



# Fine Structures in the Main Sequence Revealed by Gaia Data Release 2

Wei-Chun Jao (饒惟君)<sup>1</sup> and Gregory A. Feiden<sup>2</sup>

<sup>1</sup> Department of Physics and Astronomy, Georgia State University, Atlanta, GA 30303, USA; [jao@astro.gsu.edu](mailto:jao@astro.gsu.edu)

<sup>2</sup> Department of Physics and Astronomy, University of North Georgia, Dahlonega, GA, 30597, USA

Received 2020 April 28; revised 2020 June 29; accepted 2020 June 29; published 2020 August 6

## Abstract

The Hertzsprung–Russell diagram (HRD) is fully examined using the Fourier analysis. This work shows more stars are above the gap in the lower main sequence than below it, and this implies that stars spend more time above the gap while they undergo variability associated with the  $^3\text{He}$  instability. The enhanced HRD also shows the width of the gap is not linear and depends on the  $G_{\text{BP}} - G_{\text{RP}}$  color up until  $G_{\text{BP}} - G_{\text{RP}} = 2.7$ . Beyond this color limit, the gap is hardly seen. Besides, a new low density region is revealed for the first time centered at  $M_G \approx 10.7$  and  $G_{\text{BP}} - G_{\text{RP}} \approx 2.8$ , which is below the lower right corner of the gap. This work also shows that the main sequence appears to have fine stripes where stellar densities are relatively low or high compared to their adjacent regions on the main sequence. These stripes can be seen throughout the main sequence of stars redder than  $G_{\text{BP}} - G_{\text{RP}} = 0.8$  and are not limited to any specific color or spectral type. Slopes of these features are different from the main sequence, but are pretty consistent throughout the main sequence, with a few exceptions. We are perplexed by these new features, but the complexities of stellar atmospheric features and opacities of dwarfs may have caused these patterns.

*Unified Astronomy Thesaurus concepts:* Hertzsprung Russell diagram (725); Main sequence stars (1000); Parallax (1197)

## 1. Introduction

It has been more than a century since the Hertzsprung–Russell diagram (HRD) was first plotted by Rosenberg (1910)<sup>3</sup> and later by Hertzsprung (1911) and Russell (1914). Since then, the total number of stars with trigonometric parallaxes has increased with efforts from the ground and space. The latest Gaia Data Release 2 (DR2) provides unprecedented astrometric precision for over a billion stars (Gaia Collaboration et al. 2018) and has revolutionized the way we see this fundamental diagram in stellar astronomy. Features like new white dwarf sequences (Kilic et al. 2018), a double main sequence (Gaia Collaboration et al. 2018), and a new gap in the lower main sequence (Jao et al. 2018) can be identified on the HRD for the first time. These new features, like all other previously known features on the HRD, including the Hertzsprung gap, Hayashi track, Henyey track, instability strip, and main-sequence turnoff point, are linked to the physics of stellar interior structures and atmospheres. Upon closer inspection of the HRD, we notice one intriguing low density stripe marked in box B in Figure 1. This diagonal feature has a low stellar density compared to its adjacent region. In Sections 2 and 3, we discuss how we process the HRD to enhance this feature. In Section 4, we then apply the algorithm to different samples to validate this feature and others identified in this work. Finally, in Section 5, we highlight and investigate prominent features on the HRD. In the appendices, we elaborate on the data processing procedures and present the enhanced HRD from late-G to late-M dwarfs.

## 2. Processing the HRD

The HRD has often been presented in a two-dimensional (2D) map by using a star’s color and absolute magnitudes as

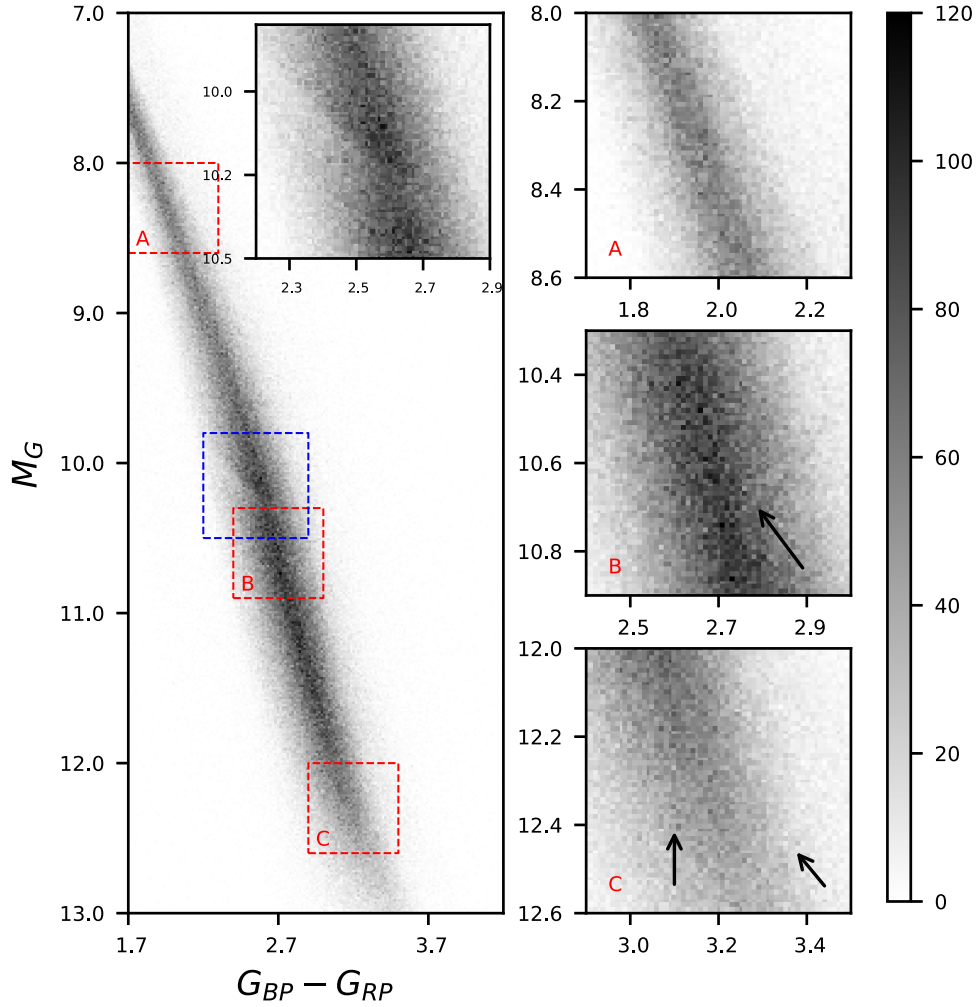
proxies of effective temperature and luminosity, even though a star’s location on this map is governed by its mass, atmospheric composition, metallicity, age, magnetic/stellar activity/spot coverage, spinning angle (pole on or edge on), or multiplicity (single or unresolved system). Given the number of parameters contributing to a star’s appearance, two single stars with the same parameters except, for example, their magnetic activities, could occupy different locations on the HRD. This means that this fundamental map contains a “heterogeneous” sample of stars, so one could naively assume that dwarfs should be randomly distributed on the main sequence in a narrow range of color and absolute magnitudes. Hence, no other feature is expected to be seen, except the gap in the lower main sequence (Jao et al. 2018).

Prior to the release of DR2, because of limited trigonometric parallaxes, the HRD of nearby stars is very sparse, especially in the lower main sequence. All works related to analysis of the HRD other than the applications of the Hipparcos catalog are limited in using a scatter plot. Since the DR2, it has been possible to make a two-dimensional (2D) density plot or three-dimensional (3D) plot to understand the population density at a given range of color and magnitude for almost all types of nearby stars for the first time. Figure 1 is an example of the 2D density plot for 1.16 million stars within 200 pc on main sequence with  $G_{\text{BP}} - G_{\text{RP}}$  color redder than 1.7.<sup>4</sup> The 1.7 color limit is selected because (1) there are at least five absolute magnitudes of void region between the main sequence and clusters of giants, (2) very few giants are redder than  $G_{\text{BP}} - G_{\text{RP}} = 1.7$ , and (3) this color limit is the end of the white dwarf sequence. It is easier for us to select a uniform main-sequence sample and we will focus on these stars in this work.<sup>5</sup> Hence, this part of the main sequence covers

<sup>3</sup> Hans Rosenberg made the first color–magnitude diagram for Pleiades in 1910. Both Iben (1991) and Valls-Gabaud (2007), which are the earliest two papers to cite Rosenberg’s work according to the ADS, also made a similar argument, but both papers incorrectly cited this paper as published in 1911. An English translated version of this paper can be seen at <http://www.leosondra.cz/en/first-hr-diagram/>.

<sup>4</sup> The selection of reliable DR2 data is based on the three-step cuts presented in Lindegren et al. (2018) and the SQL script is given in Appendix A.

<sup>5</sup> Later, we extend the main sequence to  $G_{\text{BP}} - G_{\text{RP}} = 0.8$  and it is discussed in Appendix D.



**Figure 1.** Density plot in gray scale is shown for 1.16 million stars within 200 pc. To make this density plot, the bin size is 0.01 magnitude at both abscissa and ordinate. The gap is clearly seen close to  $M_G \sim 10$  and an inset in the left figure represents the zoomed region (a blue box) close to the gap. The other three plots on the right column represent three zoomed regions, A, B, and C, highlighted as three red boxes in the left plot. The box B contains one feature we first identify and it appears to be a low stellar density band with about  $35^\circ$  angle tilted from the abscissa. The two arrows that mark features in box C will be discussed later in the text. The gray scale on the right represents the total number of stars in each bin.

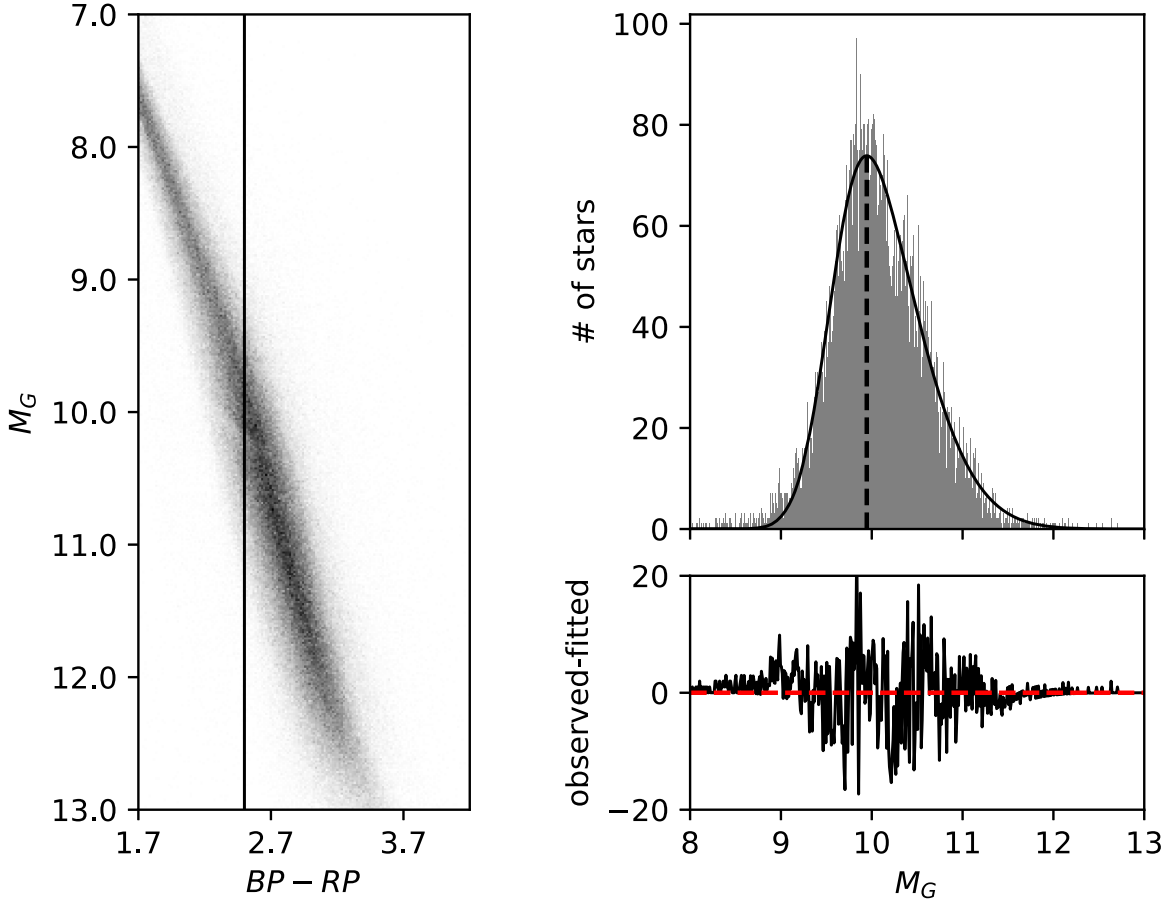
$7 < M_G < 13$  and  $1.7 < G_{BP} - G_{RP} < 4.2$  and this includes stars redder than mid-K type dwarfs.

In Jao et al. (2018), a Gaussian distribution was used to represent a stellar density distribution at a given color to highlight the gap. Here, to better fit the overall structure of the density distribution, a skewed Gaussian curve is used to represent the density distribution vertically at each color with a step size of 0.01 magnitude and the same step size is used for  $M_G$  as well. Results show using a skewed Gaussian curve yields a lower residual than a Gaussian curve and a mixture of Gaussians. An example fit is given in Figure 2. Based on this example, it shows there are slightly more stars in the center of the main sequence than the fitted curve, so the fit yields fewer stars at the peak of the curve. Also, there could be young stars at  $M_G \approx 9$  (above the main sequence), because there are more stars than the fit. Nonetheless, the gap is clearly seen at  $M_G \sim 10.3$ . After fitting the entire region of the density plot vertically slice-by-slice, a 2D residual plot can be made and is shown in Figure 3.

