Draft version September 12, 2024
Typeset using LaTeX **twocolumn** style in AASTeX631

Post-Processing High-Contrast Images to Detect and Characterize Debris Disks

Emma Davis,[1,2] Maxwell Millar-Blanchaer,[2] and Connor Vancil[2]

[1] *Departments of Physics and Astronomy, Smith College*
[2] *Department of Physics, University of California Santa Barbara*

ABSTRACT

Astronomers interested in directly observing debris disks face the challenge of imaging faint objects near much brighter objects. High-contrast imaging recovers debris disk, exoplanet, and other faint signals through a range of instrumental, observational, and post-processing techniques. State-of-the-art post-processing algorithms provide better visualization and characterization of debris disks by removing background noise and starlight, or the Point Spread Function (PSF). We utilize the method of Non-Negative Matrix Factorization (NMF) — with `JAX NumPy` expediting the process on a GPU — to perform image post-processing on the debris disks detected in Esposito et al. (2020). Employing JAX on a GPU resolves some of the more computationally expensive portions of the NMF algorithm, allowing a more feasible time frame for this process. Paired with data imputation (Ren et al. 2020), user-defined masks for each disk further supplements NMF. The binary data-imputation masks reduce the amount of disk light subtracted in post-processing by treating the disk light as missing data and thus ignoring the disks in PSF construction. The results demonstrate less over- and self-subtraction in the regions of the disks closest to the respective host stars. Additionally, by integrating these tools in pyKLIP (Ren et al. 2018), we ensure that all pyKLIP users will have access to the techniques outlined in this paper, promoting greater access to modern techniques and knowledge of debris disks and their possible accompanying exoplanets.

Keywords: Debris disks – Non-Negative Matrix Factorization – Data imputation – High-Contrast Imaging – Image post-processing – Direct Imaging

1. INTRODUCTION

Beyond the Solar System, debris disks act as large–scale versions of our Solar System's asteroid and Kuiper belts, similarly composed of dust, gas, and minor bodies. Collisions break down the bodies within the disks through a collisional cascade, producing more dust and gas that, upon reaching blowout size, are no longer contained in the disk by gravitational forces and are instead blown out by radiative and wind pressures (Hughes et al. (2018)). A debris disk's collisional cascade eventually reaches a steady state with more small bodies than large ones, producing a dominance of dust spectra (Pearce (2024)). The host stars of extrasolar systems heat the dust particles of their respective debris disks, producing blackbody radiation and leading to an excess flux from infrared wavelengths. Debris disk discoveries started from the determination of infrared excess, but advancements in observational techniques and technologies allow for the direct imaging of debris disks.

In direct imaging, the light scattered by the dust allows for a better morphological study of debris disks, such as in position angles, geometry, and, importantly for exoplanet detection, rings (Hughes et al. (2018); Pearce (2024)). Rings may represent areas where exoplanets influence the orbits of the smaller bodies within the disks, creating gaps in the disks where the exoplanets may reside. Exoplanet discovery through debris disk inference probes different regions of parameter-space for exoplanets, as seen in Figure 1.

1.1. *Direct Imaging*

This powerful tool for determining the existence and location of exoplanets is limited by the precision of the direct images. Disks and other faint objects require high-contrast imaging (HCI), a technique in direct imaging that separates a faint source from a bright one — here, the faint disk from the bright host star — with the ratio of their luminosities being the contrast ratio (Follette (2023)). While HCI encompasses a range of technologies such as adaptive optics, wavefront sensing,

