





PACMAN: A pipeline to reduce and analyze Hubble Wide Field Camera 3 IR Grism data

Sebastian Zieba ^{1,2} and Laura Kreidberg ¹

¹ Max-Planck-Institut für Astronomie, Königstuhl 17, D-69117 Heidelberg, Germany ² Leiden Observatory, Leiden University, Niels Bohrweg 2, 2333CA Leiden, The Netherlands

DOI: [10.21105/joss.04838](https://doi.org/10.21105/joss.04838)

Software

- [Review](#) 
- [Repository](#) 
- [Archive](#) 

Editor: [Juanjo Bazán](#)  

Reviewers:

- [@astrobel](#)
- [@aureliocarnero](#)

Submitted: 25 July 2022

Published: 21 December 2022

License

Authors of papers retain copyright and release the work under a Creative Commons Attribution 4.0 International License ([CC BY 4.0](#)).

Summary

The Hubble Space Telescope (HST) has become the preeminent workhorse facility for the characterization of extrasolar planets. Launched in 1990 and never designed for the observations of exoplanets, the STIS spectrograph on HST was used in 2002 to detect the first atmosphere ever discovered on a planet outside of our solar system ([Charbonneau et al., 2002](#)).

HST has currently two of the most powerful operating tools in space to characterize exoplanets over a broad spectral range: The Space Telescope Imaging Spectrograph (STIS; installed in 1997) in the UV and the Wide Field Camera 3 (WFC3; installed in 2009) in the Near Infrared (NIR). With the introduction of a spatial scan mode on WFC3 ([Deming et al., 2012](#); [McCullough & MacKenty, 2012](#)) where the star moves perpendicular to the dispersion direction during an exposure, WFC3 observations have become very efficient due to the reduction of overhead time and the possibility of longer exposures without saturation.

For exoplanet characterization, WFC3 is used for transit and secondary eclipse spectroscopy and phase curve observations. The instrument has two different grisms: G102 with a spectral range from 800 nm to up to 1150 nm and G141 encompassing 1075 nm to about 1700 nm. The spectral range of WFC3/G141 is primarily sensitive to molecular absorption from water at approximately 1.4 microns. This led to the successful detection of water in the atmosphere of over a dozen of exoplanets (e.g., [Deming et al., 2013](#); [Evans et al., 2016](#); [Fraine et al., 2014](#); [Huitson et al., 2013](#); [Kreidberg, Bean, Désert, Line, et al., 2014](#)). The bluer part of WFC3, the G102 grism, is also sensitive to water and most notably led to the first detection of a helium exosphere ([Spake et al., 2018](#)).

Here we present PACMAN, an end-to-end pipeline developed to reduce and analyze HST/WFC3 data. The pipeline includes both spectral extraction and light curve fitting. The foundation of PACMAN has been already used in numerous publications (e.g., [Kreidberg, Bean, Désert, Benneke, et al., 2014](#); [Kreidberg et al., 2018](#)) and these papers have already accumulated hundreds of citations.

Statement of need

Exoplanet spectroscopy with Hubble requires very precise measurements that are beyond the scope of standard analysis tools provided by the Space Telescope Science Institute. The data analysis is challenging, and different pipelines have produced discrepant results in the literature ([Kreidberg et al., 2019](#); e.g., [Teachey & Kipping, 2018](#)). To facilitate reproducibility and transparency, the data reduction and analysis software should be open-source. This will enable easy comparison between different pipelines, and also lower the barrier to entry for newcomers in the exoplanet atmosphere field.

What sets PACMAN apart from other tools provided by the community, is that it was specifically

designed to reduce and fit HST data. For example, there are several open-source tools that can fit time series observations of stars to model events like transiting exoplanets (e.g., EXOFASTv2 (Eastman et al., 2019), juliet (Espinoza et al., 2019), allesfitter (Günther & Daylan, 2019, 2021), exoplanet (Foreman-Mackey et al., 2021a, 2021b), starry (Luger et al., 2019)). PACMAN's source code, however, includes fitting models that can model systematics which are characteristic to HST data (like the orbit-long exponential ramps due to charge trapping or the upstream-downstream effect) which removes the need for the user to write these functions themselves. PACMAN will also retrieve information from the header of the fits files, automatically detect HST orbits and visits and use this information in the fitting models.

