


# Dust scattering halo of 4U 1630–47: high resolution X-ray and mm observations constrain source and molecular cloud distances

E. Kalemci<sup>1</sup> ,<sup>1</sup>★ M. Diaz Trigo,<sup>2</sup> E. Oztaban,<sup>1</sup> A. A. Abbasi,<sup>1</sup> T. Stanke,<sup>3</sup> J. A. Tomsick,<sup>4</sup> T. J. Maccarone,<sup>5</sup> A. Saraçyakupoğlu,<sup>1</sup> E. von Nussbaum,<sup>6,7</sup> J. C. A. Miller-Jones<sup>8</sup> and B. Bahçeci<sup>1</sup>

<sup>1</sup>Faculty of Engineering and Natural Sciences, Sabancı University, Orhanlı-Tuzla 34956, Istanbul, Turkey

<sup>2</sup>ESO, Karl-Schwarzschild-Straße 2, D-85748, Garching bei München, Germany

<sup>3</sup>Max-Planck-Institut für Extraterrestrische Physik, Gießenbachstraße 1, 85748, Garching bei München, DE

<sup>4</sup>Space Sciences Laboratory, 7 Gauss Way, University of California, Berkeley, CA 94720-7450, USA

<sup>5</sup>Department of Physics & Astronomy, Texas Tech University, Box 41051, Lubbock, TX 79409-1051, USA

<sup>6</sup>Otto von Taube Gymnasium, Germeringer Straße 31, D-82131 Gauting, DE

<sup>7</sup>Technische Universität München, D-80333 München, DE

<sup>8</sup>International Centre for Radio Astronomy Research–Curtin University, GPO Box U1987, Perth, WA 6845, Australia

Accepted 2025 September 25. Received 2025 September 25; in original form 2025 April 10

## ABSTRACT

We re-investigated the distance to the black hole X-ray binary 4U 1630–47 by analysing its dust scattering halo (DSH) using high-resolution X-ray (*Chandra*) and millimeter (*APEX*) observations. Dust scattering haloes form when X-rays from a compact source are scattered by interstellar dust, creating diffuse ring-like structures that can provide clues about the source’s distance. Our previous work suggested two possible distances: 4.9 and 11.5 kpc, but uncertainties remained due to low-resolution CO maps. We developed a new methodology to refine these estimates, starting with a machine learning approach to determine 3D representation of molecular clouds from the *APEX* data set. The 3D maps are combined with X-ray flux measurements to generate synthetic DSH images. By comparing synthetic images with the observed *Chandra* data through radial and azimuthal profile fitting, we not only measure the source distance but also distinguish whether the molecular clouds are at their near- or far-distances. The current analysis again supported a distance of 11.5 kpc over alternative estimates. While the method produced a lower reduced  $\chi^2$  for both the azimuthal and radial fits for a distance of 13.6 kpc, we ruled it out as it would have produced a bright ring beyond the *APEX* field of view, which is not seen in the *Chandra* image. The 4.85 kpc estimate was also excluded due to poor fit quality and cloud distance conflicts. The systematic error of 1 kpc, which arises due to errors in determining molecular cloud distances, dominates the total error.

**Key words:** dust, extinction – X-rays: binaries – X-rays: individual: 4U 1630–47.

## 1 INTRODUCTION

X-ray transients in outburst may span orders of magnitude in luminosity range and may temporarily become the brightest X-ray objects in the sky (E. Kalemci, E. Kara & J. A. Tomsick 2022). There is clear evidence for super-Eddington accretion in some ultraluminous X-ray sources (M. Bachetti et al. 2014; M. Brightman et al. 2019), with some debate over whether this is due to bona fide super-Eddington accretion or beaming. There is also tentative evidence that some Galactic black hole X-ray binaries reach super-Eddington phases (M. P. Muno, E. H. Morgan & R. A. Remillard 1999; M. Revnivtsev et al. 2002; P. Uttley & M. Klein-Wolt 2015; P. Jin et al. 2024). However, uncertainties in distance, mass, and inclination angle prevent firm conclusions, with only the spectral properties being fully established in systems at high luminosities in a way that is consistent with radiation pressure being important.

