1. INTRODUCTION

Understanding planetary system formation and evolution is one of the major initiatives in astronomy. Stars form surrounded by protoplanetary disks of primordial gas and dust where planets grow. The material in these disks that does not fall into the star either collects into planets or is dissipated by processes such as photoevaporation (e.g., Clarke et al. 2001) and tidal forces either collects into planets or is dissipated by processes such as photoevaporation (e.g., Clarke et al. 2001) and tidal forces (e.g., Bryden et al. 1999), typically in less than 10 Gyr. Debris disks are our best current means of studying planet system evolution over its entire duration.

About 20% of the nearby stars harbor debris disks above current detection limits (Habing et al. 2001; Trilling et al. 2008; Carpenter et al. 2009; Gáspár et al. 2013; Eiroa et al. 2013). The likelihood of a detectable debris disk at 24 μm depends on age, with a higher percentage among young stars than older ones (e.g., Habing et al. 1999; Rieke et al. 2005; Gáspár et al. 2013). The expected timescale for the evolution of the disk components that dominate in the far infrared indicates that, depending on disk location and density, there might also be a detectable decay in disk incidence at 70 μm (Wyatt 2008). However, it has proven difficult to confirm this prediction definitively (Wyatt 2008; Bryden et al. 2006; Trilling et al. 2008; Carpenter et al. 2009). Gáspár et al. (2013) have recently demonstrated a drop in IR excess on the basis of comparison with the predictions of a theoretical model. If such a decay can be confirmed by an alternative analysis, it would help substantially to constrain models and particle properties of debris disks.

In this paper, we explore debris disk evolution in the far infrared (70–100 μm) using a large sample of stars to allow for reaching statistically robust conclusions. The paper presents high-quality homogeneous data reductions for a large sample of stars observed with Spitzer, analyzes their behavior, and then combines this sample with the Herschel-observed DEBRIS (Mathews et al. 2010) and DUNES (Eiroa et al. 2013) samples. We take care in determining stellar ages, since the apparent rate of decay can be strongly influenced by misclassification of the ages (a small number of young stars mistakenly identified as old ones might dominate the detections among the “old” sample). Section 2 describes the selection of the Spitzer sample, the determination of stellar ages, and the reduction of the IR measurements. We determined debris-disk-emitted excesses as described in Section 3. Section 4 presents the analysis of the behavior of these excesses with stellar age, and Section 5 merges this sample and additional Spitzer data with Herschel observations to study debris disk evolution in a total sample of about 470 stars. Our conclusions are summarized in Section 6.

2. SAMPLE SELECTION AND DATA REDUCTION

2.1. Selection Criteria

2.1.1. Photometric Criteria

We used specific criteria to draw samples of stars from the entire Spitzer Debris Disk Database. All stars were required to have Hipparcos data (van Leeuwen 2007), with parallax errors not to exceed 10% (the smallest parallax in the sample
All Sky Survey (2MASS) stars were also required to have observations on the Two Micron through M-type dwarf stars selected on the basis of probable youth from lithium abundance, K 148 78 Unbiased sample from 3 Myr to 3 Gyr in estimated age, equally distributed over this range, with 84 3 F, G, and K stars indicated to be members of Sco/ Cen based on Hipparcos 72 8 Young based on lithium, H α A3–F8 stars in binary systems, selected for proximity, binary characteristics, low IR background Trilling et al. (2007), Low et al. (2005) 80 10 A through M-type dwarf stars selected on the basis of probable youth from lithium abundance, X-ray emission, chromospheric activity, and rotation Pllavchan et al. (2009) 84 3 F, G, and K stars indicated to be members of Sco/ Cen based on Hipparcos Chen et al. (2005) 148 78 Unbiased sample from 3 Myr to 3 Gyr in estimated age, equally distributed over this range, with masses between 0.7 and 1.5 M⊙ (roughly F2V–K5V) Meyer et al. (2006) 206 1 Eight members of Castor Moving Group, no pre-selection for excesses Program abstract 713 1 Member of Trilling et al. sample (PID 41) observed in IOC Trilling et al. (2008) 2324 14 Potential TPF target stars, spectral types F0–M3.5, no companions within 100 pc, and terrestrial planet zone exceed 50 mas, generally within 25 pc Beichman et al. (2006) 3401 1 F-type stars with some previous indication of IR excess preferred; this star has a 60 μm IRAS detection Moir et al. (2011) 3600 3 Members of nearby young associations Zuckerman et al. (2011) 20065 1 Thirteen CoROT target stars Program abstract 20707 1 Selected either for IRAS detection or for membership in young associations; because the star in question has no IRAS excess (and no 70 μm excess), there is no bias 30211 5 Unbiased sample of F5–F9 stars older than 1 Gyr Trilling et al. (2008) 30339 2 Fifteen thick disk stars within 40 pc (volume limited) Sheehan et al. (2010) 30490 57 Volume limited sample within 25 pc Koerner et al. (2010) 40078 2 Stars selected on the basis of tight binarity (<3 AU), A3–F8 types Program abstract

The original programs in which our sample stars were measured are identified in Table 1. A large majority (93%) come from seven Spitzer programs: (1) the MIPS Guaranteed Time Observer (GTO) Sun-like star observations (Trilling et al. 2008); (2) Formation and Evolution of Planetary Systems (FEPS; Meyer et al. 2006); (3) Completing the Census of Debris Disks (Koerner et al. 2010); (4) potential Space Interferometry Mission/Terrestrial Planet Finder (SIM/TPF) targets (Beichman et al. 2006); (5) an unbiased sample of F-stars (Trilling et al. 2008); and (6) two coordinated programs selecting stars on the basis of indicators of youth (Low et al. 2005; Plavchan et al. 2009). All of these programs were based on unbiased samples selected in differing ways. The remaining 7% of the sample came from a broad range of programs, but, with the single exception of HD 206893, none made selections on the basis of prior detections. Another source of bias is the sources that were left off a selection because they were committed to guaranteed programs. This issue is most important for the MIPS solar-type program, and Bryden et al. (2006) list four stars that would otherwise have been in their sample. Of them, HD 10647 is in our sample and HD 17206 is in the DEBRIS sample and will be included when we merge with it. HD 10700 (Chen et al. 2006) does not appear to have an excess at the wavelengths of interest (so it would have little influence on our results), and
\(\epsilon\) Eri is not in our sample. To first order, it and HD 206893 offset biases, because they are both young stars with substantial excesses. However, these details demonstrate that any sample compiled from a collection of other samples defined on different bases are subject to selection biases.

We checked SIMBAD and the literature for indications of stars being members of multiple star systems. Such stars were only allowed in the final sample if they had a separation of less than 1 arcsec or greater than 12 arcsec. For typical distances for the sample members (\( \geq 8\) pc), 12 arcsec corresponds to a separation \( \geq 100\) AU and hence avoids the possible bias against debris disks for binary separations less than this value (Trilling et al. 2007; Rodríguez & Zuckerman 2012). Close binaries were allowed because they do not appear to suppress outer debris disks (emitting in the far infrared; Trilling et al. 2007). The original samples from which ours is drawn were not selected on the basis of rotational characteristics, although on the basis of gyrochronology we expect systematically higher \( \text{vsini} \) for the younger stars. After assembly of a preliminary sample, we carefully screened each star repeatedly for issues such as blending or complex backgrounds. The stars rejected in these tests were biased slightly toward being young but the distribution did not differ significantly from the age distribution of the entire sample (rejecting 10 in the 0–750 Myr range, 4 in 750 Myr–5 Gyr, and 4 in >5 Gyr). Finally, we considered whether metallicity could bias our results, given the dependence of detected planets on metallicity (Fischer & Valenti 2005; Johnson et al. 2010). However, a number of studies imply that the dependence of debris disk excesses on \([\text{Fe}/\text{H}]\) is weak (Trilling et al. 2008; Maldonado et al. 2012), and that the correlation between the presence of planets and debris disks, although present, is also weak (Bryden et al. 2009; Maldonado et al. 2012). Any bias due to metallicity effects can also be discounted because the average \([\text{Fe}/\text{H}]\) of the stars of ages <5 Gyr is \(-0.13\), whereas for those at ages >5 Gyr, it is virtually identical at \(-0.15\).

Given the requirements on spectral type and apparent magnitudes, the sample is almost entirely (95%) within 50 pc of the Sun. The stars beyond this distance are representative of the sample as a whole: in average metallicity, \([\text{Fe}/\text{H}] = -0.14\); in average spectral type, F9V; in age, seven 0–700 Myr, three 700 Myr–5 Gyr, and three >5 Gyr; and in incidence of excesses, 30%.