As mentioned above, there are several open-source codes that fit light curves but PACMAN also does data reduction. Another code that can reduce HST/WFC3 observations is Iraclis (Tsiaras et al., 2016) (see Section [Similar tools](#) for more information).

Outline of the pipeline steps

The pipeline starts with the *ima* data products provided by the Space Telescope Science Institute that can be easily accessed from [MAST](#). These files created by the WFC3 calibration pipeline, *calwf3*, have already several calibrations applied (dark subtraction, linearity correction, flat-fielding) to each readout of the IR exposure.

In the following we highlight several steps in the reduction and fitting stages of the code which are typical for HST/WFC3 observations:

- **Wavelength calibration:** We create a reference spectrum based on the throughput of the respective grism (G102 or G141) and a stellar model. The user can decide if he or she wants to download a stellar spectrum from MAST or use a black body spectrum. This template is used for the wavelength calibration of the WFC3 spectra. We also determine the position of the star in the direct images which are commonly taken at the start of HST orbits to create an initial guess for the wavelength solution using the known dispersion of the grism. Using the reference spectrum as a template, we determine a shift and scaling in wavelength-space that minimizes the difference between the template and the first spectrum in the visit. This first exposure in the visit is then used as the template for the following exposures in the visit.
- **Optimal extraction and outlier removal:** PACMAN uses an optimal extraction algorithm as presented in Horne (1986) which iteratively masks bad pixels in the image. We also mask bad pixels that have been flagged by *calwf3* with data quality DQ = 4 or 512¹.
- **Scanning of the detector:** The majority of exoplanetary HST/WFC3 observations use the spatial scanning technique (McCullough & MacKenty, 2012) which spreads the light perpendicular to the dispersion direction during the exposure enabling longer integration times before saturation. The *ima* files taken in this observation mode consist of a number of nondestructive reads, also known as up-the-ramp samples, each of which we treat as an independent subexposure. [Figure 1](#) (left panel) shows an example of the last subexposure when using spatial scanning together with the expected position of the trace based on the direct image.
- **Fitting models:** PACMAN contains several functions to fit models which are commonly used with HST data. The user can fit models like in [Equation 1](#) to the white light curve or to spectroscopic light curves. An example of a raw spectroscopic light curve and fitting [Equation 1](#) to it, can be found in [Figure 2](#). Here are some examples of the currently implemented models for the instrument systematics and the astrophysical signal:

- systematic models:
 - * visit-long polynomials

¹for a list of DQ flags see <https://wfc3tools.readthedocs.io/en/latest/wfc3tools/calwf3.html>

- * orbit-long exponential ramps due to charge trapping: NIR detectors like HST/WFC3 can trap photoelectrons (Smith et al., 2008), which will cause the number of recorded photoelectrons to increase exponentially, creating typical hook-like features in each orbit
- astrophysical models:
 - * transit and secondary eclipse curves as implemented in batman
 - * sinusoids for phase curve fits
 - * a constant offset that accounts for the upstream-downstream effect (McCullough & MacKenty, 2012) caused by forward and reverse scanning

A typical model to fit an exoplanet transit in HST data is the following (used for example in Kreidberg, Bean, Désert, Benneke, et al. (2014)):

$$F(t) = T(t) (c S(t) + k t_v) (1 - \exp(-r_1 t_{\text{orb}} - r_2)), \quad (1)$$

with $T(t)$ being the transit model, c (k) a constant (slope), $S(t)$ a scale factor equal to 1 for exposures with spatial scanning in the forward direction, and s for reverse scans, r_1 and r_2 are parameters to account for the exponential ramps. t_v and t_{orb} are the times from the first exposure in the visit and in the orbit, respectively.

