Disintegrating Exoplanets: Creating Size Constraints by Statistically Peering Through the Debris*

KEITH M BAKA
UAZORENA HONORS COLLEGE

EVERETT SCHLAWIN
UAzORENA
Steward Observatory

ABSTRACT

We study two intriguing disintegrating exoplanets, Kepler-1520b and K2-22b, and attempt to constrain the size of the underlying objects. These two planets are being disintegrated by their host stars, spewing dust and debris pulled from their surface into tails that trail and precede the exoplanet in its orbit, making it difficult to discern the true nature of the object. We attempted to peer through the dust cloud to put a constraint on the maximum radii of these exoplanets. While previous studies have done this in the past by selecting shallow transit events, we attempt a new statistical approach to model the intrinsic astrophysical and photon noise distributions simultaneously. We assume that the lightcurve flux distribution is distributed as a convolution of a Gaussian photon noise component and a Raleigh astrophysical component. The Raleigh curve has a finite flux maximum, which we fit with a Hamiltonian Markov Chain. With these methods, a more accurate flux maximum may be estimated, producing a more accurate and better final value for the size of these exoplanets. To determine statistical significance, we used the python package PyMC3 to find the posterior distribution for our data with Gaussian, Rayleigh, and joint function curves and plotting it against our collected flux.

After completing this analysis, we were unable to constrain the radii of the exoplanets, as the forward scattering by dust dominates over dust extinction. However, this does mean that we were able better able to constrain the astrophysical variability and its maximum with our analysis.

Keywords: Exoplanets, Disintegrating Systems, K2-22b, Kepler-1520b

1. INTRODUCTION

Kepler-1520 is a star located approximately 620pc from Earth with an exoplanet candidate with a period of 0.65 days orbiting it. The Kepler observatory collected long cadence photometry of Kepler-1520 for 4 years, Borucki et al. (2010). In 2012, a study conducted by Rappaport et al. (2012) revealed that the orbiting exoplanet was a candidate for being a disintegrating exoplanet, as the light curve has flux dips ranging between 0.2% and 1.3%, while maintaining a highly regular interval of approximately 0.654 days. The coincidence of these behaviors makes it likely that Kepler-1520b is a disintegrating rocky exoplanet with a variable comet-like tail of debris trailing behind it.

Another similar exoplanet, K2-22b, was first investigated in 2015 by Sanchis-Ojeda et al. (2015). This study showed an exoplanet orbiting its host star with a consistent period of 0.381 days while producing transit depths between 0% and 1.3%. Additionally, analysis of the light curves revealed features at the ingress and egress points for the exoplanet transit consistent with both a leading and following dust trail. With these two traits, it is likely that K2-22b is another a disintegrating rocky exoplanet similar to Kepler-1520b.

The 2018 review by van Lieshout & Rappaport (2018) estimated the minimum size of both planets. By analyzing the variability in the extinction due to the debris trail, they were able to determine the mass-loss rates for planets of different sizes that produce a debris trail of that size. Through this and the assumed lifetime of the planet, van Lieshout & Rappaport (2018) estimated that both planets are of size on the order of 500km.
Figure 1. Phase-folded and binned lightcurve for the transit of Kepler-1520b. This data, collected by Kepler, contains approximately 1773 transits of the exoplanet. This lightcurve is the average of all those transits, and clearly shows a transit profile that is unusual for an exoplanet, as it is asymmetric and shows an increase in flux preceding transit.

While previous papers, such as Rappaport et al. (2012) and Sanchis-Ojeda et al. (2015), have studied the approximate size and nature of the exoplanets in conjunction with their trail(s) of debris, the goal of this paper is to look through the veil of the debris trails in order to characterize the underlying planets by placing a constraint on the radius of each. Another technique used by other papers is to study quiescent transits, or transits with low dust activity (Sanchis-Ojeda et al. 2015), (Rappaport et al. 2012) in order to place a minimum value on the radius of the planets. This approach often gives good results, but can be subject to error in that the selected transits may not be the minimum dust activity.