There are a few prominent features in Figure 3. The first feature is that the main sequence appears to have three bands shown as gray–black–gray shades. As discussed earlier, a

skewed Gaussian distribution is used to represent the density distribution at a given color, so this feature is a result of the fit and has no astrophysical meaning. The second obvious feature is the gap, which has been enhanced because of this process. Surprisingly, there are more stars above the gap than below it. Also, the gap is more prominent on the blue side of the main sequence than the red side, which is consistent with what Jao et al. (2018) reported. The other feature is a faint band just above the main sequence and it covers almost the entire main sequence. It is related to extra stars elevated above the main sequence and they may be young stars (Gagné & Faherty 2018) or unresolved binaries in DR2. The most prominent region for these young stars is at  $G_{BP} - G_{RP} < 2.7$ . On the opposite side of the main sequence from young stars at a given color, the overabundance of cool subdwarfs can also be identified at the lower edge of the main sequence.

The last feature is marked by an arrow in box B of the plot in Figure 3. The fuzzy stripe indicates a slice of a region where stellar density is relatively lower than this surrounding area, and this feature appears to have an opposite slope from the gap. By comparing with the original density plot in Figure 1, this faint feature can be identified in both plots, so this particular



**Figure 2.** Example fit of a slice of the density plot at a given color is shown. The location of the slice with a width of 0.01 mag is marked as a vertical line in the left figure. The upper right plot shows the fit to the density distribution and the residual is shown in the lower right plot. The vertical dashed line marks the peak of the skewed Gaussian and the assemblage of these peaks make the fitted line of the main sequence discussed in Appendix E. The gap is clearly seen in the lower right plot at  $M_G \approx 10.2$ .

feature in box B is not an artifact created by the way data are fitted.

### 3. Fourier Analysis on Residual Images

To further enhance these faint features and increase the contrast on the residual plot, a Fourier analysis is applied to the residual image to remove high frequency noises. The details of the analysis are discussed in Appendix B. Because the residual plot is a rectangular image, zero counts are padded on all sides to make this plot square with a size of  $1200 \times 1200$  pixels, expanding from  $250 \times 1000$  pixels. Extra padded areas are to ensure the discontinuity on the edges of the image, so that a Fourier transform will not generate aliases calculating from connecting one side of the image to the other. The step-by-step results are given in Figure 4 and the Fourier enhanced HRD is given in Figure 5.

Like Figure 3, all those features discussed there, including gray-black-gray bands, the gap, young stars/double stars above the main sequence, and subdwarfs below the main sequence, are clearly seen in Figure 5. It is noteworthy that the stripe marked in box B of Figures 1 and 3 is further enhanced. In addition, an oval shape of low density region surrounding this feature is identified for the first time. Hereafter, both features are combined and identified as feature “F1” in this work.

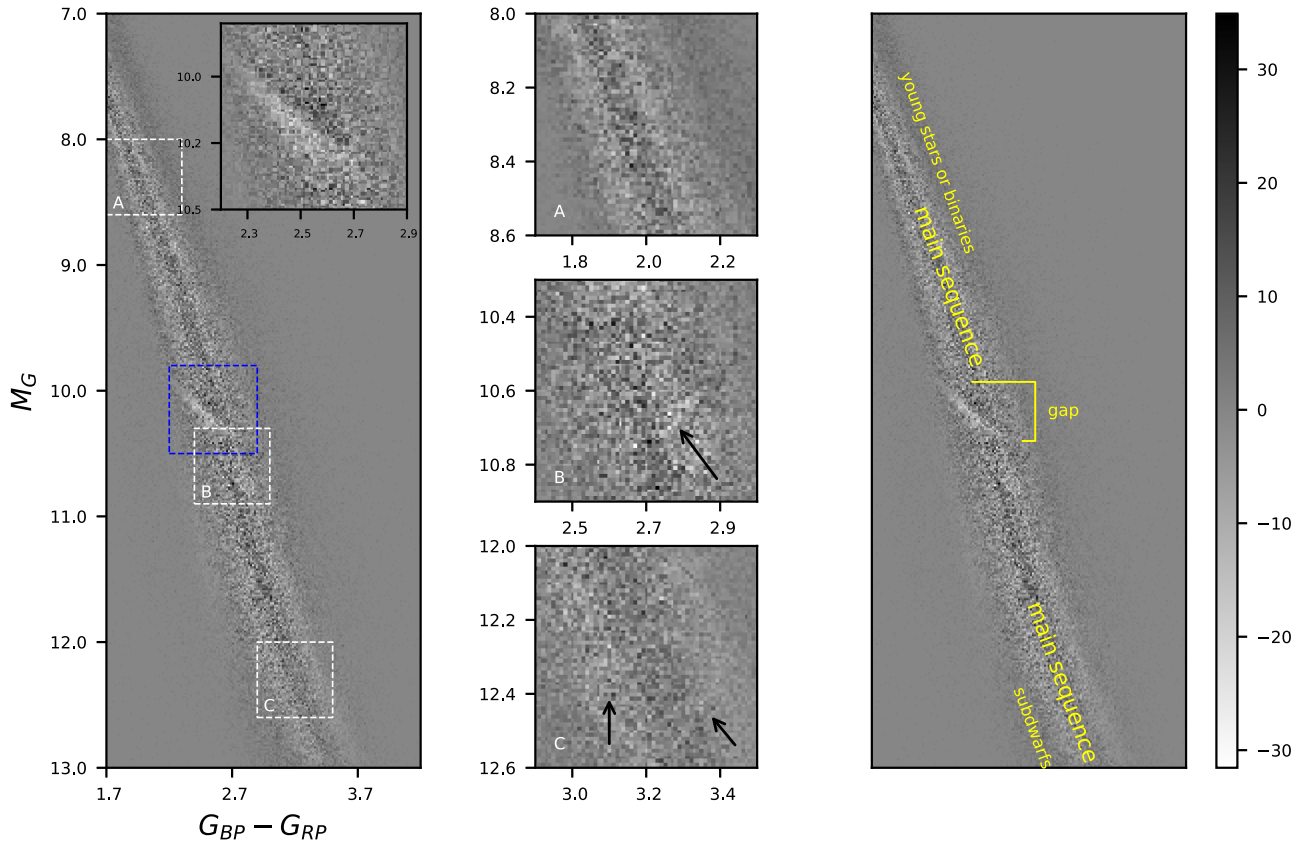
In box C, we can also see two stripes have a lighter shade, which indicates these regions have slightly fewer stars than fitted skewed Gaussian distributions. They are identified as features “F2” and “F3.” In box A, the stripes are not as prominent as the features shown in box C, but the main sequence seems been lightly brushed diagonally and these stripe-like features have different slopes from the main sequence. If the density of the main sequence is as random as the reason discussed in Section 2, we should not expect to see any pattern or feature after this process. Therefore, before we analyze these features, it is necessary to ensure that they are genuine features of the HRD.

### 4. Validation of New Features on This Enhanced HRD

#### 4.1. Simulating the Main Sequence

The fitting of a skewed Gaussian done earlier yields parameters of the density distribution at each bin, including peak location, scale, and shape. Based on these parameters, each vertical distribution is simulated slice-by-slice and random noise is added for each distribution. Then these simulated skewed Gaussian distributions are combined to make one simulated main sequence or HRD.

After this simulated HRD is generated, the same Fourier enhanced procedures are applied to the image. The difference of two simulated HRD (random1-random2 or “diff1”) would



**Figure 3.** This figure is the residual plot with a size of  $250 \times 1000$  pixels and is arranged in the same way as Figure 1. It has an additional plot on the right to label features discussed in Section 2. The arrow in box B marks that intriguing feature, which can also be identified in Figure 1. This low density stripe can also be seen in the left plot where the contrast to its surrounding regions is obvious. The two arrows marked in box C will be discussed later. Because this is a residual plot, the gray scale bar on the right contains negative counts.

still be featureless noise. If the features identified in Figure 5 are random noise, the difference between our enhanced HRD and a third randomly generated HRD (enhanced–random3 or “diff2”) would be very similar to “diff1.”<sup>6</sup> Results are shown in Figure 6 and three zoomed regions are plotted in Figure 7.

The patterns in “diff1” appear to be random speckles, which are very different from the pattern in “diff2.” Fine features identified earlier in all three boxes are not seen in the “diff1,” but can be identified in “diff2.” Consequently, these fine features are unlikely random noises.

#### 4.2. Using Different Samples

To further check these new features, two different samples are drawn. The first one contains 177,236 stars within 100 pc and the second sample has 514,515 stars within 200 pc with  $\text{mean\_flux\_error}/\text{mean\_flux}$  ratio less than 0.01 for  $G$ ,  $G_{BP}$ , and  $G_{RP}$  magnitudes. The first sample represents a subset of the original sample and the second sample is comprised of a higher quality of entries in DR2. Results from this Fourier analysis of these two samples are shown in Figure 8. Because the first sample size is significantly less than our original sample, the paucity of stellar density at each bin of the image has caused the patterns of the main sequence to be speckles, similar to the

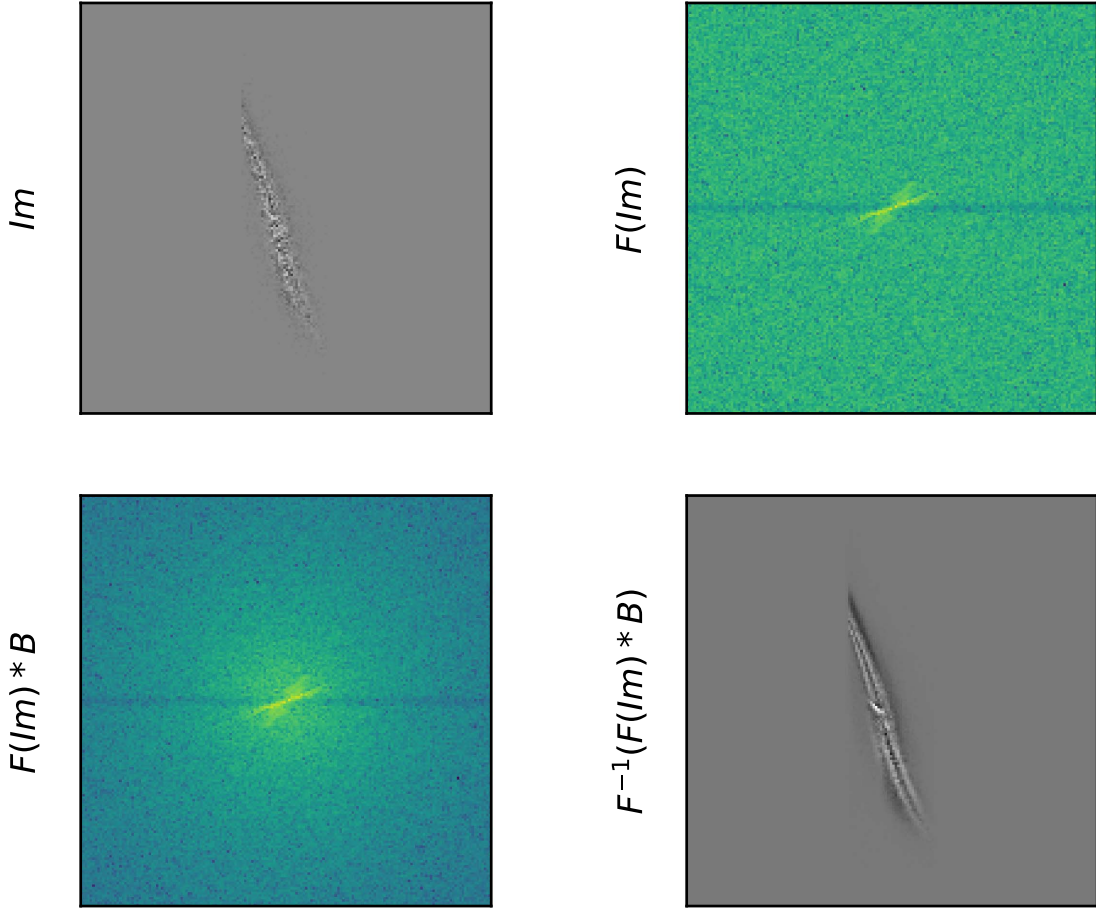
left panel in Figure 6. As for the second sample, because of a strict magnitude error limit, a significant amount of low-mass stars are missing, but the main sequence above the gap appears to be similar to the same region in Figure 5. Despite the differences in sample sizes, the feature F1 in box B can still be identified, especially for the 100 pc sample. The stripe-like features in box C are suppressed in the low main sequence in Figure 8 because of low star counts. Other samples are drawn using galactic coordinates,  $b > 0$  ( $\approx 572$  K sources) and  $b < 0$  ( $\approx 592$  K sources), but they suffer the same local paucity problems.

The other test is to acquire the 2MASS near-IR photometry to reproduce these features in the near-IR band, but it immediately encounters a difficulty. The main sequence in this  $J - K_s$  color range is almost vertical, so the only way to make a histogram is to bin the data horizontally at a given  $M_G$ , instead of using  $G_{BP} - G_{RP}$  color. However, the gap is almost horizontal and those short stripes in Figure 5 would possibly become horizontal. If the main sequence is binned horizontally and fitted using a skewed Gaussian, this analysis would remove these features entirely because the frequency of slicing and the existence of possible features would appear to be the same. To prevent this problem, this Fourier analysis is applied to the original 2D density distribution without fitting a skewed Gaussian.

We cross-match these 1.16 million Gaia sources against the 2MASS catalog to get their near-IR photometry (details of matching sources can be seen in Jao et al. 2018). To make a 2D

<sup>6</sup> A comparison between (random1–random2) and (enhanced–random2) is not performed. Because any noise from random2 would be left on (enhanced–random2), these two plots would have some noisy patterns similar to one another.





**Figure 4.** (Upper left) The residual plot as an input image for the Fourier analysis. (Upper right) A 2D Fourier spectrum after applying fast Fourier transform and shifting the zero frequency to the center of the plot. A log scale of the amplitude is shown here. (Lower left) The same 2D Fourier spectrum after applying Butterworth filter. (Lower right) An enhanced image after applying inverse fast Fourier transform. All labels are defined in Appendix B.

density distribution of the HRD, a bin size of 0.01 mag is used on both axes. Results of applying this Fourier analysis to this 2D image are shown in Figure 9. Unlike those stripes in Figure 5, only some lumps of stars or high contrast regions can be seen. The main reasons are (1) the near-IR photometry of these stars has a typical error of 0.03 mag, which is larger than the bin size used here, and (2) the span of  $M_K$  is narrower than  $M_G$ , so many points in the near-IR HRD are overlapped. Nonetheless, high contrast regions are still at same locations before and after this Fourier analysis. Although there is no obvious pattern, using near-IR photometry demonstrates this imaging enhancement method does not introduce artificial patterns.

