THE MASSIVE YOUNG STAR-FORMING COMPLEX STUDY IN INFRARED AND X-RAY: X-RAY SOURCES IN 10 STAR-FORMING REGIONS

MICHAEL A. KUHN, Konstantin V. GETMAN, Patrick S. BROOS, Leisa K. TOWNSLEY, AND Eric D. FEIGELSON

Department of Astronomy & Astrophysics, Pennsylvania State University, University Park, PA 16802, USA

Received 2013 March 19; accepted 2013 August 27; published 2013 November 14

ABSTRACT

The Massive Young star-forming complex Study in Infrared and X-ray (MYStIX) uses data from the Chandra X-Ray Observatory to identify and characterize the young stellar populations in 20 Galactic (d < 4 kpc) massive star-forming regions. Here, the X-ray analysis for Chandra ACIS-I observations of 10 of the MYStIX fields is described, and a catalog of >10,000 X-ray sources is presented. In comparison to other published Chandra source lists for the same regions, the number of MYStIX-detected faint X-ray sources in a region is often doubled. While the higher catalog sensitivity increases the chance of false detections, it also increases the number of matches to infrared stars. X-ray emitting contaminants include foreground stars, background stars, and extragalactic sources. The X-ray properties of sources in these classes are discussed.

Key words: H II regions – stars: activity – stars: formation – stars: pre-main sequence – X-rays: stars

Online-only material: color figures, figure sets, supplemental data (FITS) file

1. INTRODUCTION

The Massive Young star-forming complex Study in Infrared and X-ray (MYStIX) is a survey of 20 of the nearest (d < 4 kpc) massive star-forming regions (MSFRs) that have been observed with NASA’s Chandra X-Ray Observatory (Chandra) and with infrared (IR) survey telescopes (Feigelson et al. 2013). With ages less than ~10 million years, each star-forming complex is dominated by OB stars and most have thousands of pre-main-sequence stars. Some portions of the stellar population are embedded in molecular clouds while for other portions the natal cloud material is ionized or dissipated.

X-ray observations are an effective strategy for obtaining a census of young stars in MSFRs. The X-ray luminosities of low-mass, pre-main-sequence stars are \(10^2\)–\(10^5\) times greater than for main-sequence stars, and X-ray surveys are sensitive to both disk-bearing and disk-free, young, low-mass stars (see reviews by Güdel & Naze 2009; Feigelson 2010). Massive OB stars also produce X-rays in shocks associated with their stellar winds (Lucy & White 1980; Owocki & Cohen 1999; Townsley et al. 2003). Both classes often exhibit a hard X-ray component that penetrates high column densities of interstellar material, and the sub-arcsecond on-axis resolution of Chandra probes the crowded centers of young stellar clusters.

The MYStIX project combines Chandra studies of MSFRs lying 0.4–3.6 kpc from the Sun with near-IR surveys from the United Kingdom Infrared Telescope (UKIRT), often from the UKIDSS project (Lawrence et al. 2007), and mid-IR surveys from NASA’s Spitzer Space Telescope Infrared Array Camera. Published OB stars are also included in the MYStIX samples. A full description of the MYStIX project is given by Feigelson et al. (2013). Together, MYStIX obtains large samples of the young stellar populations with a wide range of masses and ages for improved studies of star-cluster formation, cluster dynamical evolution, triggered star formation, and other issues relating to clustered star formation in giant molecular clouds.

This work describes the X-ray observations, data analysis, and resulting X-ray source lists and properties for 10 MYStIX MSFRs. In order of increasing distance from the Sun (see Table 1 of Feigelson et al. 2013), they are: the Flame Nebula, RCW 36, NGC 2264, the Rosette Nebula, the Lagoon Nebula, NGC 2362, DR 21, RCW 38, the Trifid Nebula, and NGC 1893. Additional MYStIX Chandra targets to be treated by L. K. Townsley et al. (in preparation) are: NGC 6334, NGC 6357, the Eagle Nebula, M 17, W 3, W 4, and NGC 3576. Previous studies have been published for much of the Chandra data; the MYStIX project reanalyzes these observations obtained from the Chandra archive in a unified fashion with methods that are tuned to finding weak sources in crowded environments with overlapping exposures. The MYStIX project adopts published X-ray source lists and properties for three MSFRs that were analyzed in a similar fashion: the Orion Nebula (Getman et al. 2005), W 40 (Kuhn et al. 2010), and the Carina Nebula complex (Townsley et al. 2011; Broos et al. 2011b).

In this paper, Section 2 describes Chandra observations and data reduction, Section 3 describes the X-ray source lists, Section 4 compares our analysis to the results of prior X-ray studies of these regions, Section 5 discusses X-ray sources that are not young stars, Section 6 discusses distributions of X-ray flux and spectral hardness for various populations of X-ray sources, and Section 7 is the summary.

2. CHANDRA OBSERVATIONS AND DATA REDUCTION

X-ray observations were made with the imaging array on the Advanced CCD Imaging Spectrometer (ACIS-I; Garmire et al. 2003) on board the Chandra X-Ray Observatory (Weisskopf et al. 2002). This array of four CCD detectors subtends \(17' \times 17'\) on the sky. (We exclude data from the ACIS spectroscopic array due to Chandra’s reduced angular resolution far off axis.) The number of different Chandra pointings for each region, the total exposures for these pointings, and details of how the observations were taken are provided in Table 1. Data were acquired from the Chandra Data Archive\(^1\) and were prepared from the “Level 1” data products derived from the satellite telemetry.