The goal of this study is to put a tighter constraint on the maximum size of both planets. We will attempt to unpack the clouds of debris that surround the planets with precision in order to obtain a better understanding of the systems in their entirety. In section 2, we discuss how we modeled the transits of the disintegrating exoplanets, and in section 3 we discuss the results of our analysis. The size of these planets can help derive other properties, such as the lifetime of the system, planet density, surface gravity, composition, and other dynamics of the system that can help the scientific community better understand systems like these and better model planetary systems as a whole.

2. DATA REDUCTION

2.1. Observations

We use the publicly available data from the Kepler mission, with Kepler release 25 for Kepler-1520b and K2 data release 32 for K2-22b, and download the long cadence, SAP flux for both planets. To do this, we used the program lightkurve (Lightkurve Collaboration et al. 2018) to convert the raw data into a FITS file of the transits.

The next piece of code turned the raw lightcurve data into useable data via several steps. First, we normalized the flux of the data and removed any outliers that were above five times the standard deviation to get the lightcurves for the entirety of the data collection. We also used the lightkurve flatten function with a window length parameter of 101 to remove stellar activity in this step and produce flattened lightcurves. We then used a box least squared periodogram to find the best period for each planet, using results from Rappaport et al. (2012) and Sanchis-Ojeda et al. (2015) to place prior constraints on our periodogram, selecting the primary peak of the periodogram as the period for each planet. The next step was to use the period we obtained for each planet to phase-fold the lightcurve, and then finally binned the result to produce the results seen in Figure 1 and Figure 2. The asymmetry of the lightcurves is expected, and agrees with results from Rappaport et al. (2012) and Sanchis-Ojeda et al. (2015).

2.2. Time Slicing and Folding Data

With the knowledge that the Kepler probe recorded data once every 30 minutes, we then used the periods we obtained to split the binned data into 30 minute slices, 32 for Kepler-1520b and 19 for K2-22b. We wrote a code that saved the flux for each time slice, as well as compiled all the data into readable histograms, binned by flux with a bin size of 0.0005. This code also allows the user to modify the histograms in several ways, including the flux bin size, the number of time slices, and adds the ability to shift the time slice edges, just in case the middle of a transit fell on the division between two slices. This information is then written to a FITS file.
2.3. Joint Function for Photon and Astrophysical Noise

Following our two main assumptions that

1. extinction is larger than forward scattering, and
2. astrophysical variations follow a Rayleigh distribution,

we consider the photon uncertainty to follow a Gaussian probability distribution function (PDF):

\[ f_G(x) = \frac{1}{\sqrt{2\pi}\sigma_G} e^{-\frac{(x-\mu)^2}{2\sigma_G^2}}, \quad (1) \]

where \( \mu \) is the mean of the distribution and \( \sigma_G \) is the standard deviation.

We assume that the dust tail can only decrease the flux mid-transit and therefore follows a PDF with a sharp cutoff. We assume that it follows a Rayleigh distribution:

\[ f_R(x) = \begin{cases} \frac{x}{\sigma_R} e^{-\frac{x^2}{2\sigma_R^2}} & x < \mu_R \\ 0 & x \geq \mu_R, \end{cases} \quad (2) \]

where \( \sigma_R \) is the Rayleigh scale parameter and \( \mu_R \) is the maximum of the flux distribution, which is near 1.0. The PDF of the sum of a random variable following the Rayleigh and Gaussian PDFs will be the convolution.

\[ (f \ast g)(x) = \int_{-\infty}^{\infty} f_G(x-u)f_R(u)\,du \quad (3) \]

\[ (f \ast g)(x) = \frac{f_G(x)}{1 + \frac{\sigma_R^2}{\sigma_G^2}} + \frac{\sigma_R}{2\sigma_G^2} \left( xe^{-\frac{x^2}{2\sigma_G^2}} \text{erfc}\left( \frac{-x\mu_R}{\sqrt{2\sigma_R^2\sigma_G^2}} \right) \right), \quad (4) \]

where \( x = x - \mu_R \),

\[ \sigma_2 = \sqrt{\sigma_R^2 + \sigma_G^2}, \quad (5) \]

and \( \text{erfc} \) is the complementary error function:

\[ \text{erfc}(x) = 1 - \frac{2}{\sqrt{\pi}} \int_0^x e^{-t^2} \,dt, \quad (6) \]

and \( x \) is the random variable for the flux.