- **parameter estimation:** The user has different options to estimate best fitting parameters and their uncertainties:
 - least squared: `scipy.optimize`
 - MCMC: `emcee` (Foreman-Mackey et al., 2013)
 - nested sampling: `dynesty` (Speagle, 2020)
- **multi-visit observations**
 - PACMAN has also an option to share parameters across visits.
- **binning of the light spectrum:** The user can freely set the amount and locations of the bins. Figure 1 (right panel) shows the resulting 1D spectrum and a user-defined binning.

Figure 1 and Figure 2 show some figures created by PACMAN during a run using three HST visits of GJ 1214 b collected in GO 13201 (Bean, 2012). An analysis of all 15 visits was published in Kreidberg, Bean, Désert, Benneke, et al. (2014). The analysis of three visits here using PACMAN, is consistent with the published results.

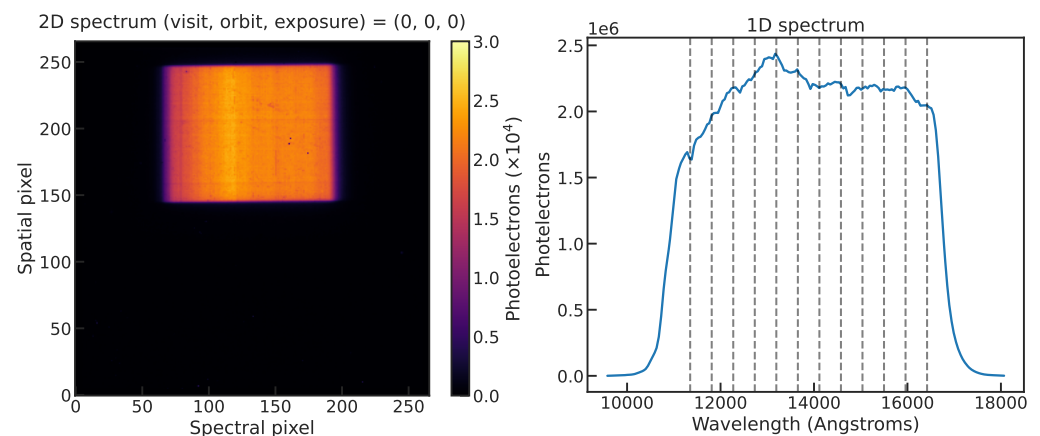


Figure 1: *Left panel:* a typical single exposure showing the raw 2D spectrum. *Right panel:* 1D spectrum after the use of optimal extraction including vertical dashed lines showing the user-set binning to generate spectroscopic light curves.

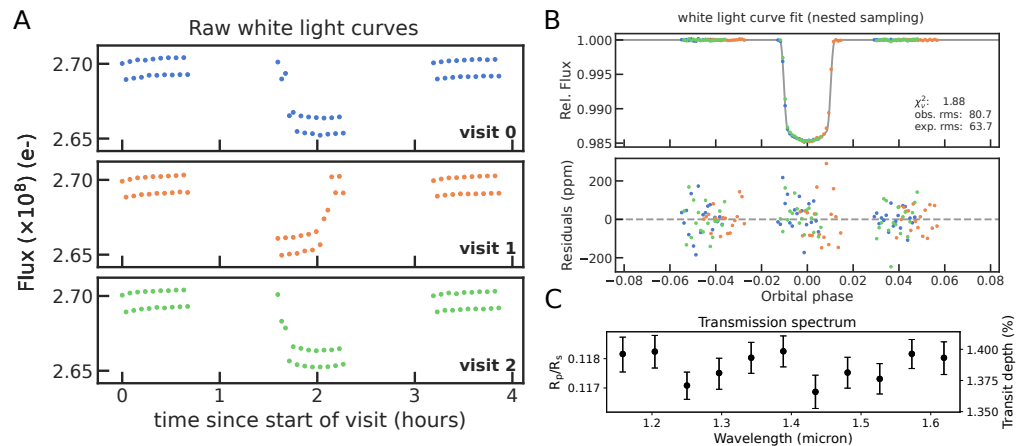


Figure 2: *panel A:* raw white light curves for each of the three visits. One can clearly see the constant offset between two adjacent exposures due to the spatial scanning mode. *panel B:* white light curve with the best astrophysical model fit using Equation 1. *panel C:* the transmission spectrum after fitting 11 spectroscopic light curves revealing the flat spectrum of GJ 1214 b as published in Kreidberg, Bean, Désert, Benneke, et al. (2014).