### 2.1.3. Ages

Accurately measured ages are essential to determine the decay of debris excesses. If the excesses decay substantially over the age range of the sample of stars, a few young stars incorrectly classified as being old can seriously affect the apparent rate of decay. To ensure the accuracy of our results, we took a conservative approach by using three different methods to estimate ages: (1) placement on the H-R diagram, (2) chromospheric activity indices (CAIs), and (3) X-ray luminosity. For the latter two indicators, we used the calibration of Mamajek & Hillenbrand (2008).

Each of these methods has weaknesses, making it desirable to use all three to verify the consistency of the results. Traditionally, using the H-R diagram can lead to relatively large errors in ages; the method is intrinsically inaccurate for stars near the main sequence, and it can also give incorrect results for reasons such as undetected stellar multiplicity. However, when used with care, the H-R diagram is useful for determining the ages of stars old enough to have left the main sequence. In comparison, CAI ages are well-calibrated for stars older than \(\sim 300\) Myr and younger than \(\sim 4\) Gyr (Mamajek & Hillenbrand 2008). For stars younger than 300 Myr, the issues are complex and will be discussed below. For stars older than 4 Gyr, the calibration of Mamajek & Hillenbrand (2008) rests on only 23 stars with accurate ages from placement on the H-R diagram. X-ray detections are also useful for young stars, but measurements of old stars are relatively uncommon. Because of their differing strengths and weaknesses, we tested the consistency of the indicators before proceeding with their use. For these tests, we drew stars from three samples studied extensively for debris disks; our combined Spitzer sample and those for the Herschel key programs DUNES and DEBRIS.

We compare the ages obtained via chromospheric activity with those from X-ray luminosity in Figure 1. The sources for the CAI, \(\log R'_{\text{HK}}\), can be found in Table 2 and in Gáspár et al. (2013). X-ray detections were taken from the ROSAT data (Voges et al. 1999, 2000), with the requirement of positional matches within 30 arcsec and \(-0.8 < HR1 < 0\); this constraint on the hardness ratio (HR1) was adopted to avoid active galactic nuclei (HR1 > 0.5), stars that are possibly flaring \((0 < \text{HR1} < 0.5)\), and background thermal sources (HR1 < \(-0.8\)) (T. Fleming, 2013 private communication). The trend line in Figure 1 (fitted by linear least squares with equal weight for all points) has a nominal slope of 1.28 ± 0.05, with a tendency for the points for the oldest stars to have larger ages from the CAI than from X-ray luminosity. This tendency could arise from a selection effect (Eddington bias). Only a minority of old stars are detected by ROSAT, and the limited ratios of signal-to-noise mean that the detected sample tends to include stars slightly overestimated in X-ray flux (due to statistics), leading to underestimated ages.

We tested this possibility on the stars in our sample with ages \(\geq 5\) Gyr based on the CAI. We used the CAI age estimates, the parallaxes, and photometric data to predict the ROSAT count rate for each star, which we compared with the ROSAT data. We evaluated the detection limits for each star from typical measurements for faint sources in the ROSAT Faint Source Catalog (Voges et al. 2000) in a field of 1 deg radius around the star. These detection limits and the predicted counts for the star were used to calculate the probability that it would have been above the detection threshold (Voges et al. 2000) for the survey. The sum of these probabilities over the entire sample of stars was 18 and is an estimate of how many detections are expected if the CAI ages agree perfectly with those from X-ray
emission. In fact, 27 stars could be identified with cataloged X-ray sources (Voges et al. 1999, 2000); however, three of them were so discordant from the count rate predictions for their ages and distances that the *ROSAT* detections are suspect and may arise from background sources. Therefore, we predicted 18 detections and found 24, indicating that the X-ray fluxes are slightly brighter than expected from the Magmajeck & Hellenbrand (2008) calibration. However, at the 1σ level, virtually no correction is indicated; in addition X-ray emission and the CAI are measures of closely related stellar parameters. We therefore used the Magmajeck & Hellenbrand (2008) calibration “as-is.” Nonetheless, it would be desirable to test the X-ray age calibration against a larger sample of old stars.

We next used the *Spitzer*–DEBRIS–DUNES sample of stars (our sample plus those in Gáspár et al. 2013) to compare CAI ages with those from the H-R diagram. We select V−K and MK for the axes of the H-R diagram, because the long wavelength baseline of V−K makes it an accurate indicator of stellar temperature (Masana et al. 2006) and the K band should be a relatively robust luminosity indicator, since it is in the Rayleigh–Jeans spectral regime for our sample of stars. To place the stars on the H-R diagram (Figure 2), we first eliminated from this expanded sample all cases with detected close companions (within 12′′) that would be included in or confuse the photometry of the star. The V−K color is quite sensitive to metallicity (Li et al. 2007); MV is also metallicity sensitive but this effect should be reduced for MK given the simpler spectrum across the K window. To calibrate the effects of metallicity on the H-R diagram, we determined a linear transformation that made the empirical metallicity dependent isochrones of An et al. (2007) coincide over the range −0.3 ≤ [Fe/H] < 0.2. We then reversed this transformation and applied it to each star so that stars of differing metallicity could be compared on a single H-R diagram:

\[
corrected M_K = M_K + 0.5 \left[ \frac{\text{Fe/H}}{\text{H}} \right]
\]

\[
corrected V - K = V - K - 0.25 \left[ \frac{\text{Fe/H}}{\text{H}} \right]
\]

Metallicity measurements were obtained from a number of sources (e.g., Marsakov & Shevelev 1995; Gray et al. 2003; Valenti & Fischer 2005; Gray et al. 2006; Ammons et al. 2006; Holmberg et al. 2009; Soubiran et al. 2010); when more than one measurement was available, they were averaged. Stars with [Fe/H] < −0.4 were eliminated in this comparison (but not in our final study if we had reliable age indicators) because our metallicity correction was not calibrated for them. We used the isochrones from the Padua group for the final comparisons (Bonatto et al. 2004; Marigo et al. 2008), because they have better coverage for older stars than those of An et al. (2007). An empirical correction of −0.14 mag was made to the absolute Ks magnitude on the isochrones to match the measured values for the main sequence; this correction removed systematic issues in matching the observational and theoretical photometric systems.

### Table 2

**Spitzer Star Sample**

| HD | STa | Age (Myr) | Ref.b | π (mas) | (Fe/H) (dex) | V (mag) | Ks (mag) | Ref.c | PID | AORKey | F24 (error) (mJy) | F70 (error) (mJy) |
|----|-----|-----------|-------|--------|-------------|--------|---------|-------|-----|--------|----------------|-----------------|
| 105 | G0V | 170 | 1, 2, 3, 4, 5, 6 | 25.4 | −0.23 | 7.51 | 6.12 | 1 | 148 | 5295872 | 27.82 (0.30) | 152.70 (9.71) |
| 377 | G2V | 220 | 1, 3, 7, 5, 6 | 25.6 | −0.10 | 7.59 | 6.12 | 1 | 148 | 5268992 | 36.31 (0.39) | 170.60 (10.62) |
| 870 | K0V | 2300 | 2, 8 | 49.5 | −0.20 | 7.22 | 5.38 | 1 | 30490 | 19827456 | 49.75 (0.51) | 22.51 (4.98) |
| 1466 | F9V | 200 | 2, 6 | 21.2 | −0.11 | 7.32 | 6.07 | 1 | 148 | 5272064 | 26.51 (0.29) | 7.78 (5.57) |
| 1237 | G8.5V | 300 | 1, 2, 6, 8 | 57.2 | −0.09 | 6.59 | 4.86 | 1 | 41 | 4060072 | 84.41 (0.85) | 11.81 (2.08) |
| 1461 | G0V | 7200 | 3, 5, 9 | 43.0 | 0.09 | 6.47 | 4.897 | 1 | 30490 | 19886336 | 80.19 (0.35) | 61.43 (6.08) |
| 1466 | F9V | 200 | 2, 6 | 21.2 | −0.27 | 7.46 | 6.15 | 1 | 72 | 4539648 | 34.47 (0.36) | 16.83 (3.09) |
| 2151 | G0V | 7000 | 2, 8, 15 | 174.1 | −0.03 | 2.82 | 1.286 | 3, 4, 5 | 30490 | 19759616 | 2225.00 (1.17) | 250.70 (5.26) |
| 4307 | G0V | 8100 | 3, 5, 8 | 32.3 | −0.23 | 6.15 | 4.622 | 1 | 41 | 4032512 | 103.00 (0.60) | 9.58 (3.21) |
| 4308 | G6Vf-09 | 5700 | 2, 6, 8 | 45.3 | −0.34 | 6.55 | 4.945 | 1 | 30490 | 19752192 | 75.79 (0.32) | 6.86 (4.67) |