The final step in our data processing accomplishes several tasks in the same code. First, we read in the information created in the previous step, and iterate through all the time slices, plotting histograms of each time slice, while also fitting both Rayleigh and joint Gaussian-Rayleigh curves to the histograms. The histogram and curve fits for each individual slice is saved separately. Then we combine and average all the out of transit data into one histogram and apply another curve fit, seen in Figures 3 and 4. While the data is being iterated, we also accumulate a list of the maximum point of flux from each time slice and plot it against the transit phase, with one example being Figure 5.

3. RESULTS & DISCUSSION

3.1. Kepler-1520b

The results we obtained for Kepler-1520b leave lots of room for interpretation and potential further study. First, when examining the maximum time series, several interesting patterns emerge. Starting with Figure 5, we see that during the transit, the maximum value for the flux actually increases. In case these time slices were too wide, and the values for maximum flux were a result of starlight scattering from the dust tail either before or...
after the transit, we chose to make two more versions of this plot with different time slice sizes.

We can also examine smaller time slices of half the exposure time, which allows us to study the time behavior at the Nyquist rate. After remaking the maximum time series with 15 minute time slices, shown in Figure 6, a new feature appears. Right at the center of the transit, a large dip occurs. Looking at the blue error bars, and keeping in mind that this graph is the summation of approximately 2000 transits, a dip of this size is not insignificant. Additionally, there is a slower build to the transit ingress, suggestive of a small leading debris trail scattering light to our detectors.

With two different sizes of time slices, it becomes necessary to compare with the transit time of Kepler-1520b for several values of the impact parameter. We can use the equation

\[ T_{\text{tot}} = \frac{P}{\pi} \sin^{-1} \left[ \frac{R_p}{a} \sqrt{(1 + k)^2 - b^2 \sin^2(i)} \right], \]  

(7)

where

\[ k = \frac{R_p}{R_*}, \ b = \frac{\cos(i)}{R_*} \]  

(8)

from Seager (2010), and use stellar parameters from Rappaport et al. (2012) and the exoplanet radius value from van Lieshout & Rappaport (2018) to obtain values for the transit durations, assuming a circular orbit. Doing this, we find that with impact parameters of 0, 0.5, and 1, the transit times are 70.4, 61.2, and 6.8 minutes respectively, or 0.075, 0.065, and 0.0036 of the orbit, meaning that for all but a nearly grazing transit, our slices should effectively capture the transit of Kepler-1520b.

In both versions of the time series though, one important pattern is visible. We can see in both graphs that the egress of the planet has a slow decline in flux that culminates in a plateau before returning to normal, suggesting a large debris tail that maintains a significant size even well after the planet has transited. We can use this to our advantage by saying that since the peak flux of each time series has the same value, we can operate under the assumption that the dip present in Figure 6 is a result of a true planet transit causing extinction amongst the forward-scattered light. Going off this assumption, we can say that the planet blocked at most 0.06% of the flux coming from the star. Thus, using the equation

\[ \Delta F = \left( \frac{R_{\text{planet}}}{R_{\text{star}}} \right)^2 \]  

(9)
along with $R_{\text{Star}} = 0.65R_\odot$ (Rappaport et al. (2012)), we can calculate that the maximum possible radius for Kepler-1520b is approximately 11000 km, which is much larger than the upper limit reported by van Werkhoven et al. (2014) using an assumed albedo and the upper limit of the eclipse. Therefore, it is far more likely that this is the radius of the part of the dust cloud that passes in front of the star.