Dependencies

PACMAN uses typical dependencies of astrophysical python codes: numpy (Harris et al., 2020), matplotlib (Hunter, 2007), scipy (Virtanen et al., 2020) and astropy (Astropy Collaboration et al., 2018, 2013).

Other dependencies required for the fitting stage depending on the model and sampler being run are: batman (Kreidberg, 2015), emcee (Foreman-Mackey et al., 2013), dynesty (Speagle, 2020) and corner (Foreman-Mackey, 2016).

For the barycentric correction, PACMAN accesses the [API to JPL's Horizons system](#).

If the user decides to use a stellar spectrum for the wavelength calibration, PACMAN will download the needed fits file from the [REFERENCE-ATLAS HLSP](#) hosted on the MAST archive (STScI Development Team, 2013).

Documentation

The documentation for PACMAN can be found at pacmandocs.readthedocs.io hosted on [ReadTheDocs](#). It includes most notably, a full explanation of every parameter in the *pacman control file* (pcf), the API, and an example of how to download, reduce and analyze observations of GJ 1214 b taken with HST/WFC3/G141.

Similar tools

The only other end-to-end open source pipeline² specifically developed for the reduction and analysis of HST/WFC3 data is [Iraclis](#) (Tsiaras et al., 2016).

Another open-source pipeline that has been for example used as an independent check of recent results presented in Mugnai et al. (2021) and Carone et al. (2021) is [CASCADe](#) (Calibration of trAnsit Spectroscopy using CAusal Data). For a more detailed discussion of CASCADe see Appendix 1 in Carone et al. (2021).

²as far as we are aware of

Future work

Additional fitting models are planned to be added to PACMAN in the future like phase curves using the open-source python package SPIDERMAN.

Another model that can be added to fit the orbit-long ramps is the [RECTE systematic model](#).

It is planned to add a limb darkening calculation if the user wants to fix limb darkening parameters to theoretical models in the fitting stage.

PACMAN could also be expanded by adding a WFC3/UVIS data reduction in the future.

Acknowledgements

We acknowledge B. Zawadzki for the creation of the PACMAN logo. We also acknowledge the comments and contributions by I. Momcheva to PACMAN.