**Notes.**

a Spectral type.

b Age references: (1) Schröder et al. 2009; (2) Henry et al. 1996; (3) Wright et al. 2004; (4) Rocha-Pinto & Maciel 1998; (5) Isaacson & Fischer 2010; (6) X-ray from RASS; (7) White et al. 2007; (8) Gray et al. 2006; (9) Gray et al. 2003; (10) Martinez-Arnáz et al. 2010; (11) Kasova & Livshits 2011; (12) Jenkins et al. 2006; (13) Zuckerman & Song 2004; (14) Monte et al. 2001; (15) log g < 4; (16) Duncan et al. 1991; (17) Buggino & Mauas 2008; (18) Rizzuto et al. 2011; (19) Barrado y Navascues 1998; (20) Karatas et al. 2005; (21) Schmitt & Lieflke 2004; (22) Ng & Bertelli 1998; (23) Mamajek 2009; (24) Mishenina et al. 2011.

c Sources of K photometry: (1) 2MASS; (2) this work; (3) Carter 1990; (4) Glass 1974; (5) Bouchet et al. 1991; (6) Johnson et al. 1966; (7) Selby et al. 1988; (8) Kidger & Martin-Ruiz 2003; (9) Alonso et al. 1998; (10) Aumann & Probst 1991; (11) Allen & Gallagher 1983; (12) UKIRT 1994.

(This table is available in its entirety in a machine-readable form in the online journal. A portion is shown here for guidance regarding its form and content.)

![Figure 2](image-url)
The Astrophysical Journal, 785:33 (13pp), 2014 April 10

The majority of stars with ages indicated from the CAI greater than 6 Gyr fell between the 6 and 11 Gyr isochrones, as shown in Figure 2. Figure 3 compares the ages; a few stars were eliminated from the sample based on this comparison because their ages from the H-R diagram were much younger than from the CAI. Because it is difficult for binarity or other issues to produce such discrepancies, the CAI ages for these stars are suspect. There are 38 stars in the sample in Figure 3 that are older than 4 Gyr; the trend line (fitted by linear least squares with equal weight for all points) has a slope of 0.985 ± 0.027, insignificantly different from 1, and there is no apparent deviation from the trend line with increasing age. We confirm that the CAI/X-ray calibration of Mamajek & Hillenbrand (2008) is consistent with isochrone ages for stars older than 4 Gyr.

With this validation, we have used chromospheric activity and X-ray luminosity as our primary age indicators. We supplemented these indicators with log g < 4—which, for stars over the relevant spectral range (F4–K2), shows that a star is above the main sequence—and with gyrochronology ages when available. We then checked that the age was consistent with the placement of the star on the H-R diagram. Stars where the initial criteria indicated ages >5 Gyr, but which fell in an inappropriate region on the H-R diagram, were not used further. The final Spitzer sample is presented in Table 2, with the V and Ks measurements, ages, and Spitzer measurements at 24 μm and 70 μm. The age distribution of the sample is shown in Figure 4; there are 255 stars with age estimates, of which 238 have useful 70 μm measurements.

We now discuss the uncertainties in the ages. Taking ages directly from Mamajek & Hillenbrand (2008) would be circular, as we use the same calibration and some of the same data sources. However, comparing with those ages is revealing; there is a scatter of 0.15 dex. These differences primarily arise because of data published since the publication of Mamajek & Hillenbrand (2008) and they illustrate that there can be significant uncertainties despite the care we have taken in estimating stellar ages. Another issue is that there is a large spread in chromospheric activity around 50–300 Myr (Mamajek & Hillenbrand 2008). As a result, a star with a low level of activity may still be within this age range. Based on Figure 5 of Mamajek & Hillenbrand (2008), it is possible that about 20% of the stars in the 50–300 Myr range are incorrectly placed in the 0.6–2.5 Gyr range, if the age is assigned purely on the basis of the CAI. This rate of incorrect age assignment is an upper limit because we use multiple indicators to check against each other. More importantly, we find that the incidence of debris disk excesses does not vary rapidly across the 0.1–2.5 Gyr age range (see Section 5.2), so a modest number of age errors within this range will have little effect on our results.

2.2. Data Reduction

2.2.1. 24 μm

We re-processed all the mid- and far-infrared data for these stars as part of a Spitzer legacy catalog, using the MIPS instrument team Data Analysis Tool (Gordon et al. 2005) for the initial reduction steps. In addition, a second flat field constructed from the 24 μm data itself was applied to all the 24 μm results to remove scattered light gradients and dark latency (Engelbracht et al. 2007). The processed data were then combined using the World Coordinate System information to produce final mosaics with pixels half the size of the physical pixel scale (which is 2′.49). We extracted the photometry using point-spread function (PSF) fitting. The input PSFs were constructed using isolated calibration stars and have been tested to ensure that the photometry results are consistent with the MIPS calibration (Engelbracht et al. 2007). Aperture photometry was also performed, but the results were only used as a reference to screen targets that might have contamination from nearby sources or background nebulosity.

The random photometry errors were estimated based on the pixel-to-pixel variation within a 2 arcmin square box centered on the source position. The measured flux densities (using 7.17 Jy as the zero magnitude flux; Rieke et al. 2008), and associated random errors are listed in Table 2. The errors from the photometry repeatability (~1% at 24 μm; Engelbracht et al. 2007) are included, which were measured on stars of similar brightness to our sample members (and hence give an estimate of photon noise as well as other systematic photometric errors). For the stars with indicated 24 μm excesses, we have compared with the WISE W3–W4 color. In all cases, where the WISE W4 measurement had adequate signal to noise, the MIPS measurement was confirmed.3 At least two of the stars

---

3 The WISE data for HD 25457 show multiple sources and appear to be confused; the MIPS data are valid.
in our sample have strong excesses at 24 μm and weak ones at 70 μm, placing them in the rare category of predominantly warm debris disks (HD 1466, for which the weak far-infrared excess is confirmed with Herschel, Donaldson et al. 2012; and HD 13246, Moór et al. 2009).

2.2.2. 70 μm

The 70 μm observations were obtained in the default pixel scale (9.85). After initial reprocessing, artifacts were removed from the individual 70 μm exposures as described in Gordon et al. (2007). The exposures were then combined with pixels half the size of the physical pixel scale. The image of each source was examined to be sure that it was free of confusing objects, such as IR cirrus emission or background sources. A number of objects were rejected at this stage. PSF photometry was used to extract flux measurements for the “clean” sources, utilizing STinyTim PSFs that had been smoothed to match the observations (accounting primarily for the effects of the pixel sampling) and with the position fixed to that found at 24 μm. The results were calibrated as in Gordon et al. (2007). The random photometry errors were based on pixel-to-pixel variations in a 2 arcmin field surrounding the source and are tabulated (including a 5% allowance for photometric non-repeatability Gordon et al. 2007) along with the flux densities in Table 2.

3. DETERMINATION OF EXCESSES

3.1. 24 μm Excesses

Excesses at 24 μm were detected using a V−K versus K−[24] plot (Figure 5). The locus of stars without excesses was determined from a total of ~1320 stars drawn from the Spitzer archive, reduced identically to the stars in our sample, and with K magnitudes from the same sources. The resulting fit is (Urban et al. 2012):

\[
K_S - [24] = 0.0055 - 0.0134x + 0.06555x^2 - 0.1095x^3 \\
+ 0.066415x^4 - 0.01458x^5 + 0.001101x^6 - 0.000003x^7
\]

for \(x > 0\), where \(x = V - K_S - 0.8\). The scatter around this fit was determined for all 1320 stars by fitting a Gaussian to the residuals. Because a significant number of stars have excesses, the fitting parameters needed to be selected to avoid a bias. We first used outlier rejection based on Huber’s Psi Function, set just to avoid rejecting any points below the photospheric value. We used the Anderson–Darling (A-D) test to confirm that the remaining distribution around the photospheric values was consistent with a normal distribution. Inspection of the results indicated that there were a number of stars with excesses (the positive side of the distribution above 0.03 in residual had 127 stars, compared with 96 for the negative side of the distribution below −0.03), so before calculating the A-D p value, we replaced the measures above 0.03 with the inverse of the ones below −0.03. The A-D test then indicated that the distribution around the photospheric value was indistinguishable from a Gaussian (\(p = 0.14\)). We therefore fitted a Gaussian to these points, finding a standard deviation of 0.0304.