3.2. K2-22b

K2-22b has a dip that goes below the out-of-transit flux, indicating a planet transit. This appears more like a conventional planet transit of a non-disintegrating body. However, the increase in flux on either side of the dip indicates forward scattering in Figures 8 and 9.

Again, we need to calculate the transit duration of K2-22b for various impact parameters. We can use 7, stellar parameters from Sanchis-Ojeda et al. (2015), the exoplanet radius from Schlawin et al. (2021), and the same values for the impact parameter of 0, 0.5, and 1 to find transit times of 53.5, 46.7 and 3.4 minutes respectively, or 0.098, 0.085, and 0.0047 of the orbit, again meaning that for all but a nearly grazing transit, our slices should effectively capture the transit of K2-22b.

When looking at these graphs, two things become obvious. First, in Figure 8, there is a clear dip in flux below unity, meaning this is most likely an exoplanet transit. Second, when looking at Figure 9, one can clearly see that there is an increase in flux both before and after the transit, characteristic of a disintegrating exoplanet with both a leading dust tail and a debris trail that both cause forward-scattering.

We first want to put a constraint on the maximum size of K2-22b. First, when looking at Figure 8, we see the flux drops down and then goes immediately back up, for an average change in flux of 0.15%. Then, using Equation 9 and $R_{\text{Star}} \approx 0.57R_\odot$ (Sanchis-Ojeda et al. (2015)), we can place an upper limit of the radius of K2-22b at just over 15000km, which almost exactly the same radius asserted by Sanchis-Ojeda et al. (2015). However, this may be the result of the statistical fluctuation so we also study the lightcurve with larger duration slices.

If we study the lightcurve with 30 minute slices (as would be caused by a low-impact parameter transit), Figure 9, we can see a similar feature to the one we studied in Figure 6 with Kepler-1520b, in which the maximum flux increases as the transit occurs due to forward scattering, then decreases as the planet transits the star. If we average the change in flux between the slices associated with the two peaks and the dip, we find a change in flux of 0.082%, a much smaller value than before, as we are using wider time slices where forward scattering and the transit are combined. This is the result of a long duration transit. Again, using Equation 9 and $R_{\text{Star}} \approx 0.57R_\odot$ (Sanchis-Ojeda et al. (2015)), we find the radius of K2-22b to be 11300km, or $\sim 1.78R_\oplus$, much smaller than the previous calculation in this paper, but slightly larger than the upper limit of $\sim 0.7R_\oplus$ found in Schlawin et al. (2021).

3.3. Bayesian Analysis

To further test our statistical assumptions about the nature of these objects, we used the software package pymc3 (Salvatier J. et al. 2016) so that we could investigate posterior distributions of the models we designed, as well as the maximum a priori solutions for both planets.
We started by modeling the out-of-transit data using a Gaussian prior for both the mean and standard deviation. As a result of the normalization process we used, we created a prior on the out-of-transit mean held constant at $\mu_P = 1$, where $\mu_P$ is the prior on the out-of-transit mean, and a prior on the standard deviation equal to the photon noise of the out-of-transit data.

For in-transit data, we returned to the joint Gaussian and Rayleigh function we created in 2.3 to model the same data. In order to create a model prior using this function, we created PyMC3 random variable objects for three variables as seen in equation 3 and equation 4: the standard deviation of the Gaussian, $\sigma_G$, along with those for the Rayleigh distribution, $\mu_R$ and $\sigma_R$. For the Rayleigh distribution, we chose the prior on the mean to be $\mu_R = 1$ and the Rayleigh scale parameter to be $\sigma_R = 0.01$. For the Gaussian component of the in-transit data, we used the results for the out-of-transit Gaussian data, holding $\sigma_G$ constant and only allowing the Rayleigh to shift.