References

- Astropy Collaboration, Price-Whelan, A. M., Sipőcz, B. M., Günther, H. M., Lim, P. L., Crawford, S. M., Conseil, S., Shupe, D. L., Craig, M. W., Dencheva, N., Ginsburg, A., VanderPlas, J. T., Bradley, L. D., Pérez-Suárez, D., de Val-Borro, M., Aldcroft, T. L., Cruz, K. L., Robitaille, T. P., Tollerud, E. J., ... Astropy Contributors. (2018). The Astropy Project: Building an Open-science Project and Status of the v2.0 Core Package. *156*(3), 123. <https://doi.org/10.3847/1538-3881/aabc4f>
- Astropy Collaboration, Robitaille, T. P., Tollerud, E. J., Greenfield, P., Droettboom, M., Bray, E., Aldcroft, T., Davis, M., Ginsburg, A., Price-Whelan, A. M., Kerzendorf, W. E., Conley, A., Crighton, N., Barbary, K., Muna, D., Ferguson, H., Grollier, F., Parikh, M. M., Nair, P. H., ... Streicher, O. (2013). Astropy: A community Python package for astronomy. *558*, A33. <https://doi.org/10.1051/0004-6361/201322068>
- Bean, J. (2012). *Revealing the Diversity of Super-Earth Atmospheres* (p. 13021). HST Proposal ID 13021. Cycle 20.
- Carone, L., Mollière, P., Zhou, Y., Bouwman, J., Yan, F., Baeyens, R., Apai, D., Espinoza, N., Rackham, B. V., Jordán, A., Angerhausen, D., Decin, L., Lendl, M., Venot, O., & Henning, T. (2021). Indications for very high metallicity and absence of methane in the eccentric exo-Saturn WASP-117b. *646*, A168. <https://doi.org/10.1051/0004-6361/202038620>
- Charbonneau, D., Brown, T. M., Noyes, R. W., & Gilliland, R. L. (2002). Detection of an Extrasolar Planet Atmosphere. *568*(1), 377–384. <https://doi.org/10.1086/338770>
- Deming, D., Wilkins, A., McCullough, P., Burrows, A., Fortney, J. J., Agol, E., Dobbs-Dixon, I., Madhusudhan, N., Crouzet, N., Desert, J.-M., Gilliland, R. L., Haynes, K., Knutson, H. A., Line, M., Magic, Z., Mandell, A. M., Ranjan, S., Charbonneau, D., Clampin, M., ... Showman, A. P. (2013). Infrared Transmission Spectroscopy of the Exoplanets HD 209458b and XO-1b Using the Wide Field Camera-3 on the Hubble Space Telescope. *774*(2), 95. <https://doi.org/10.1088/0004-637X/774/2/95>
- Deming, D., Wilkins, A., McCullough, P., Madhusudhan, N., Agol, E., Burrows, A., Charbonneau, D., Clampin, M., Desert, J., Gilliland, R., Knutson, H., Mandell, A., Ranjan, S., Seager, S., & Showman, A. (2012). Infrared Spectroscopy of the Transiting Exoplanets HD189733b and XO-1 Using Hubble WFC3 in Spatial Scan Mode. *American Astronomical Society Meeting Abstracts #219*, 219, 405.05.
- Eastman, J. D., Rodriguez, J. E., Agol, E., Stassun, K. G., Beatty, T. G., Vanderburg, A., Gaudi, B. S., Collins, K. A., & Luger, R. (2019). EXOFASTv2: A public, generalized,