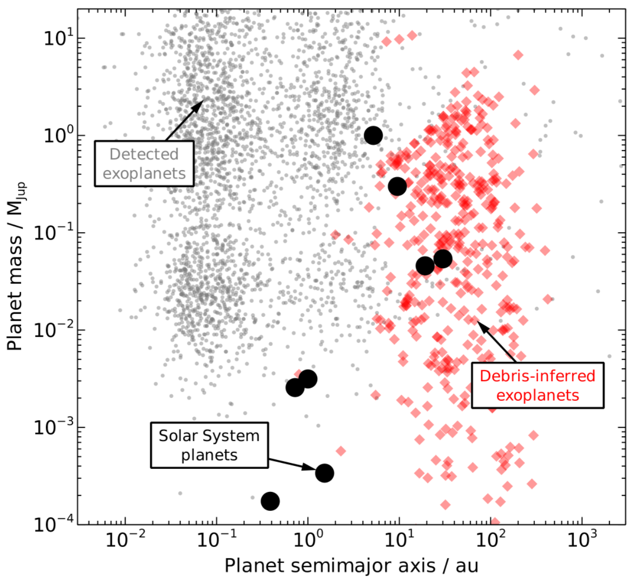

Figure 1. Reproduced from Pearce (2024), this represents exoplanets inferred from debris disks versus identified by exoplanet detection methods. The data points represent the exoplanet's distance from its host star versus its mass compared to the mass of Jupiter. The planets from the Solar System are in black. This plot demonstrates the exoplanet detection biases in parameter space for closer-in, more massive stars, which are not completely representative of planets in the Solar System, so that we know that current methods does not detect all exoplanet types. Debris-inferred exoplanets cover a different region of parameter-space, and which consists of some planets similar to Solar System planets.

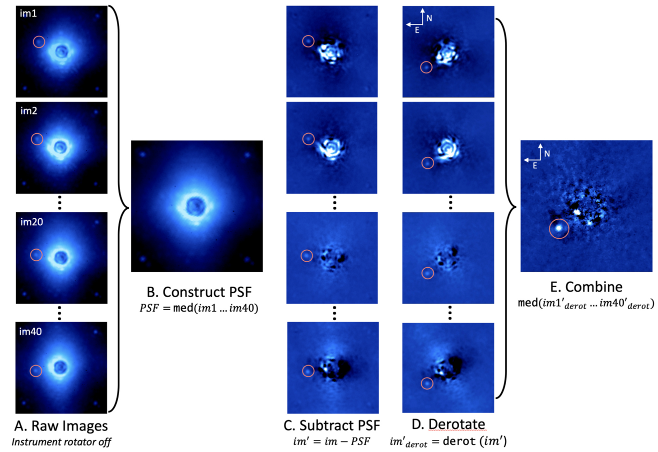

Figure 2. This figure represents step-by-step Angular Differential Imaging, reproduced from Follette (2023). First, take raw images with the instrument rotator off. Then, construct the PSF by median combining the raw images. Subtract the PSF from the raw data, then de-rotate the subtracted images so that the exoplanet or debris disk is at the same place in each image. Median combine the subtracted, derotated data for the final image.

and coronagraphy, the image produced by these methods necessitates further high contrast image processing through differential imaging.

1.2. *Image Post-Processing*

Differential imaging describes the process of removing a reference image — the distribution of light from the source star referred to as the point spread function (PSF) — and noise from a target image. The methods of differential imaging include reference differential imaging (RDI), angular differential imaging (ADI, Figure 2), spectral differential imaging (SDI), and more. ADI and RDI either assemble and combine (RDI) or construct (ADI) a PSF to subtract from the target image, and then each images is derotated and combined into a final image. SDI involves similar techniques but across different wavelengths of the image spectra with an integral field spectrograph image cube. Differential imaging techniques are typically combined with other post-processing techniques, allowing for more precise images and analysis.

While post-processing the PSFs, individuals utilize algorithms to construct more descriptive PSFs, most commonly with the Locally Optimized Combination of Images (LOCI) algorithm and the Karhunen–Loève Image Processing (KLIP) algorithm (Soummer et al. 2012) which applies principal component analysis. With KLIP, images are converted into one-dimensional vectors and cross-correlated, and the common patterns, or KL modes, are used to construct the reference images. Both KLIP and LOCI overfit and self–subtract the data, and KLIP additionally requires forward modeling, which assumes disk morphology, making these methods non ideal — especially for disks with irregular morphologies. However, the method of Non-Negative Matrix Factorization (NMF) produces promising results in these regards, as demonstrated by Ren et al. (2018).