3.3.1. Kepler-1520b
When looking at Figure 10, we can see that during the transit, there is a difference in the shape between the observed flux and the joint function due to the elongated tails in the observed flux. This is due to the Rayleigh component having an abrupt truncation at one end, and a long, continually decreasing tail on the other. This is different than the observed fluxes, which appears to have segments where the counts stay constant as the flux decreases. Even though that this is due to a shortcoming of the Rayleigh distribution and is an area for future research using more flexible functions to fit the astrophysical variation, the Rayleigh and Gaussian convolution still describes the general width and extent of the measured flux.

Additionally, in the violin plot on the right of Figure 10, we should expect that the Rayleigh component of the joint function is a Delta function centered around flux = 1 during out-of-transit phases, but this isn’t what we see in this plot. Instead, in the slices surrounding the in-transit phase, we can see a widening of the Rayleigh component of the joint function. The most likely cause of the variation of this Rayleigh component is a small preceding dust trail and a longer dust tail behind Kepler-1520b, confirming our predictions from Figure 1.

In Figure 12, we are looking at the distribution of values for $\mu_R$, or the maximum of the Rayleigh component of our joint function. We’re studying this result within our posterior distribution as the in-transit slices, where there is astrophysical variation, should be dominated by this Rayleigh component. We can see that everywhere that there is astrophysical variation, as indicated by a widening of the Rayleigh component from a delta function, the maximum of $\mu_R$ goes well above unity. This means that our assumptions that the transit is well-described by a Rayleigh and that the flux will decrease more than the increase from forward scattering during transit are flawed, as we should expect this value to go down during transit of the planet and the dust tails.

The reason we may be seeing an increase in $\mu_R$ during the in-transit phases is that the transit of Kepler-1520b may be a grazing transit, meaning that forward-scattering can have a greater impact on the amount of light collected than the obscuring of the light by the planet and the dust during transit. Additionally, our method of attempting to discern the size of the underlying planet using the average of many transits may be flawed, as forward-scattering by dust at different phases of the orbit may be clouding our results, meaning that we may be greatly underestimating the amount of forward-scattering possible.

### 3.3.2. K2-22b

The results for K2-22b are fairly similar to those of Kepler-1520b in that the maximum flux does not go below 1.0, but there is one key difference to discuss.

The most notable is found in the violin plot on the right of Figure 11, where we can see that the Rayleigh component of the joint function deviates significantly from a delta function starting at phase = -0.25 through the transit, an indication of astrophysical variation. The results of this deviation can be seen in Figure 13, where the distributions of $\mu_R$ are elevated significantly above 1. The cause for this variation is likely a leading dust trail orbiting with K2-22b, an new result only obtainable via the width and maximum of the distribution, as there is no evidence for this dust when looking at Figure 2. This result also helps explain unexpected variations in Figures 8 and 9.

### 3.3.3. Future Research

Due to our incorrect assumption that the in-transit data would be well modeled by a Rayleigh distribution, we believe that there is room for future research and improvement upon these results. We suggest that either a Rice distribution or a function with more than two parameters is used to better describe the irregular shapes of the in-transit data in order to produce better results.

### 4. SUMMARY

In this paper, we have examined the disintegrating exoplanets Kepler-1520b and K2-22b in a new way in an attempt to place a better constraint on the maximum size of the exoplanets. We stacked our data so that every period of each planet was overlaid, ensuring that
the transits were aligned properly. From there, we were able to slice this stacked lightcurve into increments of time that were one time and two times the frequency of Kepler observations. We took these time slices and made histograms of the data, calculated Rayleigh and joint Gaussian-Rayleigh curve fits, where the Gaussian component was fixed to the out-of-transit distribution, and saved a plot for each slice. We also compiled the maximum flux of each slice into a plot, as well as an average of every out-of-transit time slice, to which we fit a Gaussian curve.

We looked at the maximum time series of each exoplanet in an attempt to put a new constraint on the maximum size of the planets for two different time slice sizes, and found hints of transit features for the 15 minute time slices. Even though both plots contained features characteristic of disintegrating exoplanets, neither produced a result on the maximum radius not already obtained by previous studies.