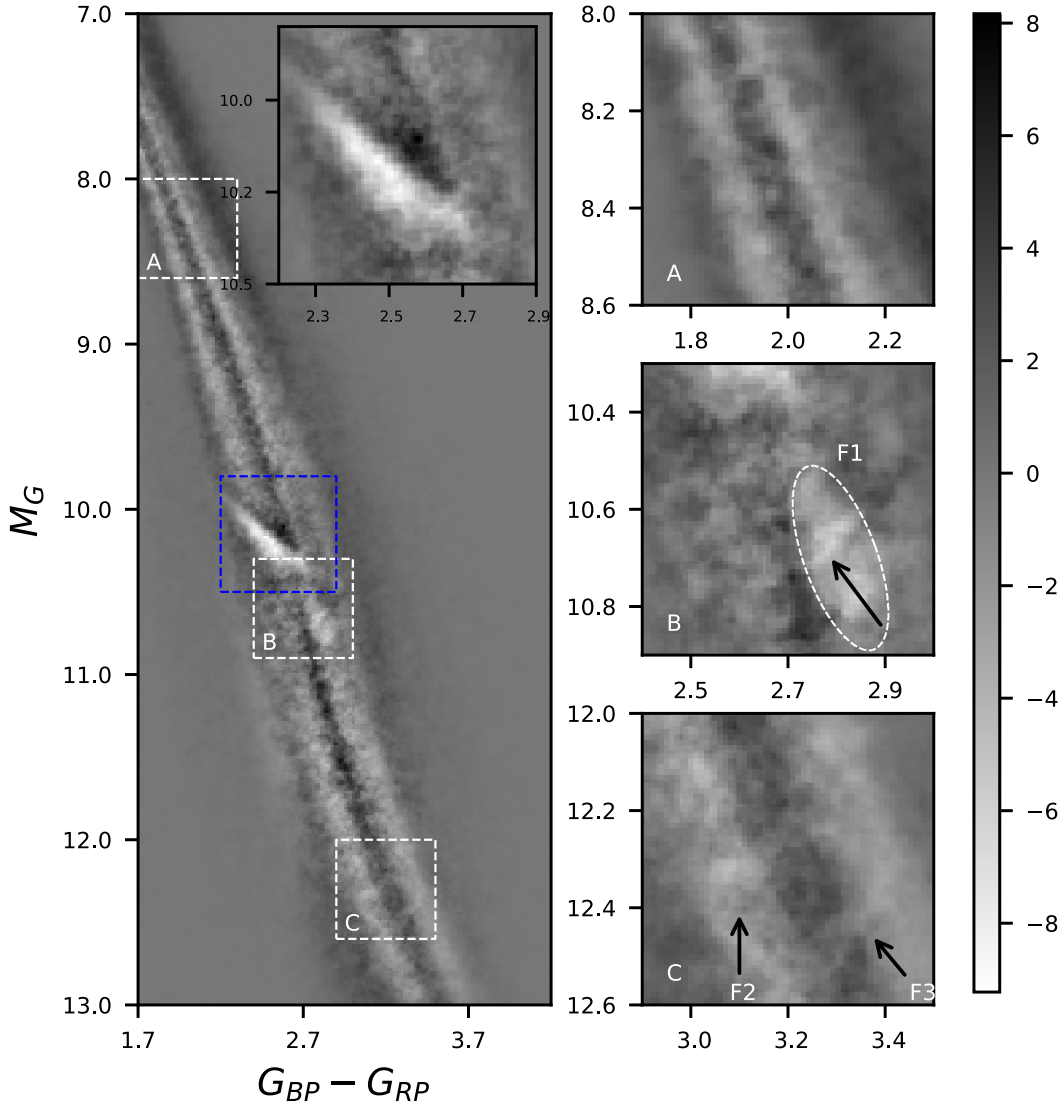
## 5. Discussion

### 5.1. The Enhanced Gap in the Lower Main Sequence

The gap reported in Jao et al. (2018) has been enhanced in this work and shows that there are more stars above the gap than below it. According the stellar evolution models proposed by van Saders & Pinsonneault (2012), MacDonald & Gizis (2018), and Baraffe & Chabrier (2018), low-mass stars with  $0.34 M_\odot \sim 0.36 M_\odot$  are expected to experience a period of structural instability caused by nonequilibrium  $^3\text{He}$  fusion in the stellar core and intermittent mixing between core and

envelope convection zones, leading to variability in stellar radii and luminosities. The underlying physics of this  $^3\text{He}$  instability are thought to produce the gap in the HRD. The overabundance of stars above the gap could then be interpreted as stars spending more time above the gap while they undergo variability associated with the  $^3\text{He}$  instability. Hence, the overdense regions should be populated by stars with similar masses and ages.

This enhanced HRD in Figure 5 also reveals that the gap is more prominent in the blue end of the gap ( $G_{\text{BP}} - G_{\text{RP}} < 2.7$ ) than in the red end. This could be the result of the gap being diluted on the red edge of the main sequence due to the presence of dwarf stars, young stars, and unresolved binaries. Because the cause of this gap in the lower main sequence is linked to stellar interior structure of low-mass stars, this feature is expected to be seen in both optical and near-IR photometries. The plot in the top row of Figure 11 demonstrates that stars selected 0.1 mag below the gap are mostly clustered in the  $J - K_s$  plot and the distribution is tilted, so it means the same underlying physics can explain both features on two different HRD plots. The other eye-catching feature of this enhanced HRD is that the width of the gap is not linear, which is not seen in Jao et al. (2018), but is a triangular shape. This shape also can be clearly identified in the nonenhanced residual plot of the HRD in Figure 3.



**Figure 5.** An enhanced residual plot after the fast Fourier analysis. This figure is arranged in the same way as Figure 1. Box A contains suspicious stripes across the main sequence. The low density stripe in box B is obvious now as it is marked by an arrow. A new extended low density region, shown as an ellipse, can also be seen centered at  $M_G \approx 10.7$  and  $G_{BP} - G_{RP} \approx 2.8$ . Both the low density stripe and this extended low density region are labeled as feature F1. Two arrows in box C mark two other low density stripe-like features, “F2” and “F3,” which have different slopes from the main sequence.

## 5.2. New Features

### 5.2.1. A New Low/High Density Region

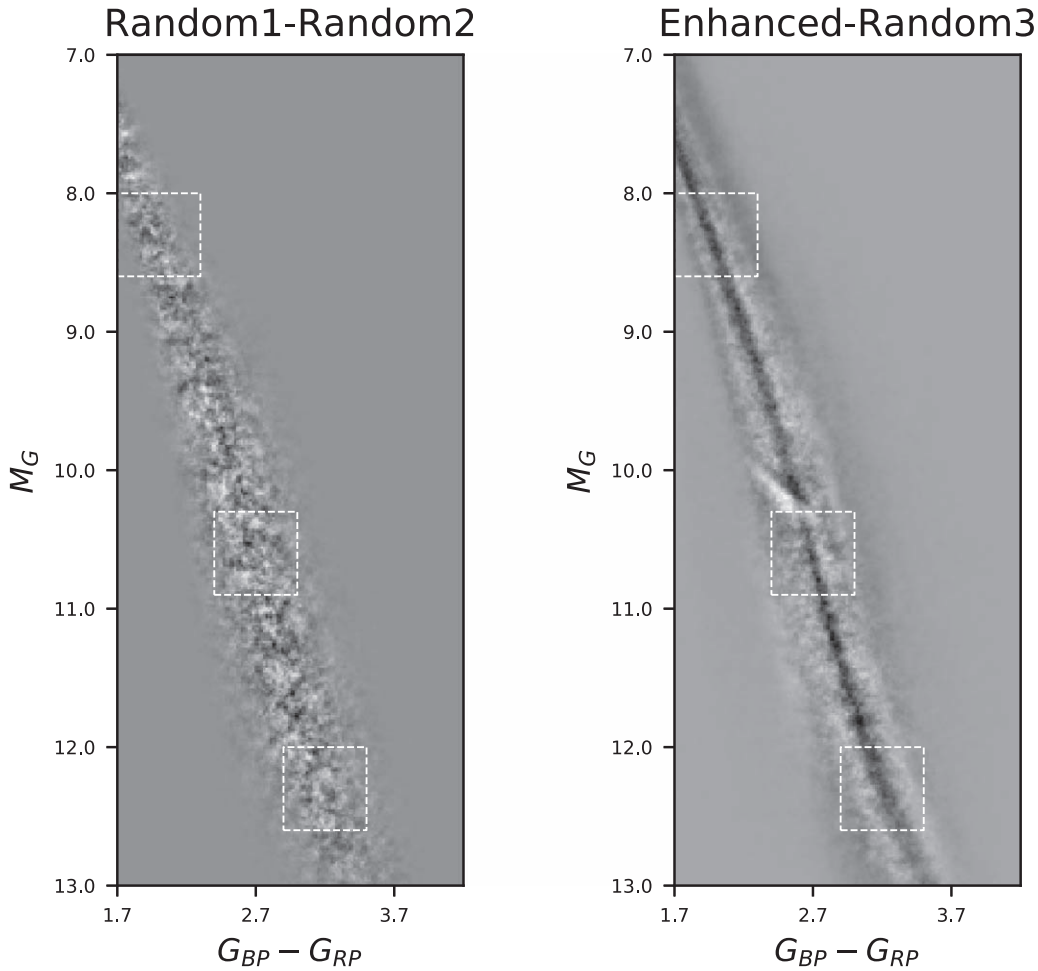
As mentioned earlier, the catalyst for this work is a diagonal low density stripe marked in box B of Figure 1 and it is enhanced in Figure 5. However, after the diagram is enhanced, a more extended low density region is seen close to its vicinity. It is located below the gap and is visible while comparing to its adjacent regions. The extended region appears to be an oval shape with the size of approximately  $0.3 \times 0.18$  magnitudes in terms of the scale of absolute  $G$  magnitudes. Both the diagonal stripe and this extended low density region are marked as feature F1 in Figure 5. This feature can also be identified in both plots of Figure 8, but is relatively subtle.

A histogram of  $M_G$  or luminosity function is made to show a dip centered at  $M_G \approx 10.7$  in Figure 10. To make this luminosity function, a bin size of 0.01 at  $M_G$  is used. A similar luminosity function was made in Jao et al. (2018), but a larger

bin size of 0.25 mag was used. Instead of using the observed  $G$  in DR2, they used converted  $V$ , so no such dip was reported in their work. By using a smaller bin size here, the dip matches the exact location of the F1 feature. In other words, Figure 10 demonstrates this Fourier analysis can enhance fine features that had previously eluded detection. Additionally, there is an overabundance/high density region just above this low density region and it appears to be associated with this low density region. The cause of this F1 feature is unknown.

### 5.2.2. New Fine Stripe-like Low Density Features

The two low density stripe-like features, F2 and F3 in Figure 5, are prominent after the Fourier analysis and they are very similar to each other. The width of these two stripes are about 0.05 mag and the two stripes are separated by about 1 magnitude in  $M_G$ . Can a group of stars in one of these new features in the  $G_{BP} - G_{RP}$  plot also be clustered in the same region in the  $J - K_s$  plot? The bottom row of Figure 11 shows



**Figure 6.** (Left) This plot shows the difference between two randomly generated HRDs, random1 and random2. (Right) This plot shows the difference between our enhanced HRD (Figure 5) and a third randomly generated HRD. Three white boxes show the same zoomed region as all previous figures and are A, B, and C from top to bottom. The left figure has speckle-like patterns, but the right figure does not have such patterns.

that a group of stars in feature F3 is highlighted in the  $G_{BP} - G_{RP}$  plot, but the same group of stars is scattered all over the main sequence in the  $J - K_s$  plot. In contrast to the gap, this particular low density feature is not linked from one tight region in the HRD to another.

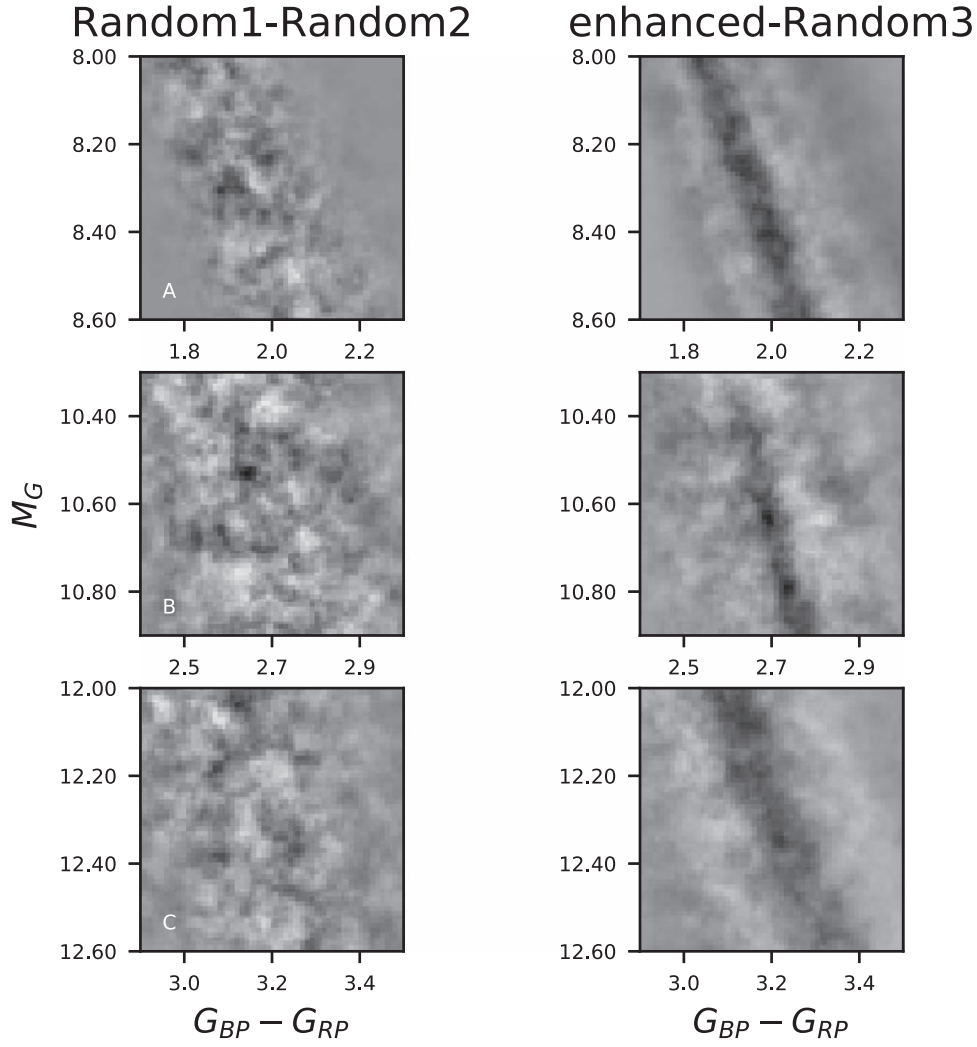
There are a few possible reasons causing these new stripe-like features. The first cause is how Gaia data are processed. If the data processing had a flaw, the gap in the lower main sequence would not appear and its location would not match models' predictions. It is also very unlikely that these features are caused by data processing by the Gaia Data Processing and Analysis Consortium. The second possible cause is that these stripe-like features are counting noise. Figure 6 shows the map would be very different if these features are noise. Also, if counting noise were random, it would be random throughout the main sequence and should not form stripes seen here. Random noise would not be associated with a specific pattern. As Figure B2 demonstrates, if the residual plot contains random noise, the enhanced image would appear like speckles, and not show any pattern after the Fourier analysis. However, if an image contains features or patterns demonstrated in Figure B3, they will be revealed after the Fourier analysis. The third possible cause is that these patterns are introduced by how data are processed here. Obviously, these stripe-like

features can also be seen in Figure 3 prior the Fourier analysis, so they are not introduced by this analysis. The last possible cause is linked to the astrophysics.