To provide a more intuitive check of this approach, we carried out a second analysis where we fitted a Gaussian excluding values more than 2σ above the photospheric trend line (i.e., above 0.06). As with use of the Psi Function, this strategy makes the fit independent of IR excesses detected in some of the stars of the sample but is more subjective than the outlier rejection approach. A demonstration of this result can be found in Ballering et al. (2013), which demonstrates the excellent fit of a Gaussian around the photospheric value, as well as showing the population of excesses. In the current case, this Gaussian indicated a standard deviation of 0.0295. That is, the two Gaussian fits to the peak of the measurement distribution around the photospheric value agree closely in width. We adopt a value of \(\sigma = 0.03\). Therefore, any star with a \(K_S - [24] > 0.09\) mag larger than predicted for its photosphere is considered to have an excess at 24 μm (nominally, the number of false identifications in our sample is then expected to be less than 0.5). None of the stars older than 1000 Myr have excesses based on this requirement; 23 of the younger stars do, a fraction of 0.136 ± 0.028 (errors from binomial statistics). Stars with excesses are listed in Table 3.

3.2. 70 μm Excesses

A method similar to that described in Trilling et al. (2008) was used to determine which stars had excesses at 70 μm. We computed an excess ratio, \(R^4\) for each star by predicting a photospheric flux density based on the measured flux density at 24 μm divided by 9.05 (the nominal ratio of flux densities between 24 and 70 μm) and then divided by the ratio of the measurement to the predicted photospheric value at 24 μm, if this ratio exceeded 1. We used the figure of merit:

\[
\chi_{70} = (F_{70} - P_{70})/\sigma
\]

(5)

to quantify the significance of an excess, where \(F_{70}\) is the measured flux density at 70 μm, \(P_{70}\) is the predicted photospheric flux density, and \(\sigma\) is the uncertainty in the flux density measurement at 70 μm (the uncertainty in \(P_{70}\) is negligible compared with that in \(F_{70}\) for our entire sample). In total 38 stars younger than 5 Gyr and 4 stars older than this age are considered to have excesses, as illustrated in Figure 6 and listed in Table 3. For the young sample, the probability of having an excess (at 70 μm)
The Astrophysical Journal, 785:33 (13pp), 2014 April 10

Sierchio et al.

Table 3
Stars with Excesses

| HD     | Excess? | $K_S - [24]$ | $R_{70}$ | $\chi_{70}$ |
|--------|---------|--------------|----------|-------------|
| 105    | 2       | 0.09         | 54.47    | 24.99       |
| 377    | 3       | 0.38         | 80.87    | 26.40       |
| 870    | 2       | -0.01        | 4.10     | 3.55        |
| 1461   | 2       | 0.02         | 7.10     | 8.65        |
| 1466   | 3       | 0.35         | 5.46     | 5.30        |
| 6963   | 2       | 0.08         | 12.72    | 8.31        |
| 7193   | 1       | 0.11         | 3.04     | 1.76        |
| 7590   | 2       | 0.04         | 35.42    | 38.02       |
| 8907   | 2       | 0.07         | 52.55    | 49.35       |
| 10647  | 3       | 0.30         | 76.26    | 73.02       |
| 13246  | 1       | 0.72         | 0.12     | -0.65       |
| 14082  | 1       | 0.11         | 0.61     | -0.33       |
| 19668  | 1       | 0.23         | 0.18     | -0.26       |
| 20794  | 2       | 0.05         | 1.34     | 4.86        |
| 20807  | 2       | 0.05         | 1.69     | 4.53        |
| 25457  | 3       | 0.34         | 18.39    | 16.90       |
| 30495  | 2       | 0.08         | 6.08     | 25.69       |
| 31392  | 2       | 0.04         | 18.39    | 16.90       |
| 33262  | 2       | 0.08         | 1.97     | 5.86        |
| 33636  | 2       | 0.00         | 8.20     | 21.80       |
| 35114  | 1       | 0.28         | 5.07     | 1.97        |
| 38529  | 2       | 0.02         | 3.61     | 6.15        |
| 43989  | 1       | 0.20         | 6.55     | 1.63        |
| 5853   | 3       | 0.84         | 483.88   | 102.78      |
| 73350  | 3       | 0.10         | 20.31    | 25.02       |
| 76151  | 2       | 0.04         | 2.53     | 5.99        |
| 85301  | 3       | 0.36         | 15.05    | 7.35        |
| 90905  | 1       | 0.11         | 4.68     | 2.52        |
| 104860 | 2       | 0.09         | 95.37    | 47.48       |
| 106906 | 3       | 2.14         | 230.90   | 30.01       |
| 107146 | 3       | 0.31         | 157.49   | 91.11       |
| 114613 | 2       | 0.02         | 1.43     | 3.52        |
| 119124 | 2       | 0.04         | 8.49     | 22.61       |
| 122652 | 2       | 0.10         | 23.24    | 16.93       |
| 134319 | 1       | 0.12         | 4.38     | 1.10        |
| 145229 | 2       | 0.09         | 21.16    | 14.26       |
| 153597 | 2       | -0.02        | 1.59     | 3.31        |
| 181327 | 3       | 2.12         | 662.75   | 52.66       |
| 183216 | 3       | 0.11         | 6.99     | 3.58        |
| 191089 | 2       | 2.04         | 267.90   | 82.14       |
| 199260 | 2       | 0.05         | 3.54     | 14.36       |
| 201219 | 2       | 0.08         | 18.49    | 8.73        |
| 202628 | 2       | 0.05         | 19.69    | 15.60       |
| 202917 | 3       | 0.46         | 46.44    | 14.92       |
| 204277 | 2       | 0.03         | 5.41     | 4.57        |
| 206374 | 2       | 0.00         | 5.34     | 3.93        |
| 206860 | 2       | 0.06         | 2.17     | 4.95        |
| 206893 | 2       | 0.06         | 56.49    | 30.30       |
| 207129 | 2       | 0.04         | 18.82    | 37.40       |
| 207575 | 1       | 0.11         | 3.75     | 2.59        |
| 218340 | 2       | 0.03         | 12.32    | 5.38        |
| HIP6276| 1       | 0.13         | 1.88     | 0.24        |

Notes.

1 $=$ excess at 24 $\mu$m only (or no 70 $\mu$m measurements), 2 $=$ excess at 70 $\mu$m only, 3 $=$ excesses at both 24 $\mu$m and 70 $\mu$m.

$^b$ 24 $\mu$m excess confirmed by WISE W3–W4 color.

is 38/169 = 0.225 ± 0.036. Similarly for the old sample, the probability of an excess is 4/86 = 0.047$^{+0.037}_{-0.020}$ (all errors are calculated from the “exact” or Clopper–Pearson method, based on binomial statistics). Three of the four detections around old stars are at $\chi_{70} > 4.5$ and should be very reliable (see Figure 6). These three systems are similar in amount of excess to typical younger stars; the ratios of the fluxes at 70 $\mu$m to the photospheric fluxes are close to the median for our entire sample.

3.3. Stars Older Than 2.5 Gyr with 70 $\mu$m Excesses

The major conclusion from this paper will rest on the small number of old stars with excesses. The reliability of this number is influenced both by the reliability of the identification of excesses, and by the accuracy of our age estimates. With regard to the reliability of the excesses, we have compared our sample with the results from the Herschel DUNES and DEBRIS surveys (as summarized in Gaspár et al. 2013 and Eiroa et al. 2013). There are 50 stars in our sample also in the Herschel samples. Using the same $\sigma$ selection criterion, 41 of the 50 stars show no evidence for excess emission with either telescope.5 Of the remaining nine stars, six show significant excesses in the data from both telescopes, (HD 20794, 30495, 33262, 76151, 199260, and 207129) and one more (HD 153597) has an excess close to 3$\sigma$ in both sets of data. The remaining two stars are HD 30652 and 117176. The first case is just right on the photospheric value, while the PACS measurement (at 70 $\mu$m) is 3.3. Stars Older Than 2.5 Gyr with 70 $\mu$m Excesses

The implied very cold spectral energy distribution for the excess suggests that it could arise from a background galaxy (Gaspár & Rieke 2014), so we have listed this star as not having an excess.