1.3. *Non-Negative Matrix Factorization*

The fundamental procedure of NMF is to decompose one matrix into a product of two non-negative matrices. In direct imaging, NMF decomposes a reference matrix with dimensions of the number of references by the number of pixels per image into a coefficient matrix and component one. The rows of the component matrix are the NMF components which represent part of the signal. The iterative approach of NMF starts with the construction of a component basis, similar to KLIP but non-negative, which approximates the reference matrix. NMF iterates over a number of components from which to compose the component basis and minimizes the Euclidean distances of the matrices which approximate the reference matrix. It then constructs a flattened target model with the previously constructed components and rescales the model due to disk contributions. With the model of the target constructed from the references, NMF then subtracts the model from the original image, or the residuals, returning a PSF subtracted target image, so for the purposes of this project an image of the disk.

1.3.1. *Data Imputation*

NMF, unlike KLIP, may also be run with data missing from the input images, which allows for the use of disk imputation as in Ren et al. (2020). In constructing the components of the references, over-subtraction occurs where disk light appears in these components and therefore the PSF model, so that the subtraction of the PSF model subtracts some of the disk. This is illustrates in the leftmost column of Figure 3. With data imputation, binary masks model the disks as zeros, flagging the light of the disk as missing data and therefore constructing the components for the model without subtracting out any of the disk.

2. METHODS

We process disks images with the NMF algorithm through the open-source Python library pyKLIP. By speeding up the process on a GPU and implementing data imputation, we improve upon previous methods of NMF. Finally, we apply this data imputation with JAX for NMF to Gemini Planet Imager (GPI) data, with this process available to all pyKLIP users.

2.1. *Debris Disk Data*

The disks evaluated here come from the Gemini Planet Imager Exoplanet Survey (GPIES), a survey of stars with suspected massive planets or debris disks in scattered light. This survey was taken by the Gemini Observatory from November 2014 to December 2018, and was previously compiled and processed in Esposito et al. (2020). We analyze 16 of these disks, with best results assumed for disks that are edge-on (elliptical inclinations approaching 90°) and with larger parallactic angle rotation. The disks are listed in Table 1, along with their on-sky parallactic angles, inclination, and total change in parallactic angles throughout the data collection.

Name	Inclination(deg)	PA (deg)	Δ PA (deg)
HD 30447	83.0	212.0	125.8
HD 32297	88.4	47.9	19.1
HD 106909	84.6	284.2	7.1, 20.3*
HD 110058	84.0	155.0	25.2
HD 111161	62.1	83.2	38.0
HD 111520	88.0	165.0	28.3
HD 114082	83.3	105.7	12.3
HD 115600	80.0	27.5	24.0
HD 117214	71	179.8	18.5, 19.8*
HD 129590	75.7	121.7	17.9
HD 131835	75.1	61.4	74.2
HD 143675	87.2	113.2	20.5, 94.3*
HD 145560	43.9	41.5	17.5, 36.0*
HD 146897	84.0	113.9	29.5
HD 157587	70.0	127.0	49.9
HD 191089	59.0	70.0	101.3

Table 1. Data from Esposito et al. (2020).
This table gives the name, inclination (in degrees), and parallactic angle (in degrees) of the disks, as well as the total rotation of the datasets, Δ PA. In some cases, more than one datasets were used, with the second representing the spectral auto-reduced datasets.
*Indicates multiple datasets

2.2. *NMF with JAX NumPy*

Previously, our group had paired NMF with `JAX NumPy` (UCSB-Exoplanet-Polarimetry-Lab 2023) based on code originally developed by Guangtun Ben Zhu and adapted by Bin Ren. In summary, this code uses `JAX NumPy`'s Just-In-Time (JIT) compiler to run NMF on a graphics processing unit (GPU), which creates parallelization and therefore faster computing time.

In this project, we integrated this NMF JAX into pyKLIP, a Python library which performs PSF subtraction through NMF, KLIP, or other PCA algorithms and which can be run with ADI, SDI, RDI, or ADI+SDI. The inclusion of NMF JAX in pyKLIP follows pyKLIP's own handling of NMF, with the same forms of inputs and outputs. Additionally, NMF JAX is integrated such that JAX installation is only required for the running of NMF JAX and not for users only interested in KLIP or traditional NMF. NMF JAX becomes available as a parameter for the input `algo` in the function `klip dataset`, which also dictates the running of the other algorithm types. We also expanded NMF and NMF JAX in pyKLIP to run with a list of component numbers, as KLIP does with KL modes. This streamlines the process of testing numbers of components, since rather than computing each component individually a list of components produces an image cube of FITS files. A user with a GPU may now run NMF JAX within pyKLIP through the same methods of running NMF or KLIP.