Jao et al. (2018) showed the equal mass or equal radius lines with different metallicities/ages have negative slopes based on theoretical models,<sup>7</sup> but these stripe-like features have opposite slopes. Therefore, stars in the F2 and F3 features should not have equal masses or radii and they are mixed with different masses, ages, and metallicities. Therefore, it seems reasonable that the features may be the result of atmospheric processes that leave an imprint on the stellar spectra.

To understand these stripe-like features spectroscopically, we select stars along F3 in box C as an example and cross-match these stars against the SDSS Data Release (DR) 14 to acquire their spectra. Four stars are retrieved and their spectra are shown in Figure 12. Based on the high signal-to-noise ratio spectra in SDSS DR14, we learn the following. (1) These spectra are basically identical at a wavelength shorter than 6000 Å and only differ at the redder end, so their overall spectral slopes are different. (2) LP164-55 (green) and SDS1415+6139 (red, SDSS J141521.76 +613946.0) are almost identical, but their  $G_{BP} - G_{RP}$  colors are

<sup>7</sup> Here we define the main sequence as a negative slope and a line with a fixed absolute magnitude as a slope of zero.



**Figure 7.** Three zoomed regions in Figure 6. The speckle-like features seen in the left three plots are different from the stripe-like features seen in the right three plots.

slightly different. *SDS1412+5441* (black, *SDSS J141204.75+544123.4*) has strong emissions at Balmer series and *SDS0817+4057* (blue, *SDSS J081733.96+405733.4*) is the faintest and bluest among all. (3) Although these four stars are nicely located orderly along this low density feature, their  $G_{BP} - K_s$  colors do not match this order. *SDS1415+6139* and *SDS1412+5441* are redder than the other two in  $G_{BP} - K_s$  color, but *LP164-55* (green) and *SDS0817+4057* (blue) preserved their relative positions.

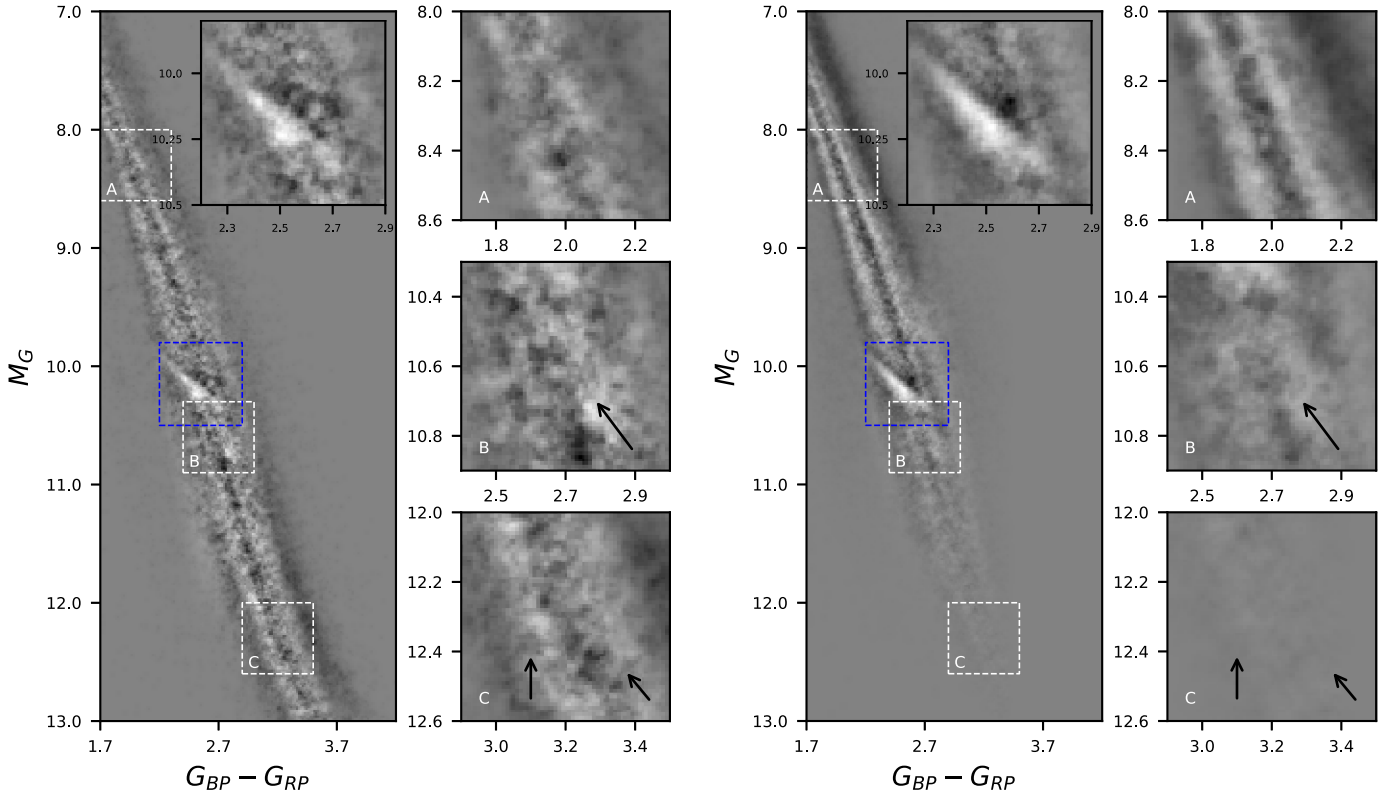
The Gaia  $G_{BP}$  and  $G_{RP}$  passbands cover 3283–6714 Å and 6296–10637 Å, respectively, so the coverage of SDSS spectra is not as wide as these two filters combined. Even though *LP164-55* and *SDS1415+6139* have almost identical spectra, they should have different fluxes beyond the SDSS spectral coverage, which causes their slight differences in  $G_{BP} - G_{RP}$  color. Also, the excess near-IR fluxes of *SDS1415+6139* and *SDS1412+5441* would imply that they may have dust, disks, or unseen low-mass companions to make them appear to be redder. The dominated opacity sources of M dwarfs in the optical band are TiO and VO (Allard & Hauschildt 1995) as shown in Figure 12, so these two molecular bands can affect the emergent flux. All four stars have been characterized as M5 in the SDSS DR14, but their spectra are quite different at a

wavelength longer than 6000 Å, especially for *SDS0817+4057*. Hence, understanding the complexity of different opacities in M dwarfs along these tight features could hold the keys to understanding the formation of these stripe-like features.

We also note that these stripe-like features can be seen throughout the main sequence as demonstrated in Figure D2 in the Appendix and they appear to cut through the main sequence like the two shown in box C. These features basically link stars (young and old) on both sides of the main sequence. More importantly, some features can even be seen for stars with  $G_{BP} - G_{RP} > 2.9$ . According to models presented in Baraffe & Chabrier (2018; MacDonald & Gizis 2018), these very low-mass stars ( $< 0.2 M_{\odot}$ ) have interior structure stable against the  $^3\text{He}$  instability, but these two low density stripes are still present and cut across the main sequence. Therefore, features F2 and F3 are unlikely to be the result of internal structure variations.

In conclusion, the Fourier analysis is applied to enhance the HRD, and it shows an excessive number of stars are just above the gap and the width of the gap is not linear. Furthermore, this work also reveals new high/low density regions at the lower right corner of the gap. Lastly, the main sequence appears to be

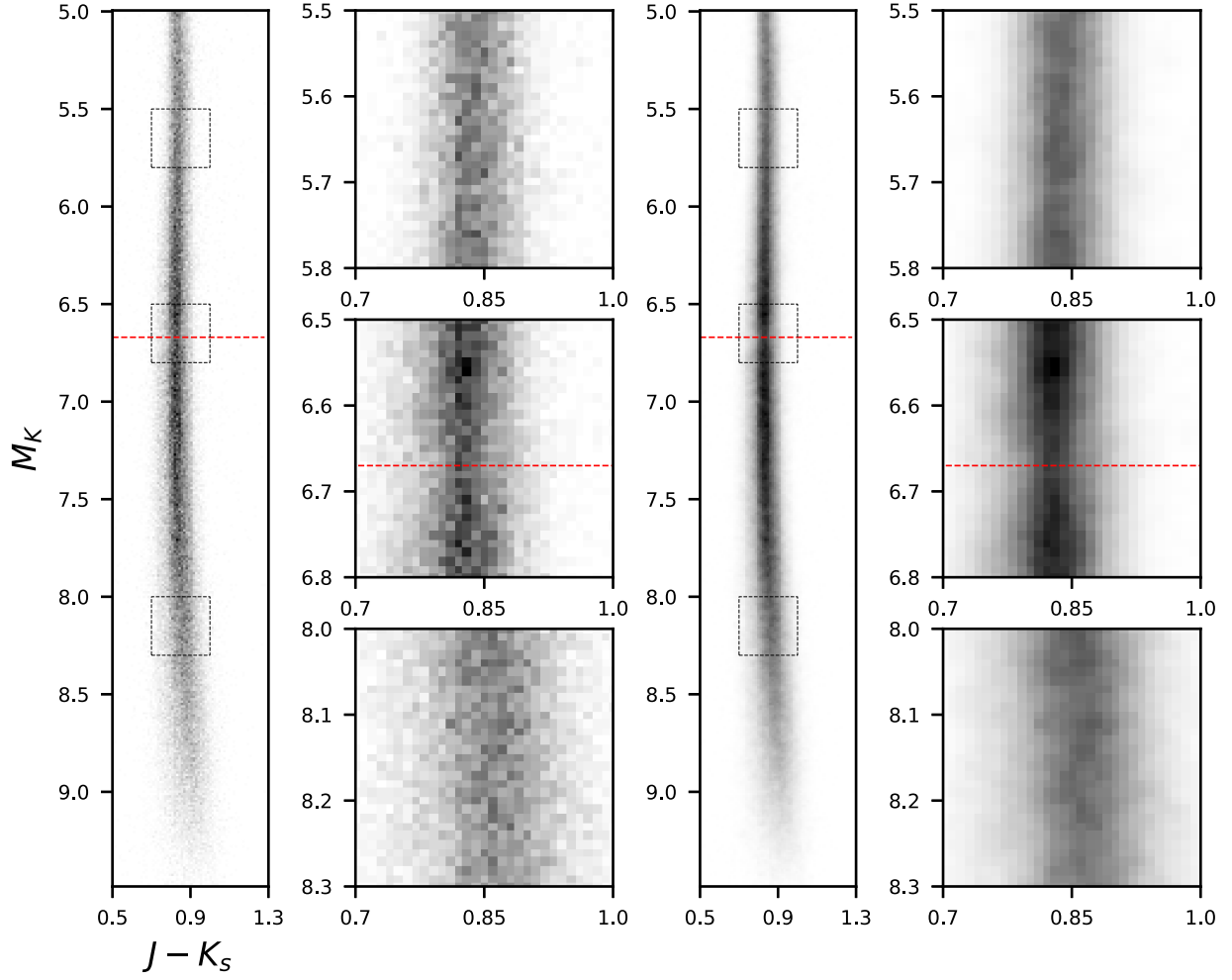




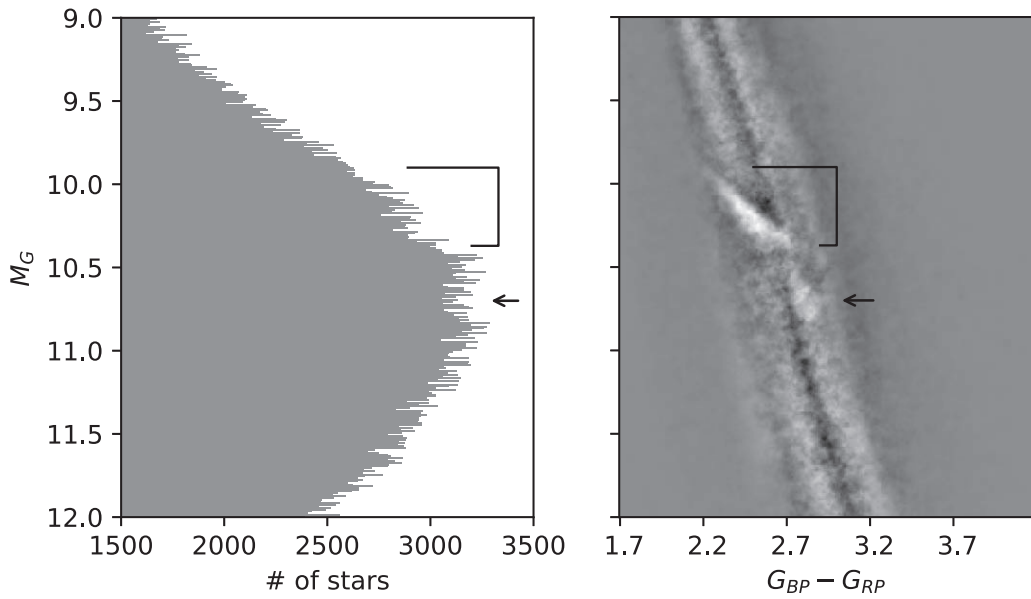
**Figure 8.** (Left) Enhanced HRD of stars within 100 pc. Because of low populations/stellar densities in this sample, the patterns on the main sequence appear to be speckles. (Right) Enhanced HRD of stars with low photometry errors within 200 pc. A significant number of stars are excluded in both samples, but the low density region in box B is still visible.

