2014.
- [6] S. a. Bagchi, "The nonuniform discrete fourier transform and its applications in filter design," *IEEE Transactions on Circuits and Systems*, vol. 43, no. 6, 1996.
- [7] J. A. Fessler, "On NUFFT-based gridding for non-cartesian MRI," *Journal of Magnetic Resonance*, vol. 188, no. 2, 2007.
- [8] I. Sullivan and al., "Fast holographic deconvolution : a new technique for precision radio interferometry," *The Astrophysical Journal*, vol. 759, no. 1, 2012.
- [9] A. R. Offringa and al., "Precision requirements for interferometric gridding in the analysis of a 21 cm power spectrum," *Astronomy & Astrophysics*, vol. 631, 2019.
- [10] F. R. Schwab, "Relaxing the isoplanatism assumption in self-calibration : applications to low-frequency radio interferometry," *The Astronomical Journal*, 1984.