2.3. *Masking Disks*

Data imputation for the disks requires an assembly of masks which the user assembles themselves. We model our disks as simple ellipses, with parameters of image center, size, disk inclination, rotation, and radius, or better thought of as the full extent to cover the disk. The parameters come from Esposito et al. (2020) and are listed in Table 1, yet we hand-tune the radial parameters to reduce the appearance of the mask and increase the disk light.

The disks are rotated onto the on-sky images and then associated with the dataset as a GPI class variable, which the user sets alongside running `GPIData` in pyKLIP. In parallelized.py — the notebook within pyKLIP required to run NMF JAX as well as other algorithms — the masks are saved to shared memory and processed alongside the references and target images, so that there is a target or science mask and a list of reference masks for each image in the dataset for ADI. Currently, RDI does not support masking techniques in pyKLIP.

The primary efforts of this project were to demonstrate the effectiveness of data imputed NMF JAX (DI-NMF JAX) as well as successfully implement the masks in NMF JAX without altering the running of any other pyKLIP processes. While the existing code suppesedly accepted mask imputs, it improperly handled the masks, often creating optical appearances of two disks, and produced size-based errors, which was resolved here. NMF JAX with data imputation was tested within pyKLIP's functions as well as with manual ADI PSF subtraction, which demonstrated preliminary success with the disk HD 111520. Moving forward, we apply this method to the 16 aforementioned disks.

3. RESULTS

Some preliminary successful results of DI-NMF JAX disks are listed in Figure 3. We compare the disks in KLIP versus DI-NMF, with the on-sky binary mask included for reference. With our testing of all 16 disks, we found that DI-NMF JAX possibly proved better than KLIP in some instances. We additionally improved the NMF code and added further options for image post-processing in pyKLIP.

3.1. *Code Improvement*

We have improved NMF in pyKLIP through the implementation of JAX, so that users interested in running NMF without masks would also see faster computational times, given access to a GPU. In addition to JAX, NMF now can be run with a list of component numbers with which to construct the images, instead of requiring NMF to run multiple times for a certain number of components. Since NMF uses previous components to construct the next one iteratively, this streamlined code improvement also utilizes the nature of NMF more efficiently, reducing computation time.

Disk	KL Modes	NMF Components	Dataset Length
HD 111520	30	30	31
HD 110058	19	19	20
HD 32297	35	37	38
HD 143675	20	51	52
HD 106909	15	42	43

Table 2. This table gives the number of KL modes and NMF components used in the results image of Figure 3.

3.2. *DI-NMF JAX*

The use of data imputation demonstrates less disk light self-subtraction and less of the over-subtraction previously noted in KLIP. The negative regions around the disk, evident with KLIP, are significantly reduced with DI-NMF JAX, seen in the right column of Figure 3. For successful implementations of DI-NMF JAX, like those examples in Figure 3, more disk light is apparent closer to the host star, and possible additional features such as the back side of the disk become more apparent — possibly in HD 111520. An area to improve is the residual PSF light towards the center, though may be an additional consequence of the positivity of NMF versus the negative weights that can be applied in KLIP. However, weighted averages discussed in section 4 may provide better results in this area.

In general, DI-NMF JAX tends to catch the light in the over-subtracted regions of KLIP, refining our imaging of debris disk morphology. In HD 111520, the inner regions of the disk appear more strongly, and the DI-NMF result may hint towards more structure to be further investigated. HD 110058 appears brighter in the inner regions, despite leftover PSF residuals. In HD 32297, the regions of self-subtraction towards the edges are reduced, and the regions closest to the star are stronger. For HD 106906, we see the strongest promise for DI-NMF JAX. The farthest regions of the disk are more prominent in regions that were previously subtracted. The presence of this feature may suggest that the disk is more inclined that previously considered. Some of these results were expected, since the purpose of the masks is to limit the disk contribution to the PSF. The number of KL modes and NMF components for Figure 3 are listed in Table 2.

The least successful cases, as previously mentioned, were the face-on disks with less rotation. Results demonstrated strong presence of the mask in the final results.