Accurate distance measurements are not only the key to establishing reliable Eddington luminosity fractions, but they are also essential in determining the size scales, inclination angles and speeds of jets observed in black hole systems (B. J. Burridge et al. 2025).

Most of the X-ray transients are in the Galactic plane behind copious amounts of dust and gas (see J. M. Corral-Santana et al. 2016, for both distance estimates and Galactic distribution). The extinction due to high column density makes it difficult to determine the distances using methods relying on optical telescopes. There are indirect methods to measure distances such as using state transition luminosities in X-rays (T. J. Maccarone & P. S. Coppi 2003; A. Vahdat Motlagh, E. Kalemci & T. J. Maccarone 2019), or other X-ray spectral analysis methods (Y. Abdulghani, A. M. Lohfink & J. Chauhan 2024). Radio observations may be used as well, including jet-parallax (J. C. A. Miller-Jones et al. 2009; P. Atri et al. 2020), which is model-independent, and H I absorption lines (J. Chauhan et al. 2021; B. J. Burridge et al. 2025). R. M. Arnason et al. (2021) discuss how the methods fare when compared to the optical parallax measurements with the *Gaia* mission. For bright sources behind

\* E-mail: [ekalemci@sabanciuniv.edu](mailto:ekalemci@sabanciuniv.edu)

moderate to high dust concentration, one can also analyse the radial profile of the halo caused by the scattering of X-rays from dust and determine the source distance in this manner (J. Trümper & V. Schönfelder 1973; P. Predehl et al. 2000; T. W. J. Thompson & R. E. Rothschild 2009; J. Xiang et al. 2011; S. Heinz et al. 2015).

In E. Kalemci, T. J. Maccarone & J. A. Tomsick 2018 (hereafter K18), we utilized *Chandra*, *Swift*, and low-resolution CO maps to constrain the distance to the black hole transient 4U 1630–47 by fitting the dust scattering halo (DSH) radial profile. We reported two possibilities: 11.5 kpc and 4.9 kpc, favouring 11.5 kpc with the main scattering molecular cloud MC-79 being at the far distance estimate obtained using a Galaxy rotation curve model. In the same study, we asserted that the resolution of CO maps is the source of the main uncertainty in the method. Recently, to overcome the difficulties caused by low-resolution images, we obtained high-resolution mm maps of the region with the Atacama Pathfinder Experiment (*APEX*; R. Güsten et al. 2006) and developed an imaging-based method to determine both the source distance and the order of molecular clouds in the line of sight. In this article, we describe the methodology and present its first results applied to 4U 1630–47.

### 1.1 4U 1630–47

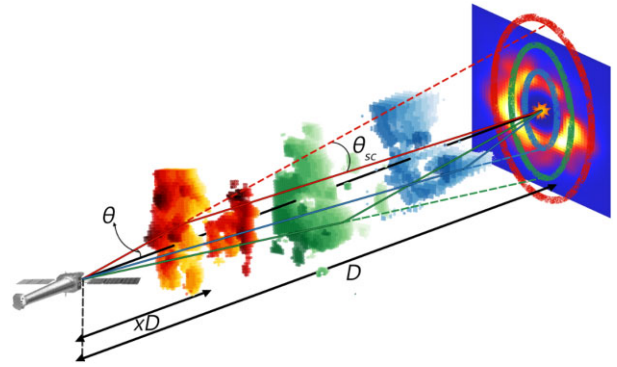
After its discovery in 1976 (C. Jones et al. 1976), 4U 1630–47 has been classified as a low-mass X-ray binary (LMXB), with recurrent quasi-periodic outbursts every 2–3 yr. It was first classified as a black hole from its spectral and timing properties by D. Barret, J. E. McClintock & J. E. Grindlay (1996). It is a peculiar Galactic black hole transient (GBHT), often not following the typical hysteresis pattern in the hardness-intensity diagram (Y. Abe et al. 2005; J. A. Tomsick et al. 2005).