The Chandra observations were configured to meet the scientific goals of the original projects, so the spatial coverage

\(^1\) http://cxc.harvard.edu/cda
and exposure durations vary among MSFRs. Sometimes a single ACIS pointing is available, while other targets have a mosaic of overlapping pointings. Furthermore, any pointing may be broken into several distinct observations, “ObsIDs,” often due to spacecraft constraints that prevented longer continuous observations. Overall, 29 Chandra ObsIDs are included with typical integration times for a pointing of 40–100 ks. Based on the X-ray luminosity function of the well-studied Orion Nebula Cluster (Feigelson et al. 2005) extrapolated to distances of 0.5–1 M_☉ for young stars are seen. The methodology of the MYStIX data reduction follows closely the procedures established by the Chandra Carina Complex Project (CCCP; Townsley et al. 2011). These procedures are described in Broos et al. (2010), Getman et al. (2010), and Broos et al. (2011b). The analysis uses codes from CIAO (Fruscione et al. 2006), MARX (Davis et al. 2012), HEASoft, and the Astronomy User’s Library (Landsman 1993), which are

### Table 1

Log of Chandra Observations

| ObsID | Sequence | Start Time (UT) | Exposure (s) | Aimpoint α_J2000 | δ_J2000 (°) | Roll Angle (°) | Mode | PI |
|-------|----------|----------------|-------------|------------------|------------|---------------|------|----|
| Flame Nebula | 1878 | 2001-08-08T06:37 | 75457 | 05:41:46.30 | −01:55:28.7 | 111 | Faint | S. Skinner |
| RCW 36 | 6433 | 2006-09-23T19:44 | 69708 | 08:59:27.49 | −43:45:27.0 | 124 | Very Faint | M. Tsuji moto |
| NGC 2264 | 2540 | 2002-10-28T14:46 | 95142 | 06:40:58.09 | +09:34:00.4 | 78 | Faint | S. Sciortino |
| | 2550 | 2002-09-07T05:10 | 48134 | 06:40:48.00 | +09:50:59.9 | 281 | Faint | J. Stauffer |
| | 9768 | 2008-03-12T17:55 | 27786 | 06:41:11.99 | +09:30:00.0 | 270 | Faint | G. Micela |
| | 9769 | 2008-03-28T14:47 | 29756 | 06:41:11.99 | +09:30:00.0 | 266 | Faint | G. Micela |
| Rosette Nebula | 1874 | 2001-05T11:54 | 15950 | 06:31:52.00 | +04:55:57.0 | 335 | Faint | L. Townsley |
| | 1875 | 2001-05T17:47 | 19503 | 06:32:40.00 | +04:43:00.0 | 335 | Faint | L. Townsley |
| | 1876 | 2001-05T23:29 | 19506 | 06:33:16.30 | +04:34:56.9 | 335 | Faint | L. Townsley |
| | 1877 | 2001-06T05:11 | 19506 | 06:34:16.49 | +04:28:00.9 | 335 | Faint | L. Townsley |
| | 1878 | 2001-05T12:02 | 74999 | 06:31:55.50 | +04:56:34.0 | 351 | Very Faint | L. Townsley |
| | 8454 | 2007-02-09T02:00 | 11844 | 06:30:50.39 | +04:59:34.0 | 286 | Very Faint | G. Garmire |
| | 12142 | 2010-12-10T04:36 | 39540 | 06:34:37.60 | +04:12:44.2 | 41 | Very Faint | G. Garmire |
| Lagoon Nebula | 977 | 2001-06-18T11:40 | 59598 | 18:04:24.00 | −24:21:20.0 | 80 | Faint | S. Murray |
| | 3754 | 2002-07-25T17:28 | 108512 | 18:03:45.10 | −24:22:05.0 | 272 | Very Faint | M. Gagné |
| | 4297 | 2002-07-24T10:08 | 14634 | 18:03:45.10 | −24:22:05.0 | 272 | Very Faint | M. Gagné |
| | 4444 | 2002-07-28T00:01 | 29060 | 18:03:45.10 | −24:22:05.0 | 272 | Very Faint | M. Gagné |
| NGC 2362 | 4469 | 2003-12-23T03:03 | 97883 | 07:18:42.79 | −24:57:18.5 | 22 | Very Faint | S. Murray |
| DR 21 | 7444 | 2004-08-08T03:28 | 48706 | 20:39:09.07 | +42:18:56.8 | 204 | Very Faint | F. Damiani |
| | 8598 | 2004-07-21T23:40 | 20480 | 20:39:09.07 | +42:18:56.8 | 303 | Very Faint | F. Damiani |
| | 9770 | 2004-07-29T15:12 | 18986 | 20:39:09.07 | +42:18:56.8 | 303 | Very Faint | F. Damiani |
| | 9771 | 2004-07-20T09:50 | 9143 | 20:39:09.07 | +42:18:56.8 | 303 | Very Faint | F. Damiani |
| RCW 38 | 2556 | 2001-12-01T10:14 | 95924 | 08:59:19.20 | −47:30:21.9 | 51 | Very Faint | S. Wolk |
| Trifid Nebula | 2566 | 2002-06-13T02:17 | 58061 | 18:02:30.30 | −23:01:29.3 | 105 | Faint | J. Rho |
| NGC 1893 | 6406 | 2006-11-09T12:51 | 115662 | 05:22:49.99 | +33:28:05.0 | 107 | Faint | G. Micela |
| | 6407 | 2006-11-15T05:31 | 126217 | 05:22:49.99 | +33:28:05.0 | 107 | Faint | G. Micela |
| | 6408 | 2006-07-23T00:12 | 102828 | 05:22:49.99 | +33:28:05.0 | 262 | Faint | G. Micela |
| | 8462 | 2006-07-11T13:33 | 42615 | 05:22:49.99 | +33:28:05.0 | 107 | Faint | G. Micela |
| | 8476 | 2006-07-17T10:55 | 53268 | 05:22:49.99 | +33:28:05.0 | 107 | Faint | G. Micela |

Notes. Exposure times are the net usable times after various filtering steps are applied in the data reduction process. The aimpoints and roll angles are obtained from the satellite aspect solution before astrometric correction is applied. Units of right ascension are hours, minutes, and seconds; units of declination are degrees, arcminutes, and arcseconds.
integrated by scripts in the IDL language. The analysis combines all imaging-mode ACIS-I data that is available from the Chandra archive to create improved Chandra point-source catalogs for the regions in our study. Compared to procedures used by many previous researchers, the methodological improvements allow up to twice as many X-ray point sources to be detected (Broos et al. 2011b). We briefly summarize the data reduction methodology, including updates to the Carina analysis procedures that are outlined here and presented in detail by L. K. Townsley et al. (in preparation).