- publication-quality exoplanet modeling code. *arXiv e-Prints*, arXiv:1907.09480. <https://arxiv.org/abs/1907.09480>
- Espinoza, N., Kossakowski, D., & Brahm, R. (2019). juliet: a versatile modelling tool for transiting and non-transiting exoplanetary systems. *490*(2), 2262–2283. <https://doi.org/10.1093/mnras/stz2688>
- Evans, T. M., Sing, D. K., Wakeford, H. R., Nikolov, N., Ballester, G. E., Drummond, B., Kataria, T., Gibson, N. P., Amundsen, D. S., & Spake, J. (2016). Detection of H₂O and Evidence for TiO/VO in an Ultra-hot Exoplanet Atmosphere. *822*(1), L4. <https://doi.org/10.3847/2041-8205/822/1/L4>
- Foreman-Mackey, D. (2016). Corner.py: Scatterplot matrices in python. *The Journal of Open Source Software*, *1*(2), 24. <https://doi.org/10.21105/joss.00024>
- Foreman-Mackey, D., Hogg, D. W., Lang, D., & Goodman, J. (2013). emcee: The MCMC Hammer. *125*(925), 306. <https://doi.org/10.1086/670067>
- Foreman-Mackey, D., Luger, R., Agol, E., Barclay, T., Bouma, L. G., Brandt, T. D., Czekala, I., David, T. J., Dong, J., Gilbert, E. A., Gordon, T. A., Hedges, C., Hey, D. R., Morris, B. M., Price-Whelan, A. M., & Savel, A. B. (2021a). *exoplanet: Gradient-based probabilistic inference for exoplanet data & other astronomical time series* (Version 0.5.1). Zenodo; Zenodo. <https://doi.org/10.5281/zenodo.1998447>
- Foreman-Mackey, D., Luger, R., Agol, E., Barclay, T., Bouma, L., Brandt, T., Czekala, I., David, T., Dong, J., Gilbert, E., Gordon, T., Hedges, C., Hey, D., Morris, B., Price-Whelan, A., & Savel, A. (2021b). *exoplanet: Gradient-based probabilistic inference for exoplanet data & other astronomical time series*. *The Journal of Open Source Software*, *6*(62), 3285. <https://doi.org/10.21105/joss.03285>
- Fraine, J., Deming, D., Benneke, B., Knutson, H., Jordán, A., Espinoza, N., Madhusudhan, N., Wilkins, A., & Todorov, K. (2014). Water vapour absorption in the clear atmosphere of a Neptune-sized exoplanet. *513*(7519), 526–529. <https://doi.org/10.1038/nature13785>
- Günther, M. N., & Daylan, T. (2019). *allesfitter: Flexible star and exoplanet inference from photometry and radial velocity* (p. ascl:1903.003). Astrophysics Source Code Library, record ascl:1903.003.
- Günther, M. N., & Daylan, T. (2021). Allesfitter: Flexible Star and Exoplanet Inference from Photometry and Radial Velocity. *254*(1), 13. <https://doi.org/10.3847/1538-4365/abe70e>
- Harris, C. R., Millman, K. J., van der Walt, S. J., Gommers, R., Virtanen, P., Cournapeau, D., Wieser, E., Taylor, J., Berg, S., Smith, N. J., Kern, R., Picus, M., Hoyer, S., van Kerkwijk, M. H., Brett, M., Haldane, A., del Río, J. F., Wiebe, M., Peterson, P., ... Oliphant, T. E. (2020). Array programming with NumPy. *585*(7825), 357–362. <https://doi.org/10.1038/s41586-020-2649-2>
- Horne, K. (1986). An optimal extraction algorithm for CCD spectroscopy. *98*, 609–617. <https://doi.org/10.1086/131801>
- Huitson, C. M., Sing, D. K., Pont, F., Fortney, J. J., Burrows, A. S., Wilson, P. A., Ballester, G. E., Nikolov, N., Gibson, N. P., Deming, D., Aigrain, S., Evans, T. M., Henry, G. W., Lecavelier des Etangs, A., Showman, A. P., Vidal-Madjar, A., & Zahnle, K. (2013). An HST optical-to-near-IR transmission spectrum of the hot Jupiter WASP-19b: detection of atmospheric water and likely absence of TiO. *434*(4), 3252–3274. <https://doi.org/10.1093/mnras/stt1243>
- Hunter, J. D. (2007). Matplotlib: A 2D Graphics Environment. *Computing in Science and Engineering*, *9*(3), 90–95. <https://doi.org/10.1109/MCSE.2007.55>
- Kreidberg, L. (2015). batman: BASic Transit Model cAlculationN in Python. *127*(957), 1161. <https://doi.org/10.1086/683602>