The source is located in the Galactic Plane with a high and varying absorption column density within and between spectral states (T. Augusteijn, E. Kuulkers & M. H. van Kerkwijk 2001; J. A. Tomsick et al. 2014), making accurate estimates of its distance difficult. Through *IXPE* observations, 4U 1630–47 has been revealed to have high polarization in X-rays (H. Krawczynski et al. 2024). The distance to the source directly affects the interpretation of these measurements, as an accurate estimate of the distance can improve the reliability of the models describing the accretion dynamics and the observed polarization of the source.

4U 1630–47 has also shown a very low luminosity ( $L/L_{\text{Edd}} = 0.0003$ ) soft state (J. A. Tomsick et al. 2014). Based on the iron line features simultaneously observed with radio emission, a baryonic content for the jet has been proposed (M. Díaz Trigo et al. 2013), but other explanations remain (J. Neilsen et al. 2014; Y. Wang & M. Méndez 2016). In K18, we tested the idea that some of the peculiarities of the source can be explained by local dust scattering through analysis of the dust scattering halos of the source observed with *Chandra* and *Swift*.

### 1.2 Dust scattering haloes

Dust scattering halos (J. W. Overbeck 1965; J. S. Mathis & C.-W. Lee 1991; P. Predehl & J. H. M. M. Schmitt 1995) are diffuse rings of light that form when X-rays are scattered by interstellar dust grains along the line of sight to a distant source. In addition to the determination of the source distance mentioned above, studies of these halos have provided valuable information on spectral variations during eclipses (M. D. Audley et al. 2006), the physical characteristics of dust grains and their distribution along the line of sight (J. Xiang et al. 2011;



**Figure 1.** Figure depicting the formation of dust scattering haloes using the 3D shapes of the molecular clouds. The source distance is given by  $D$ , and the distance to the dust cloud from the observer is given by  $xD$ .  $\theta$  is the observed scattering angle,  $\theta_{\text{sc}}$  is the physical scattering angle. The distances of 3D clouds in this figure are for better representation and do not coincide with actual distances in this work.

L. R. Corrales & F. Paerels 2015), and the measurement of X-ray extinction (P. Predehl & J. H. M. M. Schmitt 1995; L. R. Corrales et al. 2016). For a recent review, see E. Costantini & L. Corrales 2022.

When a source experiences outbursts (or flares) followed by an extended period of quiescence, the dust-scattered emission appears as distinct ring-like structures. This occurs because dust along the line of sight is concentrated in molecular clouds, and each ring structure corresponds to the delayed scattered emission from a single region of high dust density (see Fig. 1). The delay is a result of the scattered X-rays taking a longer path compared to the direct X-rays observed by the telescope (S. Heinz et al. 2016, and references therein). The DSH appears as uniform rings only if the scattering is also caused by uniform-density dust in the clouds. As explored in this work, the non-uniformity of clouds will be reflected in the azimuthal distribution of emission and absorption in the rings.

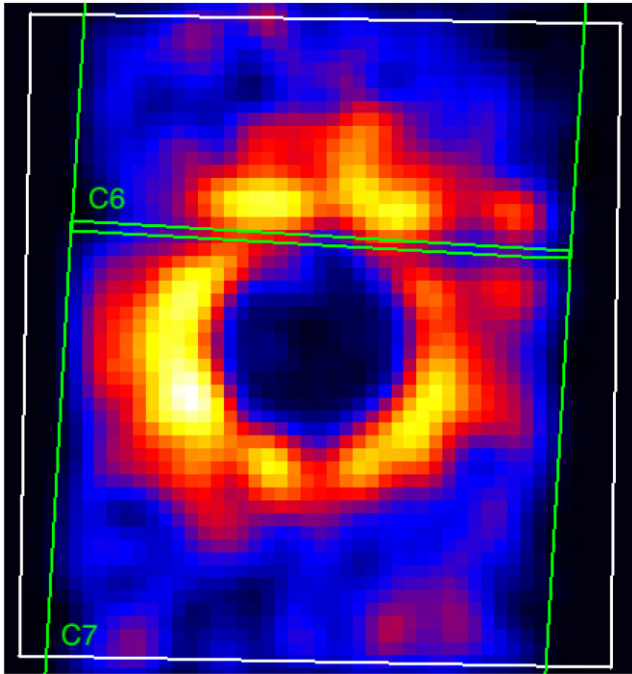
## 2 OBSERVATIONS AND ANALYSIS

In this work, we utilized data from several observatories: in X-rays *Chandra*, *Swift*, and *MAXI* and in mm band *APEX*. We also used publicly available CO maps from the Bronfman survey (L. Bronfman et al. 1989) for a larger coverage of the molecular cloud distribution and the Southern Galactic Plane Survey (SGPS; N. M. McClure-Griffiths et al. 2005) to estimate the column density of neutral atomic hydrogen.