lightly brushed in one direction in general, so many stripe-like features often have similar positive slopes. The obvious features, F1 to F4 discussed in this work, are marked in red in Figure D2 and the yellow lines show other speculated features throughout the main sequence. All these features are selected subjectively by comparing to their adjacent regions. We are perplexed about (1) the causes of all these new features, (2) why these stripe-like features are seen throughout the main

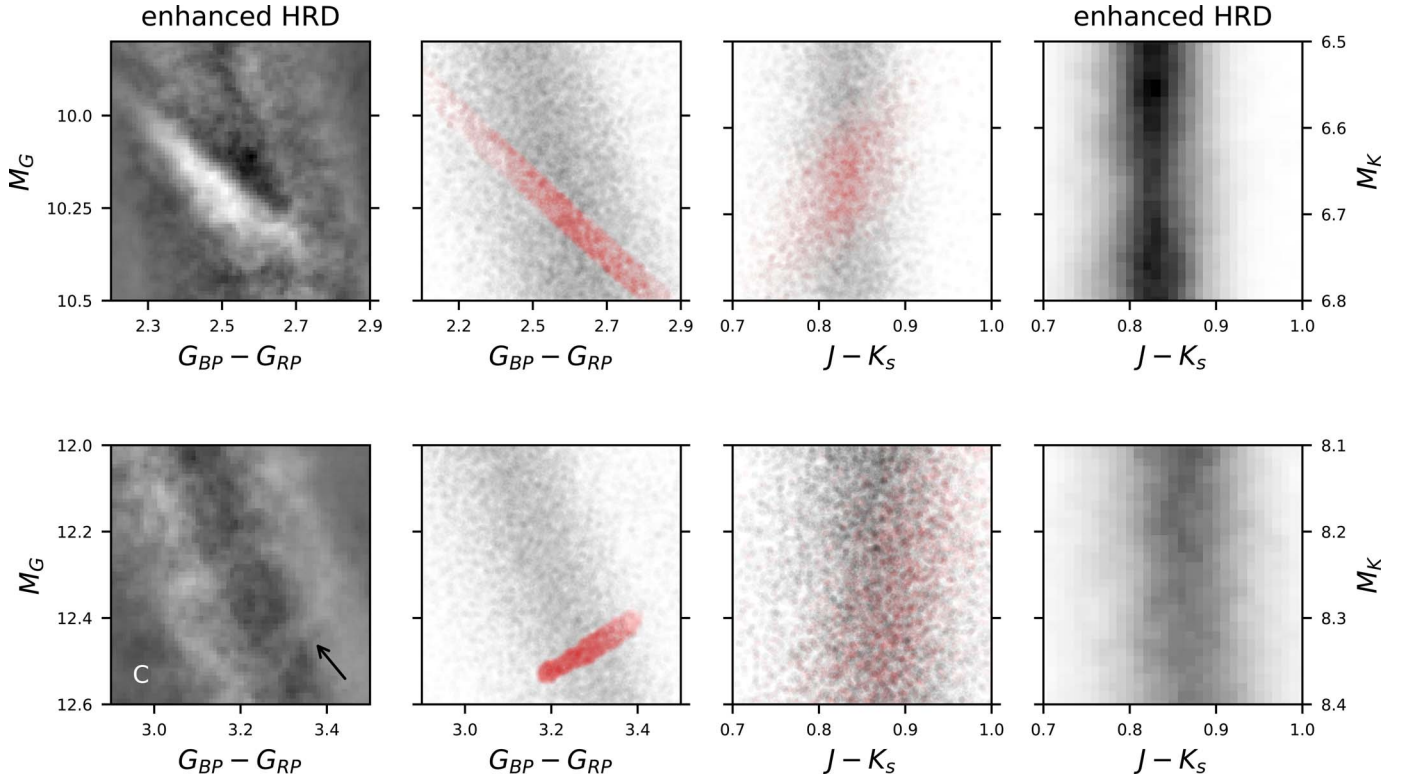
sequence, and (3) what these slopes mean. We postulate that the complex opacity sources and atmospheric features of dwarfs could make stars appear clustered and form these fine structures. A much finer theoretical model grid and detailed spectral analysis of main-sequence dwarfs would be necessary to provide insights to these new fine structures. In the forthcoming papers, we will try to present possible explanations for these new features.



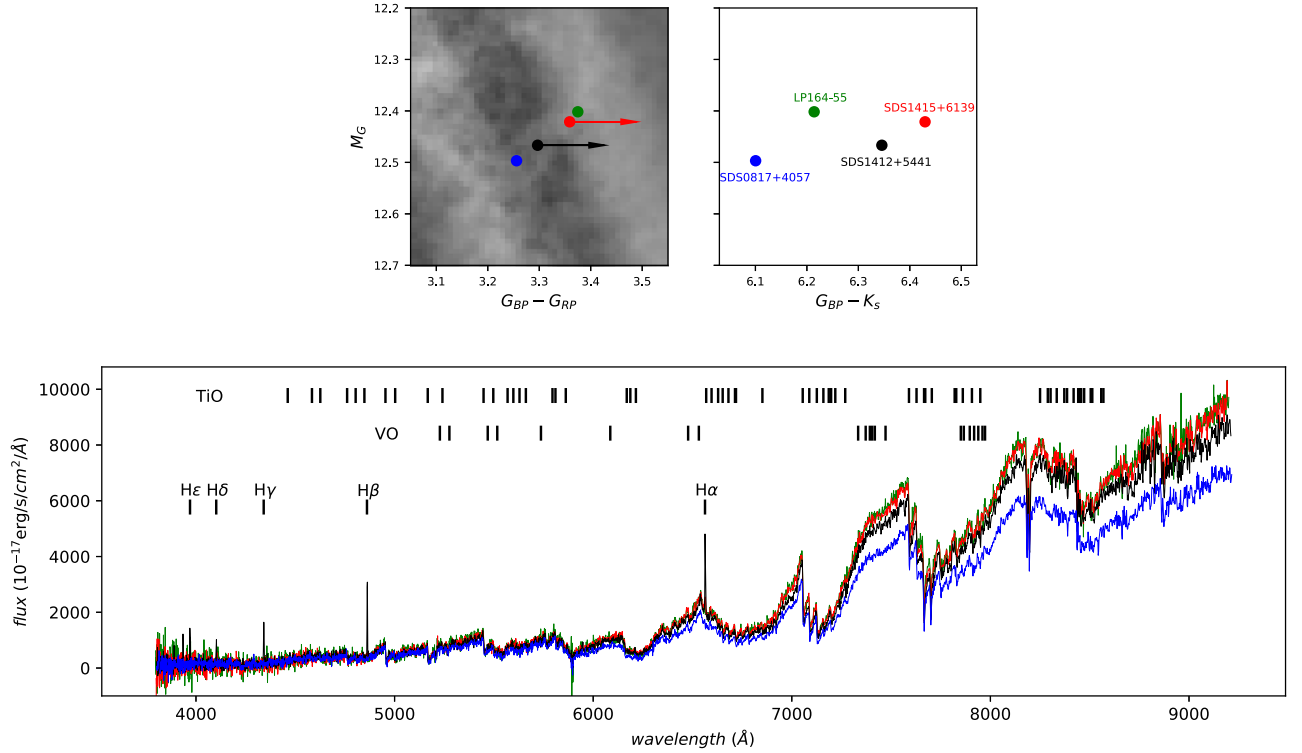
**Figure 9.** The left four figures are 2D HRD of 1.16 million stars with 2MASS near-IR photometry, boxes in the left-most plot are zoomed regions shown in the second column from the left. These zoomed plots are arranged from top to bottom relatively to the boxes in the HRD to the left. Red lines mark an approximate location of the gap. The right four figures are the enhanced 2D HRD and they are arranged the same as the four plots to the left. While comparing the three zoomed plots side by side, the local relative contrasts are kept before and after the Fourier analysis.



**Figure 10.** The stellar luminosity function (left plot) and the enhanced HRD (right plot) are shown. A bracket and an arrow mark the gap and F1 feature, respectively.



**Figure 11.** The left and right four HRDs are plotted using Gaia and 2MASS photometry, respectively. Four plots in the corners are enhanced HRDs from this work and four center plots are scatter plots of HRDs. The top row is for a region close to the gap and the bottom row is for F3 in region C in Figure 5. Red dots represent the same groups of stars in both HRDs using Gaia and 2MASS colors. Stars in the gap are consistently clustered in both HRDs, but the feature F3 in box C in Gaia color are scattered in 2MASS color.



**Figure 12.** Top row contains two HRDs using  $G_{BP} - G_{RP}$  (left) and  $G_{BP} - K_s$  (right) colors plotted against  $M_G$  for four SDSS stars along the F3 feature. The left plot also shows the enhanced HRD. The bottom plot shows the four spectra from SDSS DR14. A few important spectral features for TiO, VO, and Balmer series are labeled. The same color of text, symbol, and line is used to indicate the same star in these three plots. The two arrows indicate these two stars become redder in the  $G_{BP} - K_s$  plot. Because these stars have different distances, their fluxes are scaled to be at 10 pc using the Gaia parallaxes ( $f_{SDSS} \times (100/\varpi_{GaiaDR2})^2$ ).

We would like to thank Douglas Gies, Todd Henry, Stuart Jefferies, and Ted von Hippel for helpful discussions. We also thank Lindsey Baker for reading this manuscript. We thank an anonymous referee for many useful comments and suggestions. This work is supported by the National Science Foundation under grant AST-1715551 and G.A.F. is supported by a Presidential Semester Incentive Award from the University of North Georgia. This work has made use of data from the European Space Agency (ESA) mission Gaia (<https://www.cosmos.esa.int/gaia>), processed by the Gaia Data Processing and Analysis Consortium (DPAC, <https://www.cosmos.esa.int/web/gaia/dpac/consortium>). Funding for the DPAC has been provided by national institutions, in particular, the institutions participating in the Gaia Multi-lateral Agreement. This research has made use of the SIMBAD database, operated at CDS, Strasbourg, France. This research has made use of NASA's Astrophysics Data System. This publication makes use of data products from the Two Micron All Sky Survey, which is a joint project of the University of Massachusetts and the Infrared Processing and Analysis Center/California Institute of Technology, funded by the National Aeronautics and Space Administration and the National Science Foundation. Funding for the Sloan Digital Sky Survey IV has been provided by the Alfred P. Sloan Foundation, the U.S. Department of Energy Office of Science, and the Participating Institutions. SDSS-IV acknowledges support and resources from the Center for High-Performance Computing at the University of Utah. The SDSS website is [www.sdss.org](http://www.sdss.org). SDSS-IV is managed by the Astrophysical Research Consortium for the Participating Institutions of the SDSS Collaboration including the Brazilian Participation Group, the Carnegie Institution for Science, Carnegie Mellon University, the Chilean Participation Group, the French Participation Group, Harvard-Smithsonian Center for Astrophysics, Instituto de Astrofísica de Canarias, The Johns Hopkins University, Kavli Institute for the Physics and Mathematics of the Universe (IPMU)/University of Tokyo, the Korean Participation Group, Lawrence Berkeley National Laboratory, Leibniz Institut für Astrophysik Potsdam (AIP), Max-Planck-Institut für Astronomie (MPIA Heidelberg), Max-Planck-Institut für Astrophysik (MPA Garching), Max-Planck-Institut für Extraterrestrische Physik (MPE), National Astronomical Observatories of China, New Mexico State University, New York University, University of Notre Dame, Observatório Nacional/MCTI, The Ohio State University, Pennsylvania State University, Shanghai Astronomical Observatory, United Kingdom Participation Group, Universidad Nacional Autónoma de México, University of Arizona, University of Colorado Boulder, University of Oxford, University of Portsmouth, University of Utah, University of Virginia, University of Washington, University of Wisconsin, Vanderbilt University, and Yale University.

*Software:* TOPCAT (Taylor 2005), Matplotlib (Hunter 2007), Imfit (<https://lmfit.github.io/lmfit-py/>), scikit-learn (Pedregosa et al. 2011), NumPy (van der Walt et al. 2011), SciPy (Virtanen et al. 2020).

## Appendix A Gaia Archive Query

The following script is used to retrieve Gaia DR 2 data for this work.

---

```
SELECT source_id, ra, dec, parallax, parallax_error,
       pmra, pmra_error, pmdec, pmdec_error,
       phot_g_mean_mag, phot_bp_mean_mag, phot_rp_mean_mag,
       bp_rp, l, b
FROM gaiadr2.gaia_source
WHERE parallax > 5.0
AND parallax_over_error > 10.0
AND phot_bp_mean_flux_over_error > 10.0
AND phot_rp_mean_flux_over_error > 10.0
AND astrometric_n_good_obs_al > 5
AND (SQRT(astrometric_chi2_al/(astrometric_n_good_obs_al-5.0)) < 1.2 OR
      SQRT(astrometric_chi2_al/(astrometric_n_good_obs_al-5.0)) < 1.2*exp(-0.2*(phot_g_mean_mag-19.5)))
AND phot_bp_rp_excess_factor BETWEEN 1.0 + (0.015*POWER(phot_bp_mean_mag-phot_rp_mean_mag, 2.0))
AND 1.3 + (0.06*POWER(phot_bp_mean_mag-phot_rp_mean_mag, 2.0))
```

---

White dwarfs are excluded after the sample is retrieved. We also have tried to include an additional filter of `visibility_periods_used > 8`, but the results are the same. Results presented here do not include this additional filter.