Because of the residual uncertainties in age, we will discuss the stars with excesses and ages indicated to be greater than 2.5 Gyr (because we will show that the incidence of excesses drops substantially between 3 and 5 Gyr). At the same time, we tabulate their spectral types to probe whether they are characteristic of the full sample. We begin with the stars indicated to have ages between 2.5 and 5 Gyr—HD 20807,
G0V: the placement on the H-R diagram indicates that this star is less than 3.5 Gyr old. As indicated in Table 1, multiple sources indicate chromospheric activity levels too high for the star to be 5 Gyr or older. HD 33636, G0V: the placement on the H-R diagram indicates an age <3.5 Gyr. Isaacson & Fischer (2010) report 17 measurements of chromospheric activity that are, with a single exception, consistent with the quoted age; the age should be robust against measurement errors and variations in activity. HD 207129, G2V: there are suggestions that we have overestimated the age of this star (Montes et al. 2001; Mamajek & Hillenbrand 2008; Tetzlaff et al. 2011), but no indication that it could be older than 5 Gyr. We now discuss the stars older than 5 Gyr and with 70 μm excesses— HD 1461, G0V; two measurements (Wright et al. 2004; Gray et al. 2003) agree that the chromospheric activity of this star is very low, corresponding to an age of 7–8 Gyr. The placement on the H-R diagram indicates a lower limit of 6 Gyr. HD 20794, G8V: three measurements (Henry et al. 1996; Gray et al. 2006; Schröder et al. 2009) agree that this star has low chromospheric activity, indicating an age of ∼7 Gyr. The position of the star on the H-R diagram is inconclusive with regard to age. HD 38529, G4V: multiple measurements (Wright et al. 2004; White et al. 2007; Schröder et al. 2009; Isaacson & Fischer 2010; Kasova & Livshits 2011) agree on the low chromospheric activity of this star, corresponding to an age of about 6 Gyr. The value of the CAI reported by Isaacson & Fischer (2010) is based on more than 100 measurements. The surface gravity also places this star well off the main sequence. HD 114613, G3-4 V: four measurements (Henry et al. 1996; Jenkins et al. 2006; Gray et al. 2006; Isaacson & Fischer 2010) agree that this star has a very low level of chromospheric activity, corresponding to an age of about 8 Gyr. The value from Isaacson & Fischer (2010) is based on 28 individual measurements. The placement of the star on the H-R diagram indicates an age >5 Gyr; it also has low surface gravity, consistent with its position well above the main sequence.

4. ANALYSIS

In this section, we first show that the rate of excess detection is independent of spectral type, confirming the result from Gáspár et al. (2013). As a result, to obtain good statistics, we can consider the evolution of the debris excesses combining the results for all spectral types without loss of information. We then analyze the time evolution of the excesses for the Spitzer sample. In the following section, we first show that the stars measured in the DEBRIS and DUNES programs behave consistently with the Spitzer sample and then combine all the data to double the number of stars. This larger sample shows that the incidence of far-infrared excesses evolves substantially with stellar age (see also Gáspár et al. 2013).

4.1. Detection Statistics

Because our samples span late F-, G-, and early K-spectral types, we can report excess fractions for each type. To compute these fractions, we only considered stars less than 5 Gyr old, because stellar evolution would otherwise be an issue for the F stars. Similarly, we only considered excesses at 70 μm, as the 24 μm excesses evolve over an age range of a Gyr. We found fractions with excesses of 16/63, 20/98, and 2/8, respectively, for F4–F9, G, and K stars. We combined our results with those from DEBRIS and DUNES (Gáspár et al. 2013; Eiroa et al. 2013), using the same definition for detection of an excess (χ > 3). We omitted HIP 171, 29271, and 49908 from this combined sample because their 160 μm excesses are likely to arise through confusion with background galaxies (Gáspár & Rieke 2014); we also omitted HIP 40843, 82860, and 98959 because the reported excesses at 100 μm (Gáspár et al. 2013) may also be contaminated (Eiroa et al. 2013). The net results are 23/105, 22/133, and 10/69, respectively, for F4–F9, G, and K stars. The corresponding fractions are 0.219+0.048−0.043, 0.165+0.039−0.033, and 0.145+0.055−0.043, respectively. If the 10 stars later than K4 are excluded (all from the DEBRIS and DUNES samples and none of which have excesses), the fraction for K stars rises to 0.169+0.063−0.060; the errors are 1σ, computed using the “exact” or Clopper–Pearson method, based on the binomial theorem. Although there may be a weak trend with spectral type, the incidence of excesses is (within the errors) the same for late F through early K-type stars. This conclusion is strengthened if we take into account the possible bias of including HD 206893 (F5V) and excluding ε Eri (K2V) in our sample, as discussed in Section 2.1.2. The results of previous studies are mixed, but our result is in general agreement with Trilling et al. (2008) who reported no clear trend of IR excess detection with stellar type.

4.2. 70 μm Excess Evolution: K-M Test

Previous studies have attempted to quantify the decay rate for debris disks by one of two different approaches (Rieke et al. 2005; Trilling et al. 2008): counting the number of stars with detected excesses (i.e., basing the statistics on small excesses) or fitting the largest excesses (i.e., matching the apparent upper envelope of the excesses). Neither of these approaches takes rigorous account of all the data and each is subject to selection effects. In the first case, if the achieved sensitivity is not adequate to detect the stellar photospheres uniformly, the minimum detectable excess is not uniform, whereas in the second case the upper envelope of the excesses is ill defined and based on small number statistics. Ultimately, these studies found either no or only weak evidence of disk evolution with age. The underlying problem can be described in terms of data censoring, that is, the non-uniform detection limits result in target stars being removed from the study at the levels of excess represented by upper limits.

A standard statistical approach to such a situation is to use the Kaplan–Meier (K-M) estimator, because it naturally allows for censored data (Feigelson & Babu 2012). We used the version of the K-M estimator provided by Lowry (2013), which bases its confidence level estimates on the efficient score method (Newcombe 1998). We divide the sample into three equal parts corresponding to the age ranges <750 Myr, 750–5000 Myr, and >5000 Myr. Based on the detection limits and identifications of excesses in Table 2, we find that the probable numbers of excess ratios R > 1.5 at 70 μm are 27% with 95% confidence limits of 18% and 38% for stars of age <750 Myr and 22% with confidence limits of 14% and 33% for those between 750 and 5000 Myr. That is, the K-M test does not identify a significant change for stars younger than 5 Gyr. The overall probability for stars <5 Gyr old is 24% with 95% confidence limits of 18% and 34%, whereas the corresponding values for stars >5 Gyr in age are 5.3% with 95% confidence limits of 1.8% and 14%. The drop in incidence of excesses is a factor of 4.5, at greater than 2σ significance. Our discussion in Section 3.3 demonstrates that errors in age or in identification of excesses are unlikely to change significantly the result that there are only four excesses in the portion of our sample older than 5 Gyr.
4.3. A Test and Confirmation

In principle, the K-M estimator works under the assumption of “random” censoring, that is, that the probability of a measurement being censored does not depend on the source property being probed, in this case stellar age. We deal with debris disk measurements in terms of their significance level (e.g., $\chi_{70}$) to allow for differing depths relative to the photosphere; about half of the sample is measured deeply enough ($\geq 3\sigma$, or $\geq 3\sigma$ on the expected photospheric level) to be detected. Thus the random censoring assumption may not be completely satisfied. For comparison, we therefore tested for the evolution of 70 $\mu$m excesses using a different method based on the shape of the probability distribution of excesses, which is also able to account for non-uniform detection limits (a similar approach has been used by Gáspár et al. 2013). We also use this method to confirm that the underlying assumption for the K-M test—that the censoring is independent of age—is satisfied for our sample.

We began by assuming that a sample of stars will have some probability distribution of excess ratios. Integrating this distribution upward from the detection threshold gives the probability that a star had a detectable excess. Lower limits for the integrals are given by

$$ LL = 1 + 3(\sigma_{70}/P_{70}),$$

(6)

where 1 represents the expected flux ratio of detected to photospheric flux density for no excess, $\sigma$ is the uncertainty on detected flux density, $P_{70}$, and $P_{70}$ is the predicted photospheric flux density. The factor of three represents the $3\sigma$ threshold placed on $\chi$ to have a significant excess. For each star in a sample, the integral of the probability distribution from this lower limit to infinity gives the probability of detecting an excess. The values for all the stars can be summed to obtain the total number of predicted excesses in the sample. The probability distribution can be refined by adopting a number of detection thresholds and comparing the sums over the sample stars with the number of detected excesses, thus testing the probability distribution as a function of amount of excess. Once the final model is obtained, it can be treated as a parent distribution for comparisons with other samples of stars. That is, under the null hypothesis that a sample is identical in excess behavior to the one for which the probability distribution was determined, we can use the distribution to compute how many detections should be achieved in the second sample and compare this prediction with the number of actual detections.