- Kreidberg, L., Bean, J. L., Désert, J.-M., Benneke, B., Deming, D., Stevenson, K. B., Seager, S., Berta-Thompson, Z., Seifahrt, A., & Homeier, D. (2014). Clouds in the atmosphere of the super-Earth exoplanet GJ1214b. *505*(7481), 69–72. <https://doi.org/10.1038/nature12888>
- Kreidberg, L., Bean, J. L., Désert, J.-M., Line, M. R., Fortney, J. J., Madhusudhan, N., Stevenson, K. B., Showman, A. P., Charbonneau, D., McCullough, P. R., Seager, S., Burrows, A., Henry, G. W., Williamson, M., Kataria, T., & Homeier, D. (2014). A Precise Water Abundance Measurement for the Hot Jupiter WASP-43b. *793*(2), L27. <https://doi.org/10.1088/2041-8205/793/2/L27>
- Kreidberg, L., Line, M. R., Parmentier, V., Stevenson, K. B., Loudon, T., Bonnefoy, M., Faherty, J. K., Henry, G. W., Williamson, M. H., Stassun, K., Beatty, T. G., Bean, J. L., Fortney, J. J., Showman, A. P., Désert, J.-M., & Arcangeli, J. (2018). Global Climate and Atmospheric Composition of the Ultra-hot Jupiter WASP-103b from HST and Spitzer Phase Curve Observations. *156*(1), 17. <https://doi.org/10.3847/1538-3881/aac3df>
- Kreidberg, L., Luger, R., & Bedell, M. (2019). No Evidence for Lunar Transit in New Analysis of Hubble Space Telescope Observations of the Kepler-1625 System. *877*(2), L15. <https://doi.org/10.3847/2041-8213/ab20c8>
- Luger, R., Agol, E., Foreman-Mackey, D., Fleming, D. P., Lustig-Yaeger, J., & Deitrick, R. (2019). starry: Analytic Occultation Light Curves. *157*(2), 64. <https://doi.org/10.3847/1538-3881/aae8e5>
- McCullough, P., & MacKenty, J. (2012). *Considerations for using Spatial Scans with WFC3* (p. 8). Space Telescope WFC Instrument Science Report.
- Mugnai, L. V., Modirrousta-Galian, D., Edwards, B., Changeat, Q., Bouwman, J., Morello, G., Al-Refaie, A., Baeyens, R., Bieger, M. F., Blain, D., Gressier, A., Guilluy, G., Jaziri, Y., Kiefer, F., Morvan, M., Pluriel, W., Poveda, M., Skaf, N., Whiteford, N., ... Beaulieu, J.-P. (2021). ARES. V. No Evidence For Molecular Absorption in the HST WFC3 Spectrum of GJ 1132 b. *161*(6), 284. <https://doi.org/10.3847/1538-3881/abf3c3>
- Smith, R. M., Zavodny, M., Rahmer, G., & Bonati, M. (2008). A theory for image persistence in HgCdTe photodiodes. In D. A. Dorn & A. D. Holland (Eds.), *High energy, optical, and infrared detectors for astronomy III* (Vol. 7021, p. 70210J). <https://doi.org/10.1117/12.789372>
- Spake, J. J., Sing, D. K., Evans, T. M., Oklopčić, A., Bourrier, V., Kreidberg, L., Rackham, B. V., Irwin, J., Ehrenreich, D., Wyttenbach, A., Wakeford, H. R., Zhou, Y., Chubb, K. L., Nikolov, N., Goyal, J. M., Henry, G. W., Williamson, M. H., Blumenthal, S., ... Madhusudhan, N. (2018). Helium in the eroding atmosphere of an exoplanet. *557*(7703), 68–70. <https://doi.org/10.1038/s41586-018-0067-5>
- Speagle, J. S. (2020). DYNESTY: a dynamic nested sampling package for estimating Bayesian posteriors and evidences. *493*(3), 3132–3158. <https://doi.org/10.1093/mnras/staa278>
- STScI Development Team. (2013). *pysynphot: Synthetic photometry software package* (p. ascl:1303.023).
- Teachey, A., & Kipping, D. M. (2018). Evidence for a large exomoon orbiting Kepler-1625b. *Science Advances*, *4*(10), eaav1784. <https://doi.org/10.1126/sciadv.aav1784>
- Tsiaras, A., Waldmann, I. P., Rocchetto, M., Varley, R., Morello, G., Damiano, M., & Tinetti, G. (2016). A New Approach to Analyzing HST Spatial Scans: The Transmission Spectrum of HD 209458 b. *832*(2), 202. <https://doi.org/10.3847/0004-637X/832/2/202>
- Virtanen, P., Gommers, R., Oliphant, T. E., Haberland, M., Reddy, T., Cournapeau, D., Burovski, E., Peterson, P., Weckesser, W., Bright, J., van der Walt, S. J., Brett, M., Wilson, J., Millman, K. J., Mayorov, N., Nelson, A. R. J., Jones, E., Kern, R., Larson, E., ...

SciPy 1. 0 Contributors. (2020). SciPy 1.0: fundamental algorithms for scientific computing in Python. *Nature Methods*, 17, 261–272. <https://doi.org/10.1038/s41592-019-0686-2>