Following Broos et al. (2010), the “Level 2” data products are rebuilt from the “Level 1” products. This processing involves event energy calibration, improvements to the position of individual events, and cleaning of contamination by bad pixels, background flaring, and cosmic-ray afterglows. Heavily cleaned event lists are used for source detection and extraction of weak sources, while lightly cleaned data are used for extraction of bright sources. Alignment of the ObsID coordinate systems is adjusted using the positions of sources from the Two Micron All Sky Survey (2MASS; Skrutskie et al. 2006), and multiple ObsIDs are tiled to create a Chandra mosaic image.

The most common source-detection procedures for Chandra imaging observations of star-forming regions are based on the wavelet transformation using the wavdetect procedure developed by Freeman et al. (2002) or the PWDetect procedure developed by Damiani et al. (1997). These methods do not use calibration files that describe the complex shapes and strong spatial variations in the telescope point-spread function due to the unusual optics needed for wide-field X-ray imaging, so they are not optimal at detecting the faintest sources, resolving closely spaced sources in a crowded field, or recovering source positions for far off-axis sources. We use instead a maximum likelihood reconstruction applied to X-ray events in small tiles across the field that accurately traces spatial variations in the point-spread function, a procedure developed by Townsley et al. (2006). The reconstruction is calculated with the Lucy–Richardson algorithm (Lucy 1974) which is an implementation of the EM algorithm widely used in maximum likelihood statistical calculations (McLachlan & Krishnan 2008). A bump-hunting algorithm is then applied to the reconstructed image to give a superset of candidate point sources.

The ACIS Extract package (Broos et al. 2010, 2012) is then applied to the original data to extract events for each of the candidate point sources. This is an elaborate procedure that again uses the local point-spread function to estimate source position, create an extraction region that does not overlap neighboring extraction regions, measure a background from an optimized region that accounts for contamination by neighboring sources, and compute the null probability, ProbNoSrc_min, that no source exists given the observed local background level, assuming the events follow the Poisson distribution. Starting with the candidate point sources, insignificant sources are

### Table 2
X-Ray Observations of 10 MYStIX Star-forming Regions

| Name          | Location (α, δ) | Distance (kpc) | Area (deg² pc²) | Previous pubs | Number of X-Ray Sources |
|---------------|----------------|----------------|-----------------|---------------|-------------------------|
| Flame Nebula  | 0541−01        | 0.414          | 0.08 (4)        | 1, 2          | 547                     |
| RCW 36        | 0859−43        | 0.7 ± 0.2      | 0.08 (12)       | ...           | 502                     |
| NGC 2264      | 0641+09        | 0.913 ± 0.1   | 0.19 (52)       | 3, 4, 5, 6    | 1,328                   |
| Rosette Nebula| 0632+05        | 1.33 ± 0.05    | 0.46 (248)      | 7, 8, 9, 10   | 1,962                   |
| Lagoon Nebula | 1804−24        | 1.3 ± 0.3      | 0.13 (69)       | 11, 12        | 2,427                   |
| NGC 2362      | 0718−25        | 1.48           | 0.08 (55)       | 13, 14, 15    | 690                     |
| DR 21         | 2039+42        | 1.50 ± 0.08    | 0.09 (62)       | ...           | 765                     |
| RCW 38        | 0859−47        | 1.7 ± 0.9     | 0.08 (71)       | 16, 17, 18    | 1,019                   |
| Trifid Nebula | 1802−23        | 2.7 ± 0.5     | 0.08 (182)      | 19            | 635                     |
| NGC 1893      | 0523+33        | 3.6 ± 0.2      | 0.10 (379)      | 20, 21        | 1,442                   |

Notes. X-ray publications: 1. Skinner et al. (2003) 2. Ezoe et al. (2006) 3. Ramírez et al. (2004) 4. Sung et al. (2004) 5. Rebull et al. (2006) 6. Flaccomio et al. (2006) 7. Townsley et al. (2003) 8. Wang et al. (2008) 9. Wang et al. (2009) 10. Wang et al. (2010) 11. Damiani et al. (2004) 12. Henderson & Stassun (2012) 13. Delgado et al. (2006) 14. Damiani et al. (2006) 15. Dahm et al. (2007) 16. Wolk et al. (2002) 17. Wolk et al. (2006) 18. Winston et al. (2011) 19. Rho et al. (2004) 20. Caramazza et al. (2008) 21. Caramazza et al. (2012).

a Tot = total number of Chandra sources. “Faint” sources have <10 counts, “Mod” sources have 10–100 counts, “Strong” sources have >100 counts. These are net extracted counts after background subtraction in the Chandra total band (0.5–8 keV).

---

**Figure 1.** Adaptively smoothed ACIS-I mosaic images of the NGC 1893 MSFRs shown in logarithmic scale. Smoothing has been performed on the total 0.5–8.0 keV band. (The complete figure set (10 images) is available in the online journal.)
pruned and source positions, extraction regions, backgrounds, and $ProbNoSrc_{\text{min}}$ values are recomputed in an iterative process until convergence of the source list is achieved. Candidate sources that satisfy $ProbNoSrc_{\text{min}} < 0.01$, or a 99% probability that the source is not a random fluctuation in the background, are considered to be valid. This criterion can sometimes detect on-axis sources down to three extracted counts, considerably fainter than most wavelet methods. Thus, a fraction of the faintest X-ray sources may be spurious. We choose to produce the most sensitive X-ray catalog possible because, in our experience with MYStIX and earlier studies, even the faintest X-ray sources are often associated with IR-detected young stars. The X-ray positions typically have subarcsecond accuracy; sources within 3′ of the field center have median 1σ uncertainties around 0′.15, increasing to ~0′.8 at 10′ off-axis. Thus, spurious X-ray sources will rarely be incorrectly matched with IR sources. Our highly sensitive X-ray catalog plays a major role in producing rich MYStIX catalogs of probable members of the observed star-forming complexes (Broos et al. 2013).