To apply this approach, we evaluated several possibilities (Gaussian, lognormal, power law, and exponential) for the shape of the distribution of excess ratios for the sample of stars younger than 5 Gyr and with signal-to-noise ratios at 70 $\mu$m of at least 2 (we fitted below the $\chi = 3$ reliable detection limit to be sure the distribution was well-behaved). We refined initial guesses for the fits by computing the lower integration limit for each star, regardless of whether it was detected, as in Equation (6). We defined the upper limit for the integral to be large enough that the value had converged. We required the sum to be 38, the total number of stars with indicated excesses ($\chi > 3$). We varied the width parameter and the normalization constant, leaving the sum unchanged (for the Gaussian and lognormal models, we only varied the width-parameter $\sigma$ and calculated the appropriate normalization factor; for the exponential and power-law distributions, the exponential factor was similarly varied—it is the only other free parameter for these models). We then used least-squares minimization to fit the integral sums for each function at selected excess thresholds to the number of actual detections above that threshold. We chose 10, 20, 30, 50, 60, 90, 100, and 200 as the test thresholds. The results of these tests are given in Table 4. The lognormal and power-law shapes produced the best fits. We found that the lognormal function was very consistent with the distribution of lower limits in the sample, while the power law was not. We therefore conclude that the excess ratios and lower limits are best fit with a lognormal shape:

$$ \frac{C}{\chi} \sqrt{2\pi} e^{-\frac{(\log x - \mu)^2}{2\sigma^2}}. $$

(7)

This result is consistent with previous work (Andrews & Williams 2005; Wyatt et al. 2007; Kains et al. 2011; Gáspár et al. 2013) that also found that the distributions of young disk masses or excesses could be fitted with lognormal functions. To test our model, we divided the sample into two age sub-bins of 80 and 79 stars: age $< 750$ Myr. Within these sub-bins, 21 and 18 stars with excesses were detected, respectively. The model predicts 18.2 and 19.8 excesses for those bins, confirming the fit.

Using lower integration limits computed from the data for the stars older than 5 Gyr, the lognormal model predicts 21.6 excesses at 70 $\mu$m, assuming no decay relative to the young stars. Thus, there is no significant dependence on the predicted number of excesses with stellar age; the three age divisions, respectively, have 18.2, 19.8, and 21.6 predicted detections, from virtually identical numbers of stars in each bin. This confirms the underlying assumption in the K-M test that the probability of censoring is not a function of stellar age.

We can also use this approach to confirm the result from the K-M test. Since only four excesses were detected, we can take the ratio of the predicted number to the detected number to obtain an estimated decay by a factor of $\sim 5$. The statistical significance of this result is dominated by the poor statistics for the old star sample; it indicates a decay by a factor greater than two ($2\sigma$ limit). Our modeling therefore establishes independently that debris disk excesses decay at 70 $\mu$m over time scales of 5 Gyr. Furthermore, our division of the young sample into two parts (younger and older than 750 Myr) and the finding of an equal incidence of excess in them, indicates that the decay occurs late in the evolution of the debris disks, past 1 Gyr.

To place constraints on the accuracy of the model, we used a Monte Carlo approach. We created random samples of the same size as our young ($< 5$ Gyr) sample, using the same parameters as those in the original probability distribution. For each random sample, we recomputed the best-fit parameters. We also fit the lower limits described above to a lognormal distribution and used that distribution to generate random samples of lower limits. The recomputed fits for the random star samples and random lower limits were used to compute a predicted number of excesses for the generated random sample of stars.

| Shape of Fit | $\chi^2$ | Normalization | $\mu$ | $\sigma$ |
|--------------|---------|---------------|------|--------|
| Power law    | 20.6    | 0.219         | 1.50 | ...    |
| Lognormal    | 6.2     | 0.448         | 1.50 | 2.12   |
| Exponential  | 139.8   | 0.284         | 25.9 | ...    |
| Gaussian     | 401.1   | 0.297         | 25.4 | 24     |

Note. $\mu$ for the power law and exponential distributions represents the exponential factor.
10,000 trials, the average number of excesses predicted for the young sample was 38.0 with a variance of 5.0. We repeated the procedure for the old sample by fitting the old sample lower limits to their own lognormal distribution, generating random old sample lower limits, and re-computing the fit parameters. We still used the recomputed best-fit parameters for the young star sample, and again calculated the predicted number of excesses in the old sample. The average number of excesses predicted for the random old samples was 21.6 with a variance of 1.3 over 10,000 trials. The error in the predicted number of excesses is therefore dominated by the statistical uncertainty in the number of detected young stars. Combining the uncertainties quadratically, the predicted number of excesses is 21.6 ± 2.5, confirming that our 2σ limit is valid.

This result is in excellent agreement with the conclusions from the K-M estimator. The consistency of the two different methods gives confidence that the conclusions drawn from the K-M estimator are correct.

5. More Detailed Evolution of Debris Disks

Another issue with previous studies, e.g., Bryden et al. (2006) and Carpenter et al. (2009), is small samples and as a result, poor statistical significance; this also limits our conclusions from the Spitzer sample. To examine the evolution of debris disks with better statistical weight, we combined the sample described in this paper with that from DUNES (Eiroa et al. 2013) and DEBRIS (Mathews et al. 2010). We took the ages of these stars from Gáspár et al. (2013), discarding all of quality “1”; these ages were estimated as in this paper, except for the omission of the step to reject cases where the H-R diagram and CAI were not in agreement (we therefore added a check on the H-R diagram as with the Spitzer sample for stars older than 5 Gyr). We also took the Spitzer/MIPS 70 μm and DEBRIS 100 μm photometry from Gáspár et al. (2013), with the DUNES photometry from Eiroa et al. (2013). We did not consider disks possibly detected only at 160 μm (see Gáspár & Rieke 2014), or discard photometry listed by Eiroa et al. (2013) in their Appendix D as likely to be confused with IR cirrus or background galaxies. We merged the measurements into a single indicator of an excess at 70–100 μm, and determined the significance of each excess as a χ value analogous to the definition in Equation (5) (see Gáspár et al. 2013).

5.1. Behavior of DEBRIS and DUNES Samples

There were 231 stars in the combined DEBRIS and DUNES samples within the relevant range of spectral types and with valid ages, which is nearly as large as the purely Spitzer sample. We have therefore first analyzed it independently as a check of the results in the preceding section. In this case, to divide the sample into thirds, we adopted age ranges of 0–1500 Myr, 1500–5000 Myr, and >5000 Myr. For excesses with R > 1.5, we found the detection probabilities were 17% (95% confidence limits of 10% and 28%), 19% (limits of 11% and 19%), and 11% (limits of 5% and 21%), respectively. These values are in agreement with those from the Spitzer sample, although a bit lower, thus the drop in the incidence of excesses in the oldest age range is also indicated, but at a slightly less statistically significant level (slightly less than 2σ). The overall similarity of the two samples is not surprising, given that they include virtually identical numbers of stars and that the numbers of IR excesses detectable by Spitzer/MIPS and Herschel/PACS are also similar (see Gáspár et al. 2013).

| R Threshold | 0–3000 Myr | ≥4500 Myr |
|-------------|------------|------------|
| 1.5         | 0.34±0.06  | 0.08±0.06  |
| 2           | 0.24±0.03  | 0.05±0.05  |
| 3           | 0.21±0.05  | 0.04±0.05  |
| 4           | 0.18±0.06  | 0.02±0.04  |
| 5           | 0.17±0.05  | 0.01±0.04  |
| 10          | 0.11±0.05  | 0.00±0.01  |

Note. * Errors are for 95% confidence limits.