Lastly, we used PyMC3 and Bayesian analysis to check the statistical significance of our results. We modeled the transiting exoplanets with a joint Gaussian and Rayleigh distribution, studying the posterior distributions of the joint function after fitting it to our observed data. We noticed features characteristic to disintegrating exoplanets, but due to shortcomings in our model, were unable to put firm size limits on the underlying planets. Since the maximum of the Rayleigh distribution is greater than 1.0 during the in-transit phase, our results are dominated by forward scattering of dust, meaning that the transits may be only grazing transits, or that forward scattering from dust in other parts of the orbit obscures the transit from the underlying planet. However, the increase of the width of the Rayleigh prior to the transit of K2-22b is a new result obtained via our method, indicating astrophysical variation, or a dust trail, preceding the transit. Therefore, we believe that future research is warranted with a new distribution modeling the astrophysical variation.

Acknowledgments

Funding for ES is supported by the NASA Goddard Spaceflight center. We respectfully acknowledge the University of Arizona is on the land and territories of Indigenous peoples. Today, Arizona is home to 22 federally recognized tribes, with Tucson being home to the O’odham and the Yaqui. Committed to diversity and inclusion, the University strives to build sustainable relationships with sovereign Native Nations and Indigenous communities through education offerings, partnerships, and community service.

Software: astropy (Astropy Collaboration et al. 2018), numpy (Oliphant 2006--), lightkurve (Lightkurve Collaboration et al. 2018), pymc3 (Salvatier J. et al. 2016), and scipy (Virtanen et al. 2020)

Appendix

All code used to compute our values and generate our plots can be found at Baka (2020).

References

Astropy Collaboration, Price-Whelan, A. M., Sipőcz, B. M., et al. 2018, AJ, 156, 123, doi: 10.3847/1538-3881/aabc4f

Baka, K. 2020, Disintegrating Exoplanets, 1.0, Github.

https://github.com/bakak17/Disintegrating-Exoplanets
Borucki, W. J., Koch, D., Basri, G., et al. 2010, Science, 327, 977, doi: 10.1126/science.1185402

Lightkurve Collaboration, Cardoso, J. V. d. M., Hedges, C., et al. 2018, Lightkurve: Kepler and TESS time series analysis in Python, Astrophysics Source Code Library. http://ascl.net/1812.013

Oliphant, T. 2006–, NumPy: A guide to NumPy, USA: Trelgol Publishing. http://www.numpy.org/

Rappaport, S., Levine, A., Chiang, E., et al. 2012, ApJ, 752, 1, doi: 10.1088/0004-637X/752/1/1

Salvatier J., Wiecki T.V., & Fonnesbeck C. 2016, Probabilistic programming in Python using PyMC3, 3.11.2, PeerJ Computer Science. https://peerj.com/articles/cs-55/

Sanchis-Ojeda, R., Rappaport, S., Pallè, E., et al. 2015, ApJ, 812, 112, doi: 10.1088/0004-637X/812/2/112

Schlawin, E., Su, K. Y. L., Herter, T., Ridden-Harper, A., & Apai, D. 2021, AJ, 162, 57, doi: 10.3847/1538-3881/ac0b41

Seager, S. 2010, Exoplanets

van Lieshout, R., & Rappaport, S. A. 2018, Disintegrating Rocky Exoplanets, 15, doi: 10.1007/978-3-319-55333-7_15

van Werkhoven, T. I. M., Brogi, M., Snellen, I. A. G., & Keller, C. U. 2014, A&A, 561, A3, doi: 10.1051/0004-6361/201322398

Virtanen, P., Gommers, R., Oliphant, T. E., et al. 2020, Nature Methods, 17, 261, doi: 10.1038/s41592-019-0686-2
All Authors and Affiliations

Keith M Baka$^1$
UArizona Honors College

Everett Schlawin$^1$

$^1$UArizona
Steward Observatory