These methods are more fully presented in Broos et al. (2011b) where they were used for the identification of >14,000 X-ray sources in the CCCP. The MYStIX analysis is somewhat improved over the CCCP analysis. First, astrometric alignment is improved. Second, candidate sources within an arcsecond of a bright source are visually evaluated as to whether they are consistent with that artifact. Possible examples of an artifact found in the Chandra photopile up and readout trails associated with bright sources. Third, potential background regions now account for CCD response with MYStIX and earlier studies, even the faintest X-ray sources are often associated with IR-detected young stars. The X-ray positions typically have subarcsecond accuracy; sources within 3′ of the field center have median 1σ uncertainties around 0′.15, increasing to ~0′.8 at 10′ off-axis. Thus, spurious X-ray sources will rarely be incorrectly matched with IR sources. Our highly sensitive X-ray catalog plays a major role in producing rich MYStIX catalogs of probable members of the observed star-forming complexes (Broos et al. 2013).

These methods are more fully presented in Broos et al. (2011b) where they were used for the identification of >14,000 X-ray sources in the CCCP. The MYStIX analysis is somewhat improved over the CCCP analysis. First, astrometric alignment is improved. Second, candidate sources within an arcsecond of a bright source are visually evaluated as to whether they are possible examples of an artifact found in the Chandra point-spread function$^4$ and removed if consistent with that artifact. Possible examples of an artifact found in the Chandra point-spread function$^4$ and removed if consistent with that artifact. Possible examples of an artifact found in the Chandra point-spread function$^4$ and removed if consistent with that artifact.

Source properties such as position, flux, spectral hardness, and variability are then computed with methods adapted to the low counts seen in most MYStIX fields (see columns 7–9 of Table 2). As many MYStIX sources are too weak for statistical fitting of spectral models, nonparametric procedures to estimate hydrogen column densities and absorption-corrected fluxes described by Getman et al. (XPHOT; 2010) are used.$^5$ An important element is the use of the median energy of extracted events, the most robust estimate of spectral hardness, as an estimator of interstellar absorption. This is combined with prior empirical knowledge (e.g., from the Chandra Orion Ultradeep Project; Getman et al. 2005) of young stellar X-ray spectra to obtain good estimates of absorption-corrected source fluxes and their systematic and statistical errors. This analysis assumes the X-ray spectral shapes of young, low-mass stars. Therefore, XPHOT quantities will be unreliable for high-mass stars that are dominated by X-ray emission from a stellar wind.

### 3. LIST OF X-RAY SOURCES

A total of 11,315 X-ray sources were identified and extracted. Columns 6–9 of Table 2 summarize the resulting source populations in the 10 MYStIX fields analyzed here. Most of the X-ray sources are quite faint with typical sources exhibiting around 10 counts, although source counts range from ≤4 to >1000. Bright sources are occasionally affected by pile up, which will bias inferred X-ray properties, decreasing X-ray count rates and increasing spectral hardness. Sources with significant pile up are listed in Table 3.

| Region | IAU Source Name |
|--------|-----------------|
| Flame  | 054137.74-015351.5 |
|        | 054138.24-015309.1 |
|        | 054138.58-015322.8 |
|        | 054141.35-015327.2 |
|        | 054146.11-015414.8 |
|        | 054148.21-015601.9 |
| NGC 2264 | 064040.44+095050.4 |
|        | 064046.07+094917.3 |
|        | 064046.07+094917.3 |
|        | 064056.50+095410.5 |
|        | 064058.50+093313.7 |
|        | 064058.66+095344.8 |
|        | 064103.49+093118.4 |
|        | 064105.36+093313.5 |
|        | 064105.54+093140.6 |
|        | 064105.54+093140.6 |
|        | 064105.74+093101.3 |
|        | 064106.19+093622.9 |
|        | 064106.82+092732.2 |
|        | 064108.59+092933.7 |
|        | 064110.00+092746.0 |
|        | 064113.03+092732.0 |
| Rosette | 063155.51+045634.3 |
|        | 063231.43+044234.0 |
|        | 063327.50+043557.2 |
| Lagoon | 180352.45-242138.5 |
|        | 180407.36-242221.9 |
|        | 180414.63-242155.6 |
| NGC 2362 | 071842.48-245715.8 |
|        | 071845.25-245643.9 |
|        | 071850.46-245754.7 |
| RCW 38  | 085905.65-473040.9 |

Table 4 lists X-ray properties obtained from ACIS Extract and XPHOT. Photometry is computed for three bands: soft (0.5–2.0 keV), hard (2.0–8.0 keV), and total (0.5–8.0 keV). Descriptions of how ACIS Extract quantities are calculated are given by Broos et al. (2010, 2011b). Of particular interest are median energy and photon flux values, which may be directly compared between different Chandra ACIS-I observations. Inferred spectroscopic properties from XPHOT include hydrogen column density, X-ray flux (both incident and absorption-corrected) in the hard and total bands, and systematic and statistical errors on these quantities.

Point-source detection sensitivity is related to photon flux in the bands used for source detection.$^6$ However, sensitivity varies across the fields of view due to telescope vignetting, degradation of the point-spread function with off-axis angle, and overlapping exposures of different durations in mosaics. In particular, the “egg-crate effect” was identified by Broos et al. (2011b) where there is a higher density of X-ray sources near the focal axis of the telescope due to increased sensitivity. Similar effects are seen in MYStIX. For example, there are a number of weak, ≤4 count sources near the centers of the ACIS-I pointings, many of which have hard X-ray spectra and no IR counterpart—consistent with being extragalactic X-ray sources.

---

$^4$ http://cxc.harvard.edu/ciao4.4/caveats/psf_artifact.html

$^5$ The XPHOT software package is available from http://www2.astro.psu.edu/users/gkosta/XPHOT/. See also Getman et al. (2012).