Again, this conclusion rests on the reliability of the assigned ages, so we review those for the old stars (i.e., indicated ages >2500 Myr) with excesses. The following 12 stars have at least four independent age determinations (e.g., four measurements of the CAI) and should have relatively reliable ages: from DUNES, HIP 14954, F8V, 4500 Myr; HIP 15372, F2V, 4000 Myr; HIP 32480, G0V, 5000 Myr; HIP 65721, G4V, 8300 Myr; HIP 73100, F7V, 4000 Myr; HIP 85235, K0V, 5600 Myr; HIP107649, G2V, 4000 Myr; and from DEBRIS, HD 7570, F9V, 5300 Myr; HD 20010, F6V, 4800 Myr (there is only one measurement of the CAI, but multiple measurements put log g < 4); HD 20794, G8V, 6200 Myr; HD 102870, F9V, 4400 Myr; and HD 115617, G7V, 5000 Myr. The number of determinations for the following three stars is lower, but there is no credible evidence that the age estimates are seriously incorrect: HIP 27887 K2.5V, DUNES, 5500 Myr; HIP 71181 K3V, DUNES, 3000 Myr; and HD 38393 F6V, DEBRIS, 3000 Myr. Based on their positions on the H-R diagram and a review of other data, we revised the age in Gáspár et al. (2013) for HIP 32480, G0V, from 5000 to 3000 Myr and for HIP 62207, GOV, from 6400 to 1000 Myr. The CAI age of HD 61421, F5 IV–V, from DEBRIS, 2700 Myr, is rather uncertain because it has a highly variable activity level (Isaacson & Fischer 2010), but its position on the H-R diagram is roughly in agreement with the assigned age. In summary, the great majority of the relatively old stars with excesses appear to have reliable age estimates, as was also the case for the Spitzer sample. These stars are also distributed over the full range of spectral types in the sample.

5.2. Combined Sample

Because of the similarity in behavior between the Spitzer and Herschel samples, we could combine them directly. We used this combined sample to obtain a more detailed picture of the evolution of the far-infrared emission by debris disks. First, we tested for the overall evolution between the incidence of excesses in young and old disks. Below, we justify a division between a “young” sample with ages <3000 Myr and an “old” one with ages >4500 Myr, with the gap left as a transition. There are 235 stars in the first and 168 in the second category. We computed the excess probabilities based on different threshold values of R, and Table 5 shows the results. With the added statistical weight from combining the two samples, there is a well-detected drop in the incidence of excesses with age. The result is seen at all of the excess-detection thresholds (i.e., all values of R), although this behavior becomes stronger for larger excesses. At the lowest threshold, R = 1.5, it is significant roughly at a level of 5.5σ.

Two-thirds of this combined sample are detected at ≥3σ (or have photospheric levels that correspond to 3σ or greater), reducing the probability of censoring biases. Nonetheless, we
The Astrophysical Journal, 785:33 (13pp), 2014 April 10

Sierchio et al.

Table 6
Probable Incidence of Disks versus Age and $R$ Threshold, for the Combined Sample

| Age Range (Gyr) | Most Probable Fraction | 95% Confidence Upper Limit | 95% Confidence Lower Limit |
|----------------|------------------------|---------------------------|---------------------------|
| 280            | 0.39, 0.30, 0.19, 0.13 | 0.51, 0.41, 0.29, 0.22    | 0.29, 0.20, 0.11, 0.07    |
| 500            | 0.30, 0.24, 0.15, 0.11 | 0.41, 0.35, 0.25, 0.21    | 0.20, 0.16, 0.09, 0.06    |
| 850            | 0.21, 0.21, 0.19, 0.12 | 0.32, 0.32, 0.30, 0.22    | 0.13, 0.13, 0.12, 0.07    |
| 1400           | 0.24, 0.21, 0.15, 0.06 | 0.35, 0.32, 0.25, 0.15    | 0.15, 0.12, 0.08, 0.02    |
| 2200           | 0.32, 0.21, 0.12, 0.08 | 0.44, 0.32, 0.21, 0.16    | 0.22, 0.13, 0.06, 0.03    |
| 3300           | 0.27, 0.16, 0.08, 0.06 | 0.38, 0.26, 0.16, 0.15    | 0.18, 0.09, 0.04, 0.02    |
| 4450           | 0.12, 0.08, 0.04, 0.03 | 0.22, 0.17, 0.11, 0.10    | 0.06, 0.035, 0.01, 0.00   |
| 5400           | 0.08, 0.04, 0.01, 0.00 | 0.17, 0.12, 0.08, 0.06    | 0.03, 0.01, 0.00, 0.00    |
| 6300           | 0.09, 0.07, 0.03, 0.00 | 0.19, 0.16, 0.10, 0.06    | 0.04, 0.03, 0.01, 0.00    |
| 7500           | 0.08, 0.04, 0.01, 0.00 | 0.17, 0.12, 0.07, 0.06    | 0.04, 0.01, 0.00, 0.00    |

Note. a The successive entries in each cell are for $R \geq 1.5, 2, 5, 10$.

The important characteristics of the behavior are (1) the very large excesses decline more rapidly than smaller ones; (2) small excesses ($R \sim 2$) persist at a more-or-less constant incidence up to an age of $\sim 3000$ Myr; (3) there is a decline in the incidence of excesses between 3000 and 5000 Myr; and (4) there is a persistent (but spectral-type-independent) incidence of excesses for stars older than 5000 Myr, even though an extrapolation of the decay from 3000 to 5000 Myr would suggest they should be rare.

Previous work has found hints of the overall reduction of excesses at old ages, but has been hampered by inadequate statistics (e.g., Bryden et al. 2006; Carpenter et al. 2009), or by a number of excesses attributed to very old stars that contradicted the trend (Trilling et al. 2008). The larger sample, higher quality measurements, and emphasis on accurate age determination can account for the trend seen clearly in our study, whereas there were only hints in these previous ones.

In comparing these results with theoretical predictions, we took a fiducial debris disk radius of 30 AU (Gáspár et al. 2013) and a typical cold disk temperature of 60 K (Ballering et al. 2013). From the latter, we could estimate the fractional excesses corresponding to values of $R$, as listed in Table 7. At $R = 2$, the constant level of excess up to a drop between 3000 and 5000 Myr is then in approximate agreement in the pattern of behavior with the simple models for steady-state disk evolution in Wyatt (2008; Figure 5). A quantitative comparison with our results, however, cannot be made with the predicted decay for a single disk; a synthesis over a population of disks is required. Figure 7 shows such a model; it is for a population of debris systems around GOV stars, with an initial mass distribution width as in Andrews & Williams (2005) and a median mass of $5.24 \times 10^{-6}$ Earth masses (integrated up to particles of 1 mm radius).

Table 7
Approximate Fractional Excesses versus $R$

| $R$ at 85 $\mu$m | Fractional Excess |
|-----------------|-------------------|
| 1.5             | $4 \times 10^{-6}$|
| 2               | $9 \times 10^{-6}$|
| 3               | $1.7 \times 10^{-5}$|
| 4               | $2.5 \times 10^{-5}$|
| 5               | $3.5 \times 10^{-5}$|
| 10              | $8 \times 10^{-5}$|
| 15              | $1.2 \times 10^{-4}$|
| 20              | $1.6 \times 10^{-4}$|
| 25              | $2 \times 10^{-4}$|

Figure 7. Trend of far-infrared excesses with age (as measured by $R$, the ratio of the measured far-infrared flux density to the expected value for the stellar photosphere). The trends are smoothed with an 80 star running average; for $R = 2$ (red line) and $R = 10$ (orange), they are shown with gray zones indicating the $1\sigma$ uncertainties. The trends for $R = 5$ (blue dot–dashed line) and $R = 25$ (green dashed line) are also shown. A theoretical model for the $R = 2$ case (discussed in the text) is shown as the broad red line.

(A color version of this figure is available in the online journal.)

repeated the confirmation test described in Section 4.3 for the merged sample. The model shows that the combined sample is uniform in overall sensitivity to debris disk excesses across the entire age range (e.g., fitted values of 34–35 excesses in each of the younger and older halves of the “young sample” and to the “old” sample). The prediction of 35 for the old sample with a variance by the Monte Carlo simulation of 2.2 is significantly larger than the observed number of 11. The probability of this outcome (computed from the binomial distribution) if there is no evolution is $< 2 \times 10^{-7}$, or equivalently it is a $> 5\sigma$ indication of evolution.

Given the confirmation of the K-M results and the strong indication of evolution, we applied the K-M approach to estimate the incidence of debris excesses as a function of age. We used an 80 star running average of the combined sample to determine the most probable incidence of excesses for various values of $R$. The results are displayed in Figure 7 and selected values are listed in Table 6. As shown by the $1\sigma$ error zones, the curves are not significant.