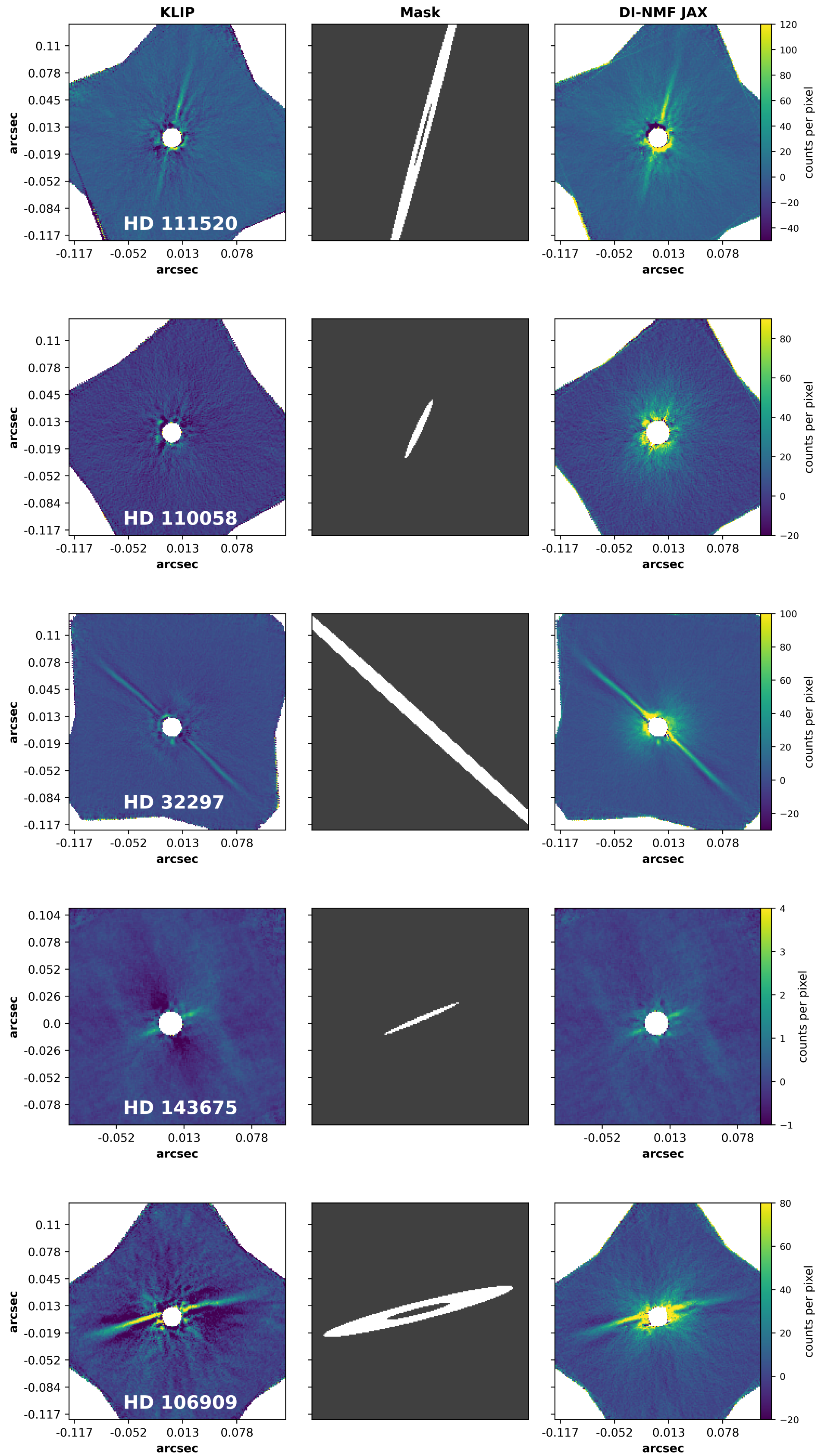

Figure 3. KLIP vs Masks vs Data Imputed NMF JAX for 5 out of 16 Disks.
The pixel value of the images are in counts per pixel, and the masks are binary, with the disk being 0 and the background being 1. The spatial coordinates are 3.6E-7 degrees per coordinate.