---
| Column Label                  | Units       | Description                                                                 |
|------------------------------|-------------|------------------------------------------------------------------------------|
| MYSTIX_SFR                   |             | name of MYStIX star-forming region                                           |
| Name                         |             | IAU source name; prefix is CXO J                                             |
| Label                       |             | source name used within the project                                          |
| RAdeg                       | deg         | right ascension (J2000)                                                      |
| DEdeg                       | deg         | declination (J2000)                                                         |
| PosErr                      | arcsec      | 1σ error circle around (RAdeg,DEdeg)                                        |
| PosType                     |             | algorithm used to estimate position (Broos et al. 2010, Section 7.1)        |
| ProbNoSrc_min               |             | smallest of ProbNoSrc_t, ProbNoSrc_s, ProbNoSrc_h                          |
| ProbNoSrc_t                 |             | p-value for no-source hypothesis (Broos et al. 2010, Section 4.3)           |
| ProbNoSrc_s                 |             | p-value for no-source hypothesis                                             |
| ProbNoSrc_h                 |             | p-value for no-source hypothesis                                             |
| ProbKS_single               |             | smallest p-value for the one-sample Kolmogorov–Smirnov statistic under the no-variability null hypothesis within a single-observation |
| ProbKS_merge                |             | smallest p-value for the one-sample Kolmogorov–Smirnov statistic under the no-variability null hypothesis over merged observations |
| ExposureTimeNominal         | s           | total exposure time in merged observations                                   |
| ExposureFraction            |             | fraction of ExposureTimeNominal that source was observed                    |
| NumObservations             |             | total number of observations extracted                                       |
| NumMerged                   |             | number of observations merged to estimate photometry properties             |
| MergeBias                   |             | fraction of exposure discarded in merge                                      |
| Theta_Lo                    | arcmin      | smallest off-axis angle for merged observations                              |
| Theta                        | arcmin      | average off-axis angle for merged observations                               |
| Theta_HI                    | arcmin      | largest off-axis angle for merged observations                               |
| PsfFraction                 |             | average PSF fraction (at 1.5 keV) for merged observations                   |
| SrcArea                     | (0.492 arcsec)² | average aperture area for merged observations                                 |
| AfterglowFraction           |             | suspected afterglow fraction                                                 |
| SrcCounts_t                 | count       | observed counts in merged apertures                                          |
| SrcCounts_s                 | count       | observed counts in merged apertures                                          |
| SrcCounts_h                 | count       | observed counts in merged apertures                                          |
| BkgScaling                  |             | scaling of the background extraction (Broos et al. 2010, Section 5.4)       |
| BkgCounts_t                 | count       | observed counts in merged background regions                                |
| BkgCounts_s                 | count       | observed counts in merged background regions                                |
| BkgCounts_h                 | count       | observed counts in merged background regions                                |
| NetCounts_t                 | count       | net counts in merged apertures                                               |
| NetCounts_s                 | count       | net counts in merged apertures                                               |
| NetCounts_h                 | count       | net counts in merged apertures                                               |
| NetCounts_Lo_t             | count       | 1σ lower bound on NetCounts_t                                               |
| NetCounts_HI_t             | count       | 1σ upper bound on NetCounts_t                                               |
| NetCounts_Lo_s             | count       | 1σ lower bound on NetCounts_s                                               |
| NetCounts_HI_s             | count       | 1σ upper bound on NetCounts_s                                               |
| NetCounts_Lo_h             | count       | 1σ lower bound on NetCounts_h                                               |
| NetCounts_HI_h             | count       | 1σ upper bound on NetCounts_h                                               |
| MeanEffectiveArea_t         | cm² count photon⁻¹ | mean ARF value                                                                |
| MeanEffectiveArea_s         | cm² count photon⁻¹ | mean ARF value                                                                |
| MeanEffectiveArea_h         | cm² count photon⁻¹ | mean ARF value                                                                |
| MedianEnergy_s              | keV         | median energy, observed spectrum                                             |
| MedianEnergy_h              | keV         | median energy, observed spectrum                                             |
| PhotonFlux_t                | photon cm⁻² s⁻¹ | log incident photon flux                                                     |
| PhotonFlux_s                | photon cm⁻² s⁻¹ | log incident photon flux                                                     |
| PhotonFlux_h                | photon cm⁻² s⁻¹ | log incident photon flux                                                     |
| X-ray Spectral Model        |             | X-ray flux, 2.8 keV                                                          |
| F_HJ                        | erg cm⁻² s⁻¹ | absorption-corrected X-ray flux, 2.8 keV                                    |
| F_HC                        | erg cm⁻² s⁻¹ | absorption-corrected X-ray flux, 2.8 keV                                    |
| SF_HC_STAT                  | erg cm⁻² s⁻¹ | 1σ statistical uncertainty on FX_HC                                          |
| SF_HC_SYST                  | erg cm⁻² s⁻¹ | 1σ systematic uncertainty on FX_HC                                          |
| F_T                         | erg cm⁻² s⁻¹ | X-ray flux, 0.5:8 keV                                                        |
| F_TC                        | erg cm⁻² s⁻¹ | absorption-corrected X-ray flux, 0.5:8 keV                                  |
| SF_TC_STAT                  | erg cm⁻² s⁻¹ | 1σ statistical uncertainty on FX_TC                                          |
| SF_TC_SYST                  | erg cm⁻² s⁻¹ | 1σ systematic uncertainty on FX_TC                                          |
| LOGNH_OUT                   | cm⁻²         | gas column density                                                          |
sources and/or spurious detections. Some MYStIX science studies require uniform-in-flux source samples across the entire regions; for example, the spatial structure study of M. A. Kuhn et al. (2013, in preparation) uses the empirical distributions of X-ray sources at various off-axis angles to calculate the incident 0.5–8.0 keV photon flux above which the sample is complete.

The determination of whether an X-ray source is a young star in the target MSFR is not made here. MYStIX Probable Complex Members (MPCMs) are classified using X-ray and IR star in the target MSFR is not made here. MYStIX Probable Complex Members (MPCMs) are classified using X-ray and IR.

### Table 4

(Continued)

| Column Label                                      | Units    | Description                        |
|---------------------------------------------------|----------|------------------------------------|
| SLOGNH_OUT_STAT_OUT                               | cm⁻²     | 1σ statistical uncertainty on LOGNH_OUT |
| SLOGNH_OUT_SYSYST_OUT                             | cm⁻²     | 1σ systematic uncertainty on LOGNH_OUT |

**Notes.** Rows are sorted by R.A. Col. (1): Column label previously published by the CCCP (Broos et al. 2011b) and produced by the ACIS Extract software. The suffixes “_t,” “_s,” and “_h” on names of photometric quantities designate the total (0.5–8 keV), soft (0.5–2 keV), and hard (2–8 keV) energy bands. Source significance quantities (ProbNoSrc_t, ProbNoSrc_s, ProbNoSrc_h, ProbNoSrc_min) are computed using a subset of each source’s extractions chosen to maximize significance (Broos et al. 2010, Section 6.2). Source position quantities (RAdeg, DEdeg, PosErr) are computed using a subset of each source’s extractions chosen to minimize the position uncertainty (Broos et al. 2010, Sections 6.2 and 7.1). All other quantities are computed using a subset of each source’s extractions chosen to balance the conflicting goals of minimizing photometric uncertainty and of avoiding photometric bias (Broos et al. 2010, Sections 6.2 and 7).