6 The different distribution of ages in the combined sample, together with the more rapid evolution for large excesses, results in a different lognormal fit.
The debris is placed at a distance of 32.5 AU from the stars in narrow rings (10% width and height) and with optical and mechanical properties for icy grains. Details of the model are as in Gáspár et al. (2013). This model matches well the initial relatively constant incidence of disks up to ~3 Gyr and the decay from that age to ~5 Gyr. However, it falls below the observations at ages >5 Gyr. In general, the decay timescale is roughly inversely proportional to fractional excess, but with indications of more excesses than predicted for old stars. It is likely that the oldest disks represent late phases of dynamical activity, as also derived by Gáspár et al. (2013) with a detailed digital model of steady-state evolution that reproduces the number of excesses, but yields a distribution of excess amounts that falls significantly short of the observations.

6. CONCLUSIONS

We present a sample of 238 F4–K2 stars with high-quality Spitzer observations through 70 μm, and for which we have been able to determine accurate ages from the chromospheric activity index, X-ray emission, and placement on an H-R diagram. Consistent with previous work, we find that the incidence of excesses at 24 μm decays dramatically from 13.6% ± 2.8% for stars <5 Gyr in age to none for older stars. At 70 μm, we also find a significant reduction in the incidence of excesses, from 38, or 22.5% ± 3.6% to 4, or 4.7+3.7−2.2%. We have evaluated the significance of this apparent reduction in the incidence of 70 μm excesses in two ways, both of which correctly account for the variable detection limits and resulting censoring of the data. The first is the traditional K-M test, which indicates a decay by greater than 2σ significance, nominally by a factor of 4.5. In the second method, we determined the distribution of excess ratios among the young (ages <5 Gyr) stars and assumed it could be treated as a probability distribution to determine if a specific star is likely to have an excess above its individual detection limit. This model was used to predict the number of excesses that would have been detected in the old (>5 Gyr) sample if they behaved identically to the young sample. This approach also predicts a decay at greater than 2σ significance, and nominally by a factor of five. The two methods therefore agree well on the existence and amount of this decay. Within this sample, there are two stars with excesses at 24 μm but no strong excess at 70 μm—HD 1466 and HD 13246.

We then combined the Spitzer-only stars with the measurements for the DEBRIS and DUNES programs (for the former using both Spitzer and Herschel data to improve the weight of the measurements). This combined sample is doubled in size, allowing us to reach robust conclusions about debris disk evolution in the far infrared. Within this larger sample, we examine the relative incidence of excesses around stars of types F4–K4, finding no statistically significant trend with spectral type. We therefore analyze the evolution of the entire sample together. We find that, at fractional debris disk luminosity of just under 10−5, the incidence of far-infrared excesses declines significantly between 3000 and 5000 Myr, whereas for fractional luminosity of just under 10−4, the decline is at an order of magnitude younger ages. This behavior is consistent with a theoretical model for passive disk evolution. However, there appears to be more excesses for the oldest stars than predicted by passive-evolution models. These excesses are distributed over the full range of spectral types in our sample. This behavior suggests that the oldest debris disks are activated by late-phase dynamical activity.

We thank the anonymous referee for a very helpful report, and the editor for suggesting improved statistical approaches. We have made use of the SIMBAD database and the VizieR catalog access tool, operated at CDS, Strasbourg, France. We also used 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. This work is also based, in part, on observations made with Spitzer Space Telescope, which is operated by the Jet Propulsion Laboratory, California Institute of Technology under NASA contract 1407. We thank Patricia Whitelock and the South African Astronomical Observatory for assistance and use of facilities to obtain near-infrared photometry. This work supported by contract 1255094 from Caltech/JPL to the University of Arizona. J.M.S. is supported by the National Science Foundation Graduate Research Fellowship Program, under grant No. 1122374.

REFERENCES

Allen, D. S., & Gallagher, J. S. 1983, MNRAS, 203, 777
Alonso, A., Arribas, S., & Martínez-Roger, C. 1998, A&A, 311, 209
Ammons, S. S., Robinson, S. E., Strader, J., et al. 2006, ApJ, 638, 1004
An, D., Terndrup, D., & Pinsonneault, M. 2007, ApJ, 671, 1640
Andrews, S. M., & Williams, J. P. 2005, ApJ, 631, 1134
Aumann, H. H., & Probst, R. G. 1991, ApJ, 368, 264
Ballering, N. P., Rieke, G. H., Su, K. Y. L., & Montiel, E. 2013, ApJ, 775, 55
Barrado y Navascues, D. 1998, A&A, 339, 831
Beichman, C. A., Bryden, G., Stapelfeldt, K. R., et al. 2006, ApJ, 652, 1674
Bonatto, Ch., Bica, E., & Girardi, L. 2004, A&A, 415, 571
Bouchet, P., Schneider, F. X., & Manfroid, J. 1991, A&AS, 91, 409
Bryden, G., Beichman, C. A., Carpenter, J. M., et al. 2009, ApJ, 705, 1226
Bryden, G., Beichman, C. A., Rieke, G. H., et al. 2006, ApJ, 646, 1038
Bryden, G., Chen, X., Lin, D. N. C., Nelson, R. P., & Papaloizou, J. C. B. 1999, ApJ, 514, 344
Buccino, A. P., & Mauas, P. D. J. 2008, A&A, 483, 903
Carpenter, J. M. 2001, AJ, 121, 2851
Carpenter, J. M., Bouwman, J., Mamajek, E. E., et al. 2009, ApJS, 181, 197
Carter, B. S. 1990, MNRAS, 242, 1
Chen, C. H., Jura, M., Gordon, K. D., & Blaylock, M. 2005, ApJ, 623, 493
Chen, C. H., Sargent, B. A., Bohen, C., et al. 2006, ApJS, 166, 351
Clarke, C. J., Gendrin, A., & Soomayoy, M. 2001, MNRAS, 328, 485
Donaldson, J. K., Roberge, A., Chen, C. H., et al. 2012, ApJ, 753, 147
Duncan, D. K., Vaughan, A. H., Wilson, O. C., et al. 1991, ApJS, 76, 383
Eiroa, C., Marshall, J. F., Mora, A., et al. 2013, A&A, 555, 11
Engelbracht, C. W., Blaylock, M. S., & K. Y. L., et al. 2007, PASP, 119, 994
Feigelson, E. D., & Babu, G. J. 2012, Modern Statistical Methods in Astronomy (Cambridge: Cambridge Univ Press)
Fischer, D. A., & Valenti, J. 2005, ApJ, 622, 1102
Gáspár, A., Rieke, G. H., & Balog, Z. 2013, ApJ, 768, 25
Gáspár, A., & Rieke, G. H. 2014, ApJ, 784, 33
Glass, I. S. 1974, MNSA, 33, 53
Gordon, K. D., Engelbracht, C. W., Fadda, D., et al. 2007, PASP, 119, 1019
Gordon, K. D., Rieke, G. H., Engelbracht, C. W., et al. 2005, PASP, 117, 503
Gray, R. O., Corbally, C. J., Garrison, R. F., McFadden, M. T., & Robinson, P. E. 2003, AJ, 126, 2048
Gray, R. O., Corbally, C. J., Garrison, R. F., et al. 2006, AJ, 132, 161
Habing, H., Dominik, C., Jourdain de Muizon, M., et al. 1999, Nature, 401, 465
Habing, H., Dominik, C., Jourdain de Muizon, M., et al. 2001, A&A, 365, 545
Habermehl, P. 1998, A&A, 335, 173
Henry, T. J., Soderblom, D. R., Donahue, R. A., & Baliunas, S. L. O. 1996, AJ, 111, 439
Hollnberg, J., Nordström, B., & Andersen, J. 2009, A&A, 501, 941
Isaacson, H., & Fischer, D. A. 2010, ApJ, 725, 875
Jenkins, J. S., Jones, H. R. A., Tinney, C. G., et al. 2006, MNRAS, 372, 163
Johnson, J. L. 1966, ARA&A, 4, 193
Johnson, J. L., Mitchell, R. I., Iriarte, B., & Wisniewski, W. Z. 1966, CoLPL, 4, 99
Johnson, J. A., Aller, K. M., Howard, A. S., & Crepp, J. R. 2010, PASP, 122, 905
Kains, N., Wyatt, M. C., & Greaves, J. S. 2011, MNRAS, 414, 2486
Karatas, Y., Böhringer, J., & Schuster, W. J. 2005, MNRAS, 360, 1345