## Appendix B Fast Fourier Transform on Images

The goal of this Fourier analysis is to enhance features on an image, and those features are often hidden by human eyes or plotting programs. The following are steps of how an image is processed.

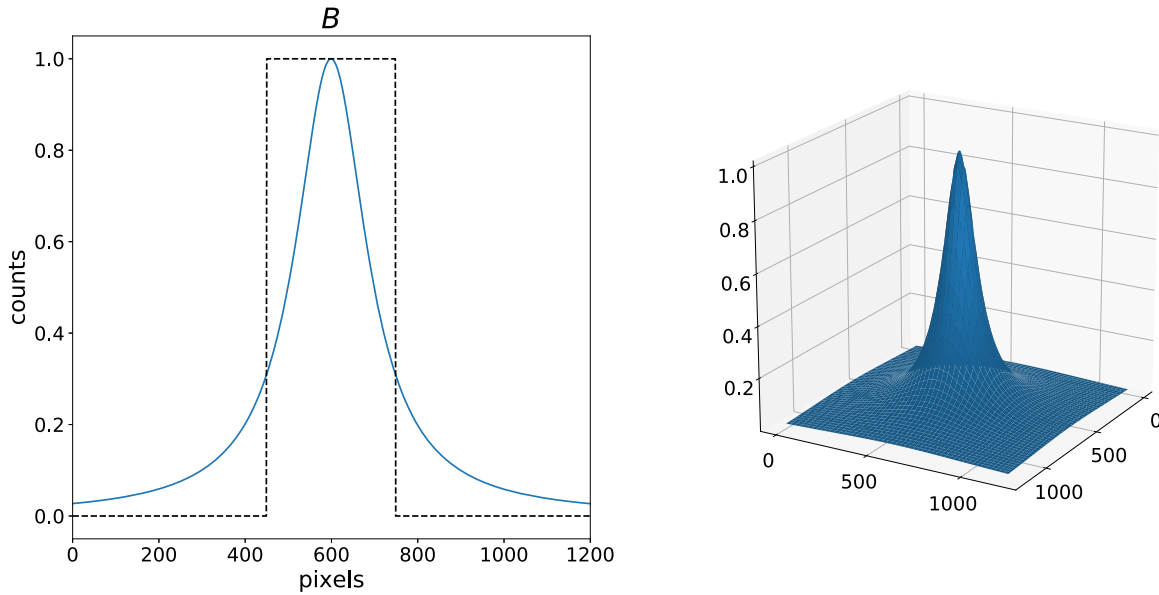
1. A fast Fourier transform is applied to an image ( $Im$ ). Then, we shift the zero-frequency component to the center of spectrum to be ready for the next step. This result is labeled as  $F(Im)$ .
2. A 3D Butterworth filter ( $B$ ) is applied to the shifted spectrum,  $F(Im) * B$ . The reason this filter is used is because (1) we need to filter out high frequency signals or noise, and (2) unlike a box-like low-pass filter, the transition of this Butterworth filter is smooth or curved so we prevent introducing extra ring-like signals. The Butterworth filter we used is given below and is plotted in Figure B1 in both 2D and 3D.

$$B_{\text{Butterworth}}[u, v] = \frac{1}{1 + (((u - u_0)^2 + (v - v_0)^2) / D_0^2)^n}$$

$$= \begin{cases} 1 & u = u_0 \text{ and } v = v_0 \\ 0.5 & (u - u_0)^2 + (v - v_0)^2 = D_0^2 \\ \text{else} & \end{cases}$$

The  $(u, v)$  is the frequency and  $(u_0, v_0)$  defines the center of the filter.  $D_0$  is the size of the filter and  $n$  is the order. Here we use optimal  $n = 1$  and  $D_0 = 100$  in this work to preserve most of the signals.





**Figure B1.** (Left) A 2D Butterworth filter is plotted as a blue line and a box-like low-pass filter is plotted as a dashed line. The first-order ( $n = 1$ ) Butterworth filter is shown here. (Right) A 3D Butterworth filter.

3. An inverse fast Fourier transform is applied to the spectrum given in the previous step, and is labeled as  $F^{-1}(F(Im) * B)$ .

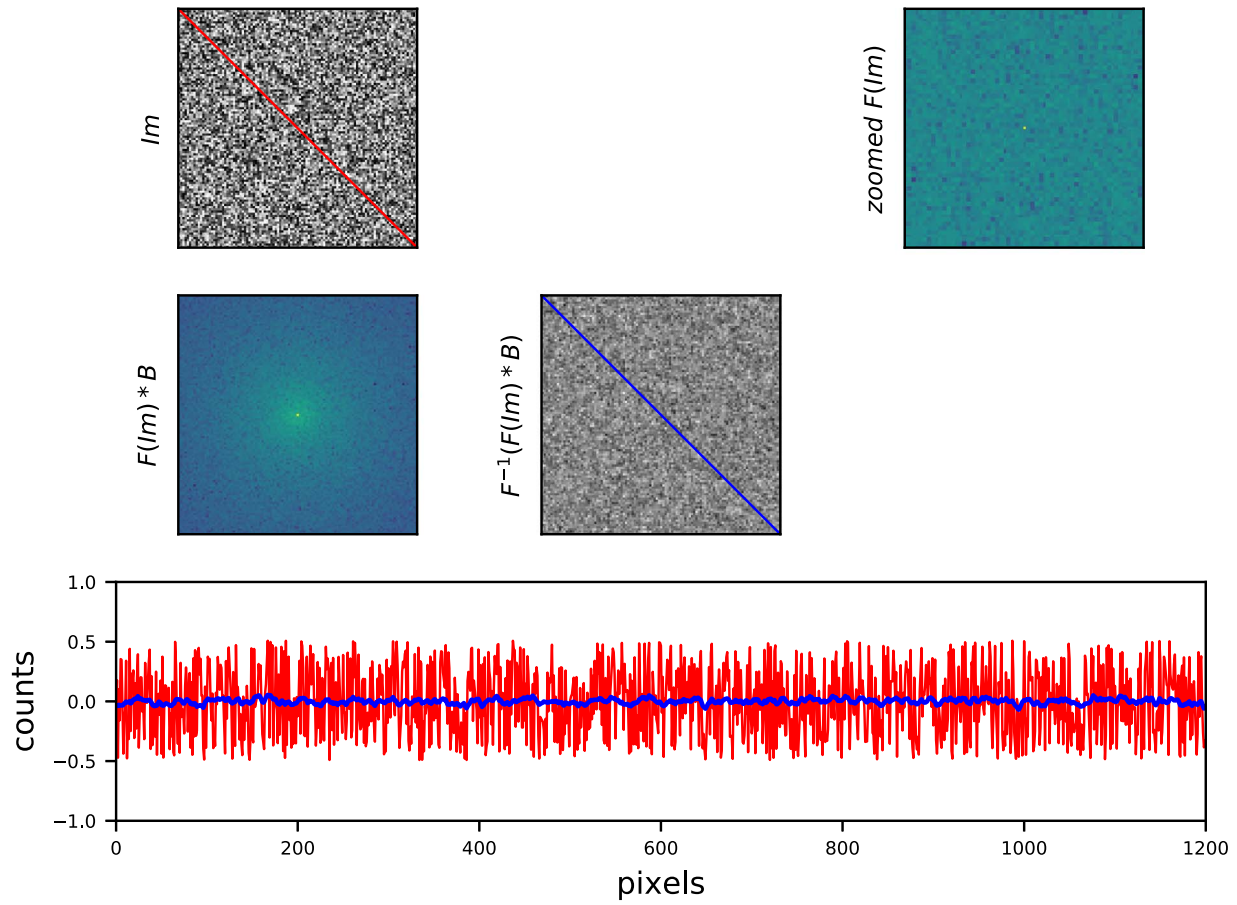
To test this analysis, we apply this routine to three different test images, a noisy image without any signal, a noisy image with a sinusoidal pattern, and an image of the M82 galaxy. Results are presented in Figures B2, B3 and B4, respectively. In Figure B2, a noisy image is preserved as a noisy image after this analysis, and no artificial signal is introduced as seen in the center plot. In the bottom of Figure B2, we highlight counts from a single slice of the image and compare it to the counts at the same location in the original image. It shows high frequency signals or noise have been filtered out.

Figure B3 shows this analysis is applied to an image with a sinusoidal pattern. This pattern can easily be identified in its original image. However, after Gaussian random noise is added, the pattern is embedded. After applying this Fourier analysis, this sinusoidal pattern is enhanced in a very noisy image as shown in the center of Figure B3. The pattern of vertical stripes is produced because noise is added vertically slice by slice and this vertical frequency can be seen as a horizontal line shown in the upper right plot of Figure B3.

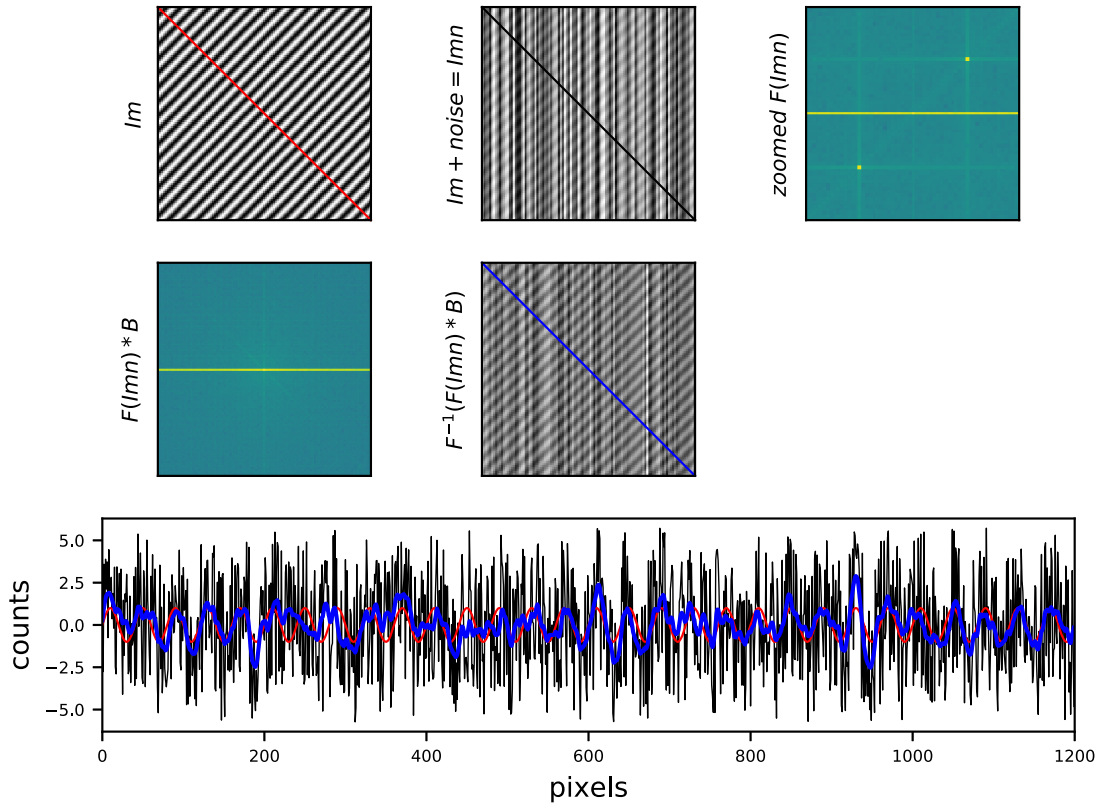
While applying the Butterworth filter, parts of the vertical patterns are kept, so these vertical stripes are still preserved in the inverse Fourier transform image, shown in the center of the plot.

Other than applying this analysis to these simulated images, we also use an M82 (NGC3034) image<sup>8</sup> as a test. The main reason an M82 image is selected is because after rotating the image, this galaxy resembles the HRD and it also has glowing hydrogen filaments resembling the gap in the center of the HRD. The step-by-step Fourier analysis result is plotted in Figure B4. It shows the majority of signals in this image are low frequency and are right in the center of the Fourier frequency domain after the fast Fourier transform. Visually, the original image ( $Im$ ) and the inverse fast Fourier transform image ( $IFIm$ ) are almost identical, but the  $IFIm$  is not as sharp as the  $Im$ . To compare these two images, we plot counts at the exact locations on these two images and compare them on the bottom of Figure B4. Like the analysis for an image with a sinusoidal pattern, overall features on the image are preserved and enhanced without introducing additional aliases. Consequently, based on example images discussed here, this analysis can enhance and preserve unseen patterns and signals on an image.