1 Source labels identify a Chandra pointing; they do not convey membership in astrophysical clusters.

In statistical hypothesis testing, the p-value is the probability of obtaining a test statistic at least as extreme as the one that was actually observed when the null hypothesis is true.

3 See Broos et al. (2010, Section 7.6) for a description of the variability metrics, and caveats regarding possible spurious indications of variability using the ProbsKsMerge metric.

Due to dithering over inactive portions of the focal plane, a Chandra source is often not observed during some fraction of the nominal exposure time. (See http://cxc.harvard.edu/ciao/why/dither.html.) The reported quantity is FRACEXPO produced by the CIAO tool mkarf.

Some background events arising from an effect known as “afterglow” (http://cxc.harvard.edu/ciao/why/afterglow.html) may contaminate source extractions, despite careful procedures to identify and remove them during data preparation (Broos et al. 2010, Section 3). After extraction, we attempt to identify afterglow events using the tool ae_afterglow_report, and report the fraction of extracted events attributed to afterglow; see the ACIS Extract manual (http://www.astro.psu.edu/xray/acis/acis_analysis.html).

Confidence intervals (68%) for NetCounts quantities are estimated by the CIAO tool aprates (http://asc.harvard.edu/ciao/ahelp/aprates.html).

The ancillary response file (ARF) in ACIS data analysis represents both the effective area of the observatory and the fraction of the observation for which data were actually collected for the source (ExposureFraction).

MedianEnergy is the median energy of extracted events, corrected for background (Broos et al. 2010, Section 7.3).

PhotonFlux = (NetCounts / MeanEffectiveArea / ExposureTimeNominal) (Broos et al. 2010, Section 7.4).

XPHOT assumes X-ray spectral shapes of young, low-mass stars, which come from coronal X-ray emission. XPHOT quantities will therefore be unreliable for high-mass stars, for which X-ray emission is associated with the stellar wind (Getman et al. 2010).

(This table is available in its entirety in FITS format in the online journal. A portion is shown here for guidance regarding its form and content.)

4. COMPARISON TO PUBLISHED SOURCE CATALOGS

As indicated in Table 2 (column 5), most of these Chandra fields have been previously analyzed. Generally, the MYStIX analysis emerges with considerably larger source lists than previous studies. This is partly due to improved methods for super-resolving crowded regions using maximum likelihood image reconstruction and improved treatment of local background levels. However, the MYStIX procedure also sets a lower threshold for source existence (ProbNoSrc_min = 1%) compared to many other analyses. MYStIX will thus have more false positive (spurious) sources near the detection threshold. This was purposely done because the lower threshold also picks up many real faint sources. Possible spurious sources can be excluded from the MPCM sample of probable members because they will not have the necessary properties (e.g., match to infrared source, X-ray flaring) for classification by Broos et al. (2013). These sources are generally ranked as “unclassified.” However, past experience indicates that many of the faintest sources have IR counterparts, and improved near-IR data has often led to new IR counterparts to the faintest ACIS Extract X-ray sources that had previously been “unclassified” (e.g., Broos et al. 2011a). This is expected because the pre-main sequence X-ray luminosity function rises steeply around log L_X ~ 30 erg s⁻¹ where ProbNoSrc_min ~ 0.1–1% in most MYStIX fields.

Global comparisons with previous work can be made directly for some MYStIX MSFRs. Other cases used different datasets in past studies, or used data analysis procedures similar to those in MYStIX.

7 Monte Carlo simulation of source detection with artificial sources is a common procedure for evaluating a detection algorithm; however, several reasons why this procedure will not produce useful information in this case are described by Broos et al. (2011b, Section 6.2). Briefly, the complexities of the Chandra survey and detection procedures limit the accuracy of simulations. Furthermore, propagation of simulated spurious sources through additional sample selection (e.g., X-ray/IR matching, source classification) and science analysis would be difficult. Broos et al. (2011b, their Figure 9) show the number of X-ray sources with and without near-IR counterparts stratified by ProbNoSrc_min, where the number of X-ray sources without counterparts is an upper limit to the number of spurious sources. The fraction without counterparts increases with larger ProbNoSrc_min, but even near the ProbNoSrc_min = 0.01 detection threshold there is no sudden increase in this category. Similar trends are found in MYStIX.
The Flame Nebula. Skinner et al. (2003) analyzed this early Chandra field using standard procedures recommended by the Chandra X-Ray Center. The wavdetect wavelet source-detection algorithm (Freeman et al. 2002) was used with different scale factors and thresholds, followed by visual inspection to remove spurious detections and add missed sources. They emerged with 283 X-ray sources, of which 248 were spatially associated with IR stars or radio sources indicating likely memberships in the Orion/Flame young stellar population. Our MYStIX analysis of this dataset obtained 547 X-ray sources, of which 422 are identified later as MPCMs (Broos et al. 2013).

NGC 2362. Damiani et al. (2006) obtained 387 sources using the PWDetect wavelet source-detection algorithm (Damiani et al. 1997) with a conservative threshold set for <1 spurious detection rate in the Chandra field of view. This is a single Chandra exposure without significant crowding, and thus does not present difficulties to traditional source detections. MYStIX obtained 690 sources, mainly due to the choice of a lower threshold for faint source existence, of which 467 are classified as MPCMs.

RCW 38. Wolk et al. (2006) and Winston et al. (2011) used PWDetect and wavdetect source-detection algorithms with a nominal false detection rate of <1% to find 518 sources. In a separate analysis, Evans et al. (2010) found 410 sources with ≥10 net counts using the Chandra Source Catalog algorithm. Our MYStIX analysis gives 1019 sources, nearly twice the number found by Winston et al. (2011) from the same dataset, of which 813 are classified as MPCMs.