For those cases, KLIP is still the better choice over DI-NMF JAX. For the successful cases, the edge-on and highly rotated disks, DI-NMF JAX proves to be a useful post-processing algorithm.

4. FUTURE WORK

This procedure requires some trial and error for mask placement, so a larger range of masks that vary in size could be applied to these disks. Also, to eliminate some of the residual PSF starlight from the images, one could apply weights to the outputs based on the standard deviation of the output, and creating a weighted average image that is inversely proportional to that standard deviation. One could also use the signal-to-noise ratio as a weight, where the average value in the disk annulus is divided by the standard deviation of the rest of the image. While these are all direct ways to improve the algorithm, future work in automating this process would also make this algorithm more accessible and versatile.

5. CONCLUSION

Direct imaging of debris disks and other spread out sources produce problems in the differential imaging process, whereby the disk may appear in the same pixel(s) in several frames, causing some of the disk to be added to the point spread function of the starlight and noise. When the disk light gets incorporated into the PSF, parts of the disk are subtracted out during PSF subtraction. By implementing masks of the disks in PSF construction, disk light is ignored and thus left in the final image. The mask data imputation method is a special feature of Non-Negative Matrix Factorization, an algorithm for approximating a reference image to subtract out of the data so that the final image shows the target, in this case a debris disk.

NMF JAX with data imputation improves upon self-subtraction in these disks and provides an improvement in resolving disk features in some cases, especially for edge-on disks with significant rotation. Overall, we successfully implement data imputation for NMF with JAX in the Python library pyKLIP, and we provide a tutorial for this procedure. The data imputation consists of creating a binary mask which assigns zeros to the regions where the disk resides and ones where it does not, so the light from the disk does not contribute to the PSF construction and subtraction. We minimize over- and self-subtraction due to the positive nature of NMF, and use the data imputation to compensate for any disk scattered light lost in the NMF method.

Some of the most successful disks, as seen in the right column of Figure 3, provide more disk light than their KLIP counterparts. This result demonstrates promise for DI-NMF JAX, especially in determining the morphology of the disks, as we can see more of the disk. While more work must be done to optimize the masks chosen per disk and reduce leftover PSF residuals, DI-NMF JAX improves upon certain established problems in KLIP and may provide a better post-processing method for astrophysical studies of debris disks.

ACKNOWLEDGEMENTS

I firstly thank my primary investigator and group advisor, Dr. Millar-Blanchaer, for the expert guidance, encouragement, and consistent advice he provided to this research as well as for my own academic and career development. Dr. Blanchaer far exceeded the responsibilities of a PI, going above and beyond for the research, his group members, and in fostering a good research environment. I additionally thank my graduate student advisor, Connor Vancil, for the assistance and accompaniment in the research, as well for exemplifying the best values of graduate student research and life. I also thank every member of the Millar-Blanchaer lab, who demonstrated the values of collaboration and support in a research group and who constantly improved my days in the lab. Importantly, I would like to thank Dr. Sathya Guruswamy for providing this research opportunity with her management of the UCSB Physics REU program and for her continual investment in the futures of every member of the 2024 Physics REU cohort. Finally, I thank the National Science Foundation for their financial and ideological support through the REU grant PHY-2349677.

REFERENCES

Esposito, T. M., Kalas, P., Fitzgerald, M. P., et al. 2020, The Astronomical Journal, 160, 24

Follette, K. B. 2023, Publications of the Astronomical Society of the Pacific, 135, 093001

Hughes, A. M., Duchêne, G., & Matthews, B. C. 2018, Annual Review of Astronomy and Astrophysics, 56, 541

Pearce, T. D. 2024, arXiv preprint arXiv:2403.11804

Ren, B., Pueyo, L., Chen, C., et al. 2020, The Astrophysical Journal, 892, 74

Ren, B., Pueyo, L., Zhu, G. B., Debes, J., & Duchêne, G. 2018, The Astrophysical Journal, 852, 104

Soummer, R., Pueyo, L., & Larkin, J. 2012, The Astrophysical Journal Letters, 755, L28

UCSB-Exoplanet-Polarimetry-Lab. 2023, NMF imagingJAX, https://github.com/UCSB-Exoplanet-Polarimetry-Lab/NMF_imagingJAX, GitHub