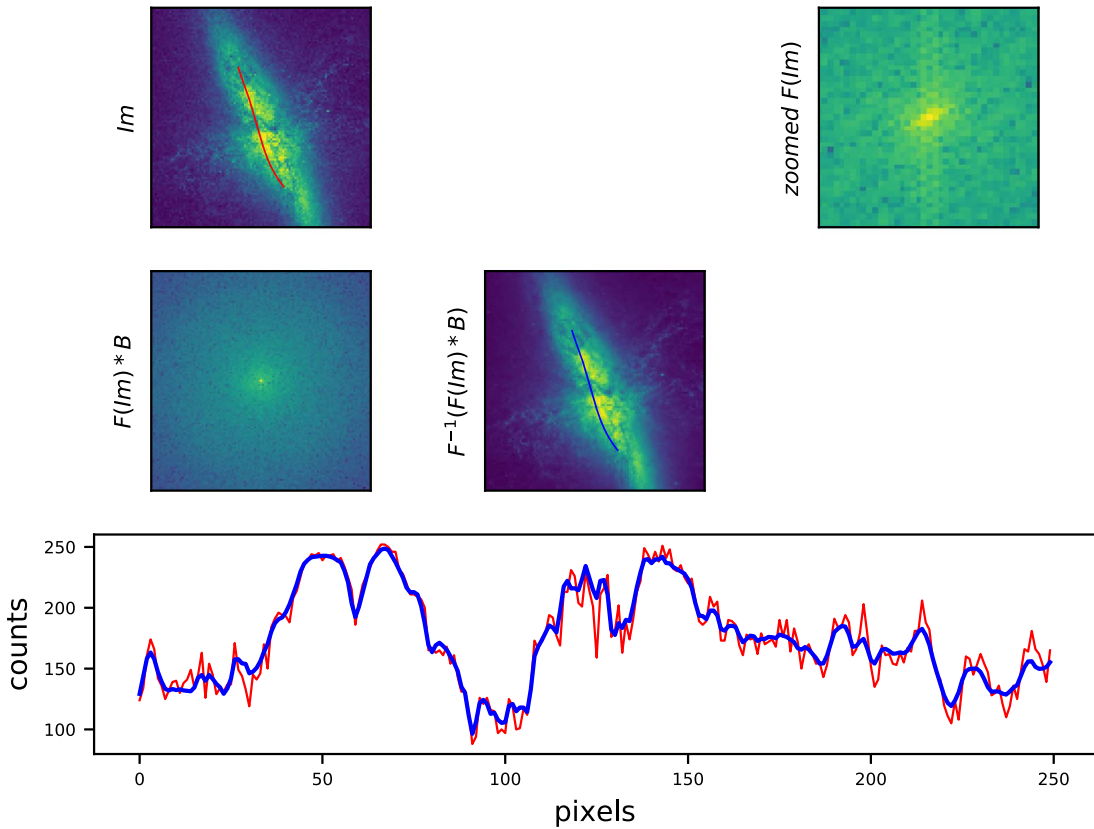
<sup>8</sup> The high-resolution M82 image is downloaded from <http://hubblesite.org/image/1876/gallery>.



**Figure B2.** (Top left) An original image. (Top center) This space is reserved for an image with added noise. We leave this space blank on purpose in this figure. (Top right) A 2D frequency or Fourier spectrum after we apply fast Fourier transform on this image. The zero frequency is at the center of the plot. The log scale of the amplitude is shown here. This plot only highlights the center portion of the spectrum where main signals are. A single point signal is seen in the center. (Middle left) A Fourier spectrum after we apply a Butterworth filter and the entire spectrum is shown. A log scale is shown as well. (Middle center) An inverse fast Fourier transform image from the spectrum shown to the left of this figure. The real part of the signals is shown. (Bottom) We compare counts along two lines plotted in the original image (red) and the enhanced image (blue). We label the relative math on the left side of each figure.



**Figure B3.** This figure is arranged and plotted in the same way as Figure B2. A sinusoidal pattern is seen on an image. Unlike Figure B2, the upper center subplot represents this image with extra noise to decrease the visibility of the pattern and we refer to this noisy image as  $Imn$ . A few signals can be identified in the spectrum shown in the upper right plot. The frequencies in the horizontal line indicate frequencies from the vertical stripes in  $Imn$ . The red, black, and blue lines diagonally trace the counts from three images,  $Im$ ,  $Imn$ , and  $F^{-1}(F(Imn) * B)$ , respectively.



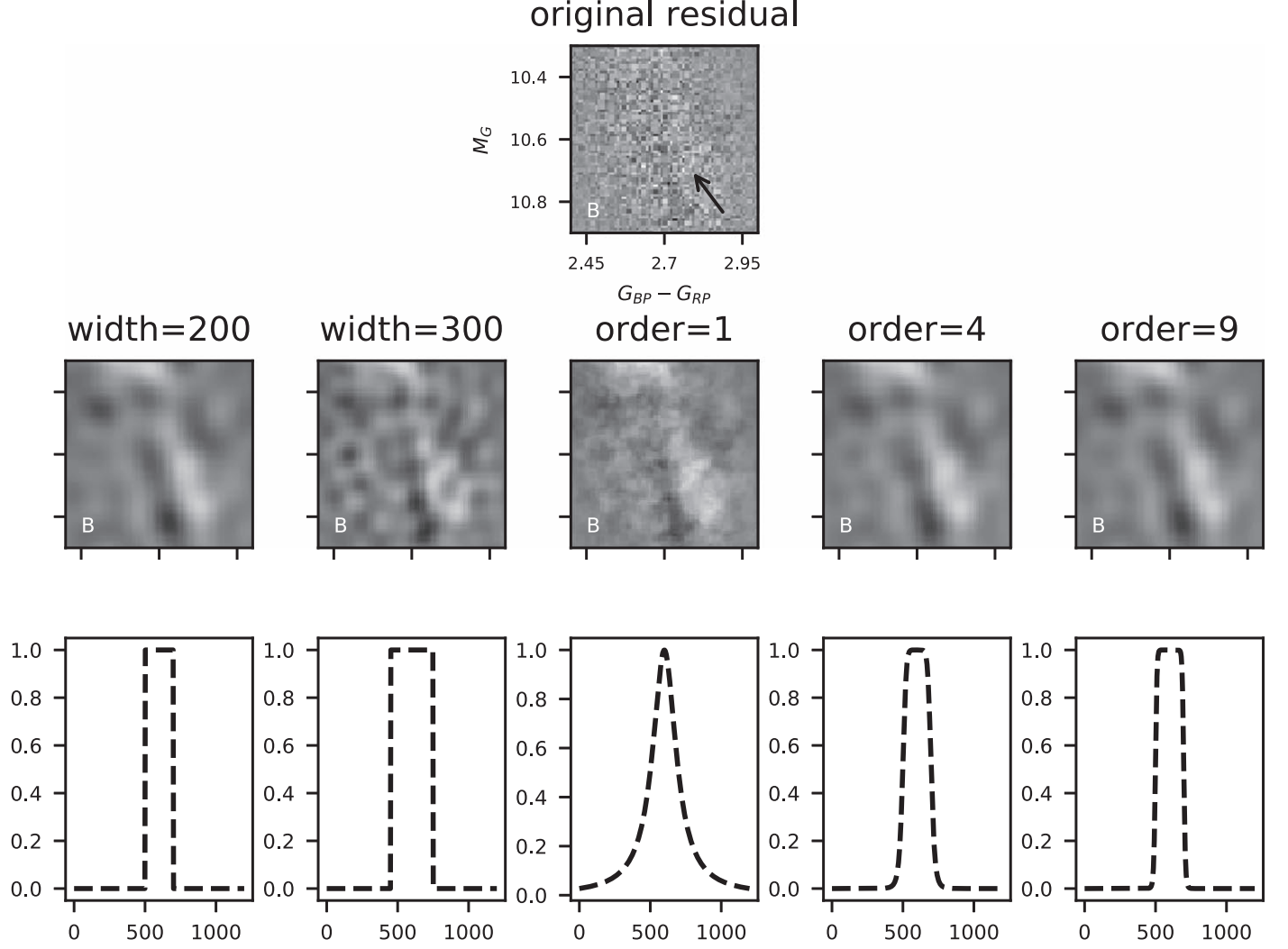
**Figure B4.** This figure for the M82 galaxy is arranged in the same way as Figure B2.

### Appendix C

#### Enhanced HRD Using Different Filters

Example results for this Fourier analysis from applying various filters are summarized in Figure C1 and it shows most filters produce consistent patterns. Other than the box-like filter

with a order of 300 producing extra small circular features, all other filters have enhanced the contrast on the residual plot. The Butterworth filter we used here has the first order, which generally preserves the constrasts.



**Figure C1.** Example results from various filters. The original residual plot at region B in Figure 3 is on the top. A single stripe feature is highlighted by an arrow. The second row contains enhanced HRD by using various filters shown on the bottom of this figure. The basic parameters for each filter are labeled. The two filters on the left are box-like filters. The three right filters are Butterworth filters with different orders, but all have  $D_0 = 100$ . All filters are 3D, but we only illustrate simplified 2D filters. The center filter (order = 1) is the one we used in this work.



### Appendix D

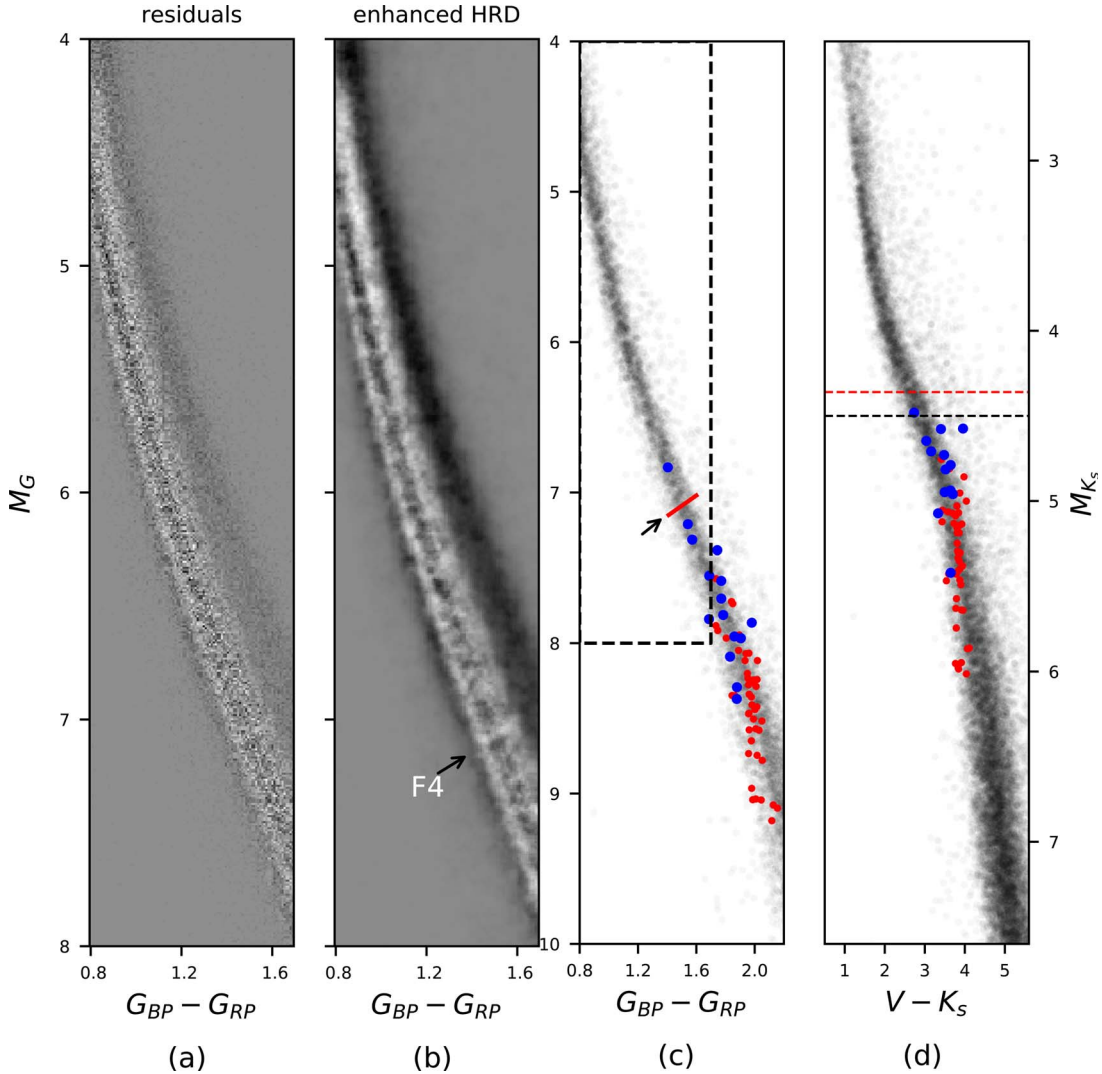
#### Main-sequence Features for Stars Bluer than $G_{BP} - G_{RP} = 1.7$

Our initial  $G_{BP} - G_{RP}$  color cutoff is set to be at 1.7, but later we extend the cutoff color to 0.8. The main sequence bluer than  $G_{BP} - G_{RP} = 0.8$  includes a significant amount of evolved and turnoff stars, so there are two types of stars at a given color. Hence, we arbitrarily use 0.8 as the new cutoff to explore features in this part of the main sequence. The enhanced HRD for this extended segment of the main sequence is shown in Figure D1. It is clear that similar patterns can be identified on the main sequence redder than  $G_{BP} - G_{RP} = 1.7$ , and the stellar density on the main sequence is not as random as expected.