The Trifid Nebula. Rho et al. (2004) made several runs of wavdetect with different wavelet scales on this image and a conservative faint source threshold to obtain 353 Chandra sources. The MYStIX analysis gives 633 sources, of which 418 are classified as MPCMs.

NGC 1893. Application of PWDetect with a threshold to give ~10 spurious sources resulted in 1025 Chandra sources (Caramazza et al. 2008, 2012). MYStIX analysis produced 1442 sources, of which 1110 are classified as MPCMs.

Here we make three detailed comparisons between previous Chandra source detections and the MYStIX procedures to further understand the source detection capabilities.
5. CONTAMINATING X-RAY POPULATIONS

Every Chandra exposure of a star-forming complex in the Galactic Plane will have three classes of unrelated X-ray sources, in addition to the young stellar population of interest: field stars in the foreground of the MSFR, more distant background stars, and extragalactic sources. The Monte Carlo simulations of their flux, median energy, and spatial distributions were performed using the procedures of Getman et al. (2011), with minor modifications described in Appendix A of Broos et al. (2013). Each MYStIX MSFR was treated individually with its unique mosaic of Chandra exposures, Galactic longitude and latitude, absorption for interstellar clouds, and distance to the complex. We use RCW 38 to illustrate the contamination here (Figure 5). The molecular clouds in this region are projected on the west side of the ACIS-I field of view, near the core of the young stellar cluster.

Extragalactic sources. Mostly active galactic nuclei with some starburst galaxies, extragalactic sources are seen through the obscuring Galactic interstellar medium (Broos et al. 2007). A typical MYStIX exposure should have dozens of extragalactic sources, and the deeper mosaics should have hundreds. Realistic X-ray flux distributions (Moretti et al. 2003) and spectra (Brandt et al. 2001) of extragalactic sources were simulated, and were then subject to spatially dependent absorption that was calculated using line-of-sight H I column density through the Galaxy (Dickey & Lockman 1990) and molecular cloud maps derived from CO observations or IR reddening of background stars (see Appendix A.2 in Broos et al. 2013). The simulated, absorbed sources were then superposed on the observed local background level and subject to the local MYStIX X-ray detection limit. The result of the simulation for RCW 38 is a prediction of ~120 extragalactic sources with a maximum source density of ~0.8 sources (arcmin)−2. Note from Figure 5 (left panel) that the extragalactic source spatial distribution is expected to be concentrated eastward of the center of the Chandra field, as the molecular cloud should obscure extragalactic sources westward of the field center. The contaminating source density is also higher toward the center of the field where the sensitivity to faint sources is higher.

Foreground Galactic sources. Magnetically active, Galactic field stars, mostly younger main sequence stars, populate the line-of-sight toward RCW 38 at d = 1.7 kpc. These mostly have soft X-ray spectra and do not show spatial variations due to absorption by the molecular clouds in the star-forming complex. The Monte Carlo calculation of their population is based on the Besançon Galactic structure model (Robin et al. 2003) and ROSAT surveys establishing X-ray luminosity functions as a function of stellar mass and age (Schmitt et al. 1995; Schmitt 1997; Hünsch et al. 1999). The simulation for the MYStIX RCW 38 observation predicts ~60 foreground stellar X-ray sources with peak surface density around 0.4 sources (arcmin)−2. These are assumed to be distributed uniformly across the sky, but, due to the higher sensitivity on-axis and intersecting CCD chip gaps in the ACIS-I detector, the expected
spatial distribution of foreground stars shows a distinct pattern (Figure 5, right panel).

**Background Galactic sources.** Field stars in the distant Galaxy are modeled including molecular cloud absorption similar to the extragalactic sources. Active and quiescent accretion binary systems are not included in the model. The simulation predicts ~30 background X-ray sources in the MYStIX X-ray catalog for RCW 38 with a peak surface density around 0.2 sources (arcmin)$^{-2}$ (Figure 5, center panel). While always fewer than the extragalactic contaminants, the predicted background Galactic source population depends strongly on Galactic latitude and longitude.

Thus, the simulation for the RCW 38 Chandra field predicts ~210 contaminating X-ray sources dominated by extragalactic sources. The contaminants typically have very little effect in the cores of rich stellar clusters, but can be a significant fraction of X-ray sources in peripheral regions of the Chandra fields. For example, in the northern corner of the RCW 38 field, about 1/3 of the X-ray sources are predicted to be contaminants.

These spatial maps of predicted contaminants, as well as their predicted X-ray median energies and counterpart $J$ magnitudes, are used in the “naive Bayes” classifier described by Broos et al. (2013). Table 8 of that paper gives the predicted numbers of each class of contaminants for each MYStIX star-forming complex. Generally, there is good agreement between the total number of predicted contaminants and the number of X-ray sources classified as contaminants or as “unclassified” due to insufficient information: for example, in RCW 38 the simulations predict 212 contaminants and the classifier identifies 206 sources as contaminants or “unclassified.” The classifier also identifies 813 Chandra sources as probable young stars in the RCW 38 complex. Thus, 20% of the X-ray sources in this field are likely contaminants and 80% are likely scientifically interesting young stars. The agreement between contamination predictions and classifier results across the MYStIX sample (see Table 8 of Broos et al. 2013) gives confidence that the predictions are reasonably reliable.

### 6. X-RAY “COLOR–MAGNITUDE DIAGRAMS”

Figure 6 shows the X-ray “color–magnitude diagram” (median energy, $\text{MedianEnergy}_J$, versus incident energy flux in the total band, $F_T$) for each of the 10 regions. Sources have been color-coded based on their classification as young stellar members (MPCMs), foreground stellar contaminants, background stellar contaminants, extragalactic contaminants, or unclassified sources by Broos et al. (2013). The distributions show that $F_T$ varies over more than three decades, while $\text{MedianEnergy}_J$ varies from $<1$ to $\sim7$ keV due to absorption. Absorption by interstellar gas increases $\text{MedianEnergy}_J$ and decreases $F_T$, shifting sources to the lower right.

Several effects can be seen in the distributions of the X-ray sources. Because detection is based on photon flux (photon cm$^{-2}$ s$^{-1}$) rather than energy flux (erg cm$^{-2}$ s$^{-1}$), an upward slope to the sensitivity limit as a function of $\text{MedianEnergy}_J$ is seen in the lower part of the diagram. Most of the “unclassified” X-ray sources from Broos et al. (2013) are located near this...
The MYStIX project is supported at Penn State by NASA grant NNX09AC74G, NSF grant AST-0968038, and the Chandra ACIS team contract SV4-74018 (G. Garmire & L. Townsley, Principal Investigators), issued by the Chandra X-Ray Center, which is operated by the Smithsonian Astrophysical Observatory for and on behalf of NASA under contract NAS8-03060. This research made use of data products from the Chandra Data Archive. This research has also made use of SAOImage DS9 software developed by Smithsonian Astrophysical Observatory.

REFERENCES

Brandt, W. N., Alexander, D. M., Hornschemeier, A. E., et al. 2001, AJ, 122, 2810
Broos, P., Townsley, L., Getman, K., & Bauer, F. 2012, Astrophysics Source Code Library, 1203.001 http://asteroscience.apod.com/viewtopic.php?f=35&t=27673 &hilit=acis
Broos, P. S., Feigelson, E. D., Townsley, L. K., et al. 2007, ApJS, 169, 353
Broos, P. S., Getman, K. V., Povich, M. S., et al. 2011a, ApJS, 194, 4
Broos, P. S., Townsley, L. K., Feigelson, E. D., et al. 2010, ApJ, 714, 1582
Broos, P. S., Townsley, L. K., Feigelson, E. D., et al. 2011b, ApJS, 194, 2
Broos, P. S., Getman, K. V., Povich, M. S., et al. 2013, ApJS, 209, 32
Caramazza, M., Micela, G., Prisinzano, L., et al. 2008, A&A, 488, 211
Caramazza, M., Micela, G., Prisinzano, L. et al. 2012, A&A, 539, A74
Dahm, S. E., Simon, T., Proszkow, E. M., & Patten, B. M. 2007, AJ, 134, 999
Damiani, F., Flaccomio, E., Micela, G., et al. 2004, ApJ, 608, 781
Damiani, F., Maggio, A., Micela, G., & Scintini, S. 1997, ApJ, 483, 350
Damiani, F., Micela, G., Scintino, S., et al. 2006, A&A, 460, 133
Davis, J. E., Bautz, M. W., Dewey, D., et al. 2012, Proc. SPIE, 8443, 84431A
Delgado, A. J., González-Martín, O., Alfaro, E. J., & Yun, J. 2006, ApJ, 646, 269
Dickey, J. M., & Lockman, F. J. 1990, ARA&A, 28, 215
Evans, I. N., Primini, F. A., Gol’tsef’’ly, K. J., et al. 2010, ApJS, 189, 37
Ezoe, Y., Kobukun, M., Makishima, K., Sekimoto, Y., & Matsuzaki, K. 2006, ApJL, 649, L123 Feigelson, E. D. 2010, Proc. NAS, 107, 7153 Feigelson, E. D., Getman, K., Townsley, L., et al. 2005, ApJS, 160, 379 Feigelson, E. D., Townsley, L.K., Broos, P.S., et al. 2013, ApJS, 209, 26 Flaccomio, E., Micela, G. & Scintino, S. 2006, A&A, 455, 903 Freeman, P. E., Kashyap, V., Rosner, R., & Lamb, D. Q. 2002, ApJ, 138, 185 Fruscione, A., McDowell, J. C., Allen, G. E., et al. 2006, Proc. SPIE, 6270, 62701V Garmire, G. P., Bautz, M. W., Ford, P. G., et al. 2003, Proc. SPIE, 4851, 28 Getman, K. V., Broos, P. S., Feigelson, E. D., et al. 2011, ApJS, 194, 3 Getman, K. V., Feigelson, E. D., Broos, P. S., Townsley, L. K., & Garmire, G. P. 2010, ApJ, 708, 1760 Getman, K. V., Feigelson, E. D., Broos, P. S., Townsley, L. K., & Garmire, G. P. 2012, Astrophysics Source Code Library, 1212.002, http://asteroscience.apod.com/viewtopic.php?f=35&t=30309 &p=188857 &p=188857

7. SUMMARY

We have described the identification of 11,315 faint X-ray sources in Chandra imaging observations of 10 massive star-forming regions in support of the MYStIX project. Together with the X-ray sources reported by L. K. Townsley et al. (in preparation) and previously published for three regions (Orion Nebula, W 40, and the Carina complex), they provide an empirical foundation to the MYStIX effort to combine multiwavelength samples of young stars that range from the youngest protostars to older, disk-free, pre-main-sequence stars. The X-ray source lists will also be used to excise point-source X-rays from the Chandra data to reveal diffuse X-ray emission from the winds and supernovae of massive stars in these regions.

The data analysis is designed to be highly sensitive with sources as faint as 3 counts. Candidate sources are found from maximum likelihood reconstruction of the Chandra images that accounts for the spatially variable point-spread function, and the candidates are listed as confirmed sources when they satisfy a probabilistic threshold based on the local background. Source events are extracted using local point-spread functions and are characterized using nonparametric techniques (Broos et al. 2010; Getman et al. 2010). The procedures are closely based on those used in the recent Chandra Carina Complex Project (Townsley et al. 2011). Codes for many of these methods are publicly available.

Our X-ray catalogs are thus considerably larger than most previously published catalogs from the same datasets, allowing more X-ray sources to be associated with IR sources. X-ray analysis is thus crucial for the production of rich samples classified as MPCMs (Broos et al. 2013). These MPCM samples can then be used in MYStIX science studies such as the identification and study of subclustering (Kuhn et al. 2013), derivation of X-ray/IR age estimators (K. V. Getman et al., in preparation), and many other future planned studies. On the X-ray “color–magnitude diagram” the locus of lightly absorbed MPCMs can be identified for some regions; however, it is difficult to distinguish between these sources and Galactic stellar contaminants or between highly absorbed MPCMs and extragalactic contaminants from the X-ray data alone.