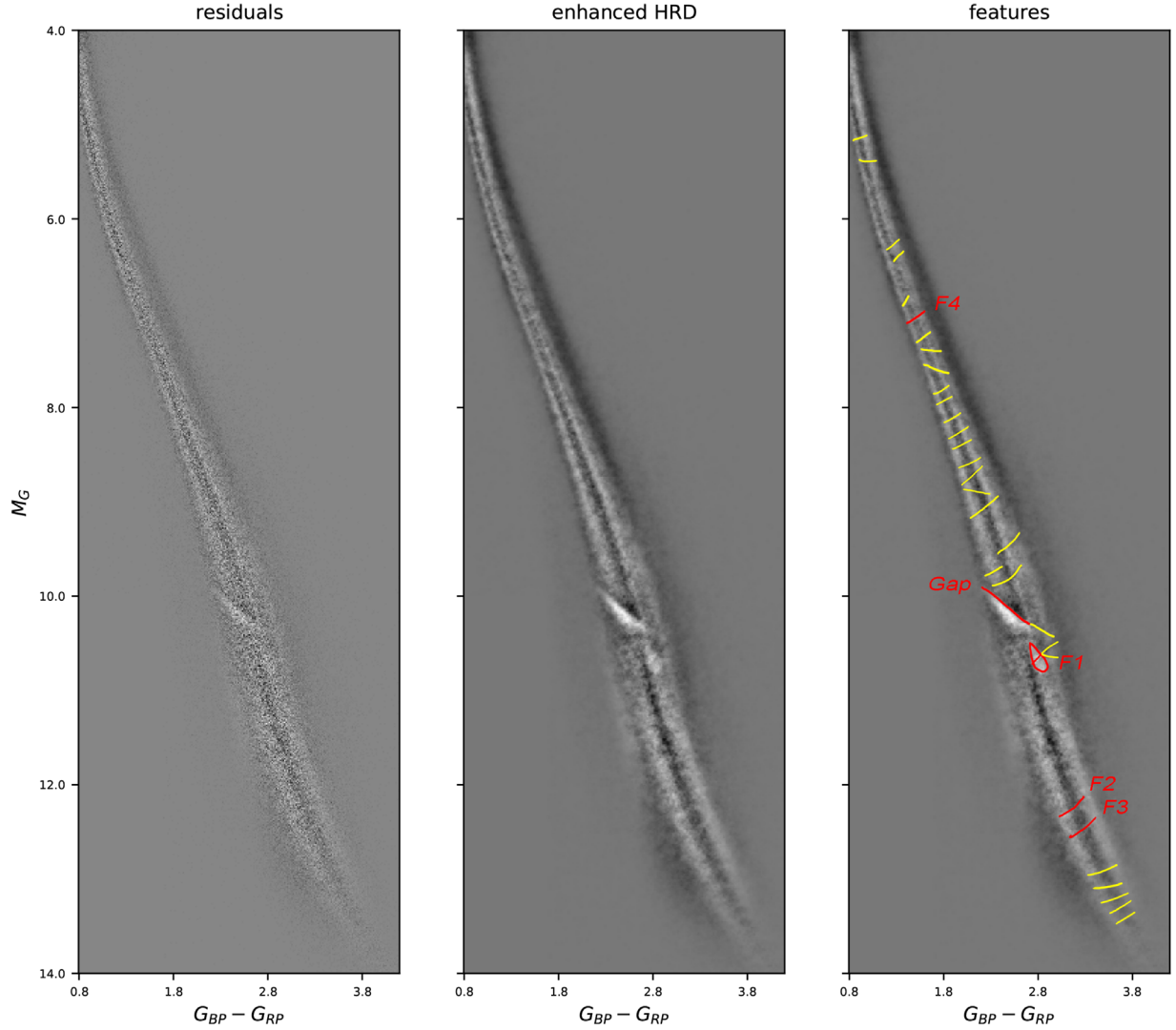
Like Figures 3, (1) the gray-black-gray bands, (2) young/double stars and (3) a band of excess subdwarfs below the main sequence are seen in Figure D1. One obvious new low density stripe-like feature in this region is labeled as “F4” at  $M_G \sim 7$ .

This feature is close to the transition from K to M dwarfs. In Figure D1, known late K8/9 and M0V dwarfs within 25pc are plotted as comparisons and it shows the F4 feature is above this empirical transition defined spectroscopically. To classify K and M dwarfs, the strengths of CaH and TiO bands are usually used as indicators (Boeshaar 1976; Kirkpatrick et al. 1991). The CaH bands are strong in K dwarfs, but TiO bands are prominent in M dwarfs. Generally one would use these band strengths and the slope of the spectrum to classify and assign proper subtypes by comparing to a set of spectral standard stars. Hence, results of the classification depend on the spectral coverages, resolution and the spectral standard stars. In the plots (c) and (d) of Figure D1, we can see that the transition from K to M is not sharp spectroscopically, and is overlapped in both  $G_{BP} - G_{RP}$  versus  $M_G$  and  $V - K_s$  versus  $M - K_s$  plots.

However, Casagrande et al. (2008) reported the main sequence has a discontinuity at K and M transitions as demonstrated in their Figure 12 using the  $V - K_s$  versus



**Figure D1.** (a) Residuals after fitting the skewed Gaussian distribution for stars bluer than  $G_{BP} - G_{RP} = 1.7$ . (b) The enhanced HRD is shown and stripe-like features can be seen. A prominent stripe-like feature is marked as “F4.” (c) Stars within 50pc in DR2 are plotted as gray points. Late K8/9 and M0 dwarfs within 25 pc are plotted as blue and red dots respectively. The dashed box shows the enhanced HRDs on the left. A red line indicates the F4 feature. (d) Gray points are the same 50 pc stars in plot (c), but are plotted in  $V - K_s$  and  $M - K_s$ . The blue and red dots represent the same K and M dwarfs in plot (c). A black dashed line indicates the location of discontinuity at  $M_{K_s} \sim 4.5$  in Casagrande et al. (2008). A red dashed line marks the F4 feature after converting  $M_G$  to  $M - K_s$  using the relation in Evans et al. (2018).



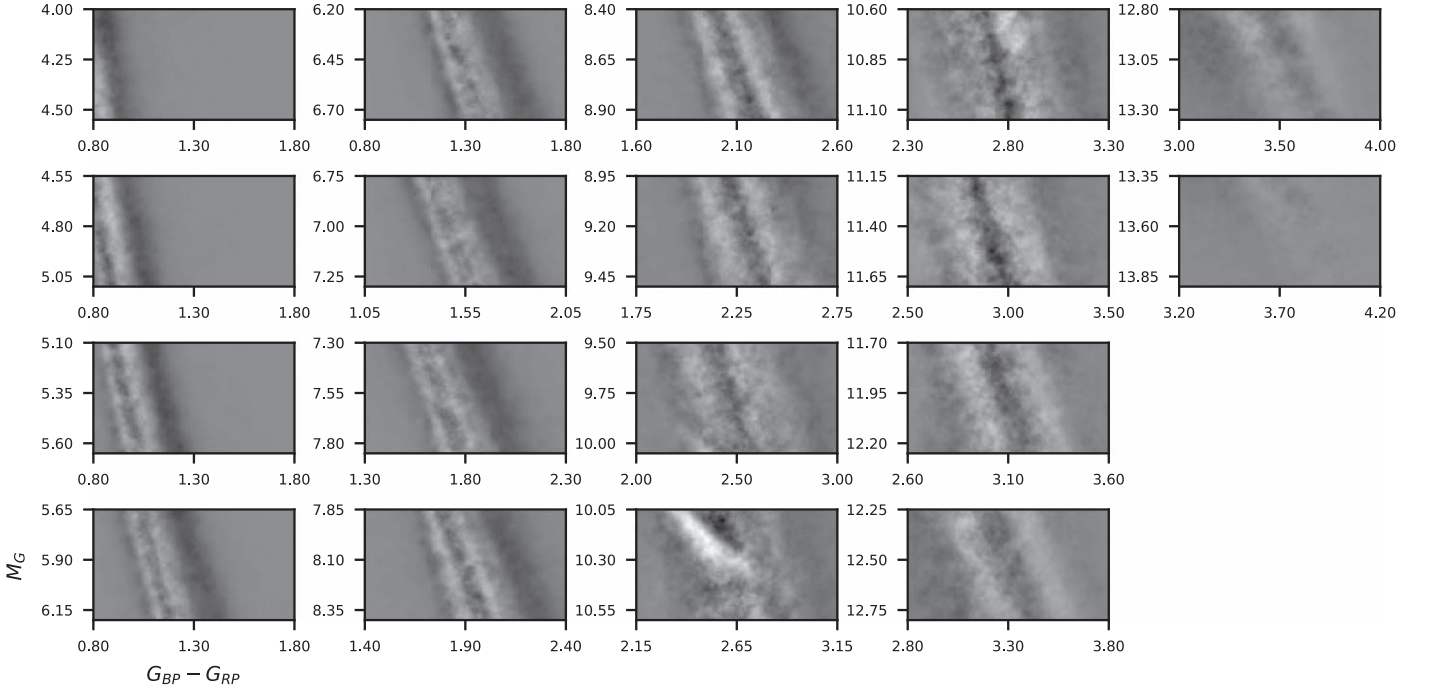
**Figure D2.** Residuals after fitting the skewed Gaussian distribution for all stars redder than  $G_{BP} - G_{RP} = 0.8$  are shown on the left figure. The enhanced HRD is shown at the center. Subjectively identified features are labeled and shown in the right figure. Prominent features discussed in this work are plotted in red. All speculated features are identified by eyes and plotted in yellow. Four prominent features, F1–F4, have approximate locations at (2.8, 10.7), (3.2, 12.25), (3.3, 12.5), and (1.5, 7.1), respectively. Note that these lines are hand drawn and they are for illustration only.

$M - K_s$  colors. Their main sequence appears to be horizontally shifted to be redder as the spectral type transits from K to M dwarfs at  $M_{K_s} \sim 4.5$ . They proposed molecular opacity could be the cause of this discontinuity, but suggested more works are needed.

In plot (d) of Figure D1, the main sequence plotted in  $V - K_s$  versus  $M - K_s$  colors is shown to match the scale in Casagrande et al. (2008), but the reported horizontal shift is not seen. However, the shift or discontinuity reported in Casagrande et al. (2008) is

very close to where F4 is after converting  $M_G$  to  $M - K_s$  using the relation reported in Evans et al. (2018). Yet, both are still brighter than the empirical transition using spectral classification, so the exact cause of this feature F4 is still to be learned.

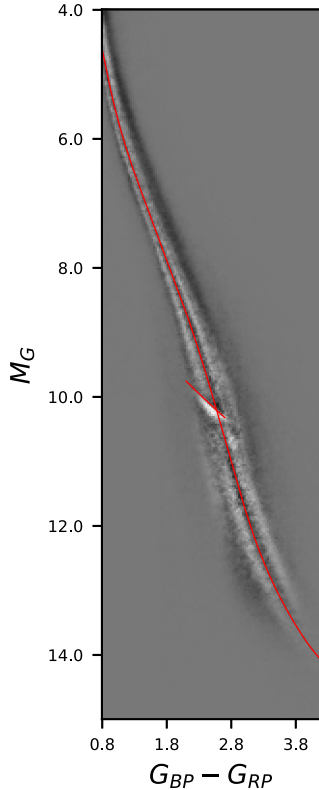
In order to get a broader view of the enhanced HRD, we merge the two segments of the main sequence in this work and present it as a whole in Figure D2. The zoomed plot of the enhanced HRD is given in Figure D3, so that fine features can be identified.



**Figure D3.** Zoomed regions of the enhanced HRD are shown, so fine features can be identified. Note that some features have been split into two plots.

### Appendix E Fitted Main Sequence and the Gap

This work also allows us to determine polynomial fits to represent the main sequence and the gap in  $G_{BP} - G_{RP}$  and  $M_G$ . The best-fitted line for the main sequence is generated



**Figure E1.** Two red lines represent fitted main sequence and the gap in the  $G_{BP} - G_{RP}$  and  $M_G$  domain.

**Table E1**

Coefficients of the Polynomials Used to Generated Fitted Main Sequence and the Gap

Coeff.	MS	Gap
a0	52.14339224	7.3626
a1	-301.8765068	1.2941
a2	778.85912205	-0.0735
a3	-1112.51951268	
a4	997.85111284	
a5	-593.08084754	
a6	238.68292406	
a7	-64.97687396	
a8	11.68974829	
a9	-1.31535951	
a10	0.08227295	
a11	-0.00210899	

**Note.** The relations are valid at  $0.8 \leq G_{BP} - G_{RP} \leq 4.0$  and  $2.1 \leq G_{BP} - G_{RP} \leq 2.7$  for MS and the gap, respectively. Both relations have the format of  $M_G = \sum_{n=0}^{11} a_n \times (G_{BP} - G_{RP})^n$ .

based on results of applying the skewed Gaussian function discussed in Section 2.

To generate the best-fitted line for the gap, we determine the slope of the residuals shown in Figure 2 at a given color first. The gap is then defined where the sign of a slope changes at a given color. Because the changing of slope is not obvious at  $G_{BP} - G_{RP} > 2.7$ , the fitted gap ends at  $G_{BP} - G_{RP} = 2.7$ . Both fitted lines are shown in Figure E1 and their coefficients are listed in Table E1.

### ORCID iDs

Wei-Chun Jao (饒惟君) <https://orcid.org/0000-0003-0193-2187>

Gregory A. Feiden <https://orcid.org/0000-0002-2012-7215>

## References

- Allard, F., & Hauschildt, P. H. 1995, [ApJ](#), **445**, 433
- Baraffe, I., & Chabrier, G. 2018, [A&A](#), **619**, A177
- Boeshaar, P. C. 1976, PhD Thesis, The Ohio State University
- Casagrande, L., Flynn, C., & Bessell, M. 2008, [MNRAS](#), **389**, 585
- Evans, D. W., Riello, M., de Angeli, F., et al. 2018, [A&A](#), **616**, A4
- Gagné, J., & Faherty, J. K. 2018, [ApJ](#), **862**, 138
- Gaia Collaboration, Babusiaux, C., van Leeuwen, F., et al. 2018, [A&A](#), **616**, A10
- Gaia Collaboration, Brown, A. G. A., Vallenari, A., et al. 2018, [A&A](#), **616**, A1
- Hertzsprung, E. 1911, [POPot](#), **22**, A1
- Hunter, J. D. 2007, [CSE](#), **9**, 90
- Iben, I., Jr. 1991, [ApJS](#), **76**, 55
- Jao, W.-C., Henry, T. J., Gies, D. R., & Hambly, N. C. 2018, [ApJL](#), **861**, L11
- Kilic, M., Hambly, N. C., Bergeron, P., et al. 2018, [MNRAS](#), **479**, L113
- Kirkpatrick, J. D., Henry, T. J., & McCarthy, D. W. 1991, [ApJS](#), **77**, 417
- Lindgren, L., Hernández, J., Bombrun, A., et al. 2018, [A&A](#), **616**, A2
- MacDonald, J., & Gizis, J. 2018, [MNRAS](#), **480**, 1711
- Pedregosa, F., Varoquaux, G., Gramfort, A., et al. 2011, [JMLR](#), **12**, 85
- Rosenberg, H. 1910, [AN](#), **186**, 71
- Russell, H. N. 1914, [Natur](#), **93**, 252
- Taylor, M. B. 2005, in ASP Conf. Ser. 347, Astronomical Data Analysis Software and Systems XIV, ed. P. Shopbell, M. Britton, & R. Ebert (San Francisco, CA: ASP), 29
- Valls-Gabaud, D. 2007, in IAU Symp. 240, Binary Stars as Critical Tools and Tests in Contemporary Astrophysics, ed. W. I. Hartkopf, E. F. Guinan, & P. Harmanec (Cambridge: Cambridge Univ. Press), 281
- van der Walt, S., Colbert, S. C., & Varoquaux, G. 2011, [CSE](#), **13**, 22
- van Saders, J. L., & Pinsonneault, M. H. 2012, [ApJ](#), **751**, 98
- Virtanen, P., Gommers, R., Oliphant, T. E., et al. 2020, [NatMe](#), **17**, 261