### 2.1 X-ray imaging analysis

We used the same data set in K18 from *Chandra* (obsid 19004) to recreate the main dust scattering halo image. Fig. 2 shows the background subtracted, point sources removed, and rebinned *Chandra* image (matching the *APEX* resolution) in the 2.25–3.15 keV band. The same figure also shows the *Chandra* ACIS-S chip boundaries as well as the *APEX* field of view. A larger image in three energy bands can be found in K18.

The details of the *Chandra* analysis to obtain the halo image are given in K18, except for one particular difference. In K18, we used blank-sky images to determine the diffuse X-ray background in each band. The resulting profile showed a flat surface brightness up to 80 arcsec. However, none of our generated images show emission



**Figure 2.** The rebinned and background-subtracted *Chandra* image in the 2.25–3.15 keV band. All point sources (including 4U 1630–47) have been removed. A Gaussian smoothing with 3 pixel radius was applied. The white box shows the APEX field of view, and green boxes show *Chandra* ACIS-S chip boundaries with chip numbers shown at the lower left corners.

inside 80 arcsec, the  $\Delta t$  values (see equation 4) corresponding to the scattering angles  $<80$  arcsec are small, and the corresponding input flux is very low. However, as discussed in Section 4.4.3, some of the excess emission may be caused by double scattering, which is not negligible. The blank-sky method possibly underestimates the background in surface brightness. Therefore, in Chip 7, we picked an annular region between 30 and 60 arcsec around the source, and in Chip 6, a circular region of size 20 arcsec away from the halo to determine new background figures in each chip.

## 2.2 X-ray spectral analysis and light curve

Another key difference between K18 and this work lies in the treatment of X-ray flux that interacts with molecular clouds. Unlike K18, where clouds are modelled as homogeneous spheres, this study adopts a more detailed approach, requiring a precise determination of extinction per pixel and scattering angle (see equation 3 and Fig. 1). This means computing the unabsorbed flux from the source that passes through the molecular clouds, accounting for both scattering and absorption effects more rigorously.

We used two sets of data to achieve our goal: the *MAXI* 2–4 keV light curve, which covers the entire 2016 outburst; and pointed *Swift* XRT observations, covering both the first half of the outburst in the windowed timing (WT) mode (MJD 57630.5–57682.5, see Table 1) and the decay part (MJD 57769.8–57777.4), mostly in the Photon Counting (PC) mode (see K18). We used High Energy Astrophysics Software (HEASOFT) v6.31 and produced photon event lists and exposure maps using *xrtpipeline*. For both WT and PC observations, we extracted photons from a circular region with a 20-pixel (47 arcsec) radius centred on the source. For WT mode observations, the background region is an annulus between 90–110 pixels (211.5–258.5 arcsec). An appropriate

**Table 1.** *Swift* observations used in the analysis.

Date (MJD)	obsid	$N_H$	$T_{in}$
57630.5	00031 224 020	$13.99 \pm 0.21^*$	$1.295 \pm 0.025$
57635.3	00031 224 022	$14.72 \pm 0.74$	$1.554 \pm 0.106$
57636.0	00031 224 023	$13.77 \pm 0.14$	$1.481 \pm 0.021$
57637.1	00031 224 024	$14.29 \pm 0.32$	$1.495 \pm 0.048$
57638.3	00031 224 025	$13.14 \pm 0.19$	$1.582 \pm 0.033$
57639.3	00031 224 026	$11.10 \pm 0.66$	$1.359 \pm 0.111$
57640.0	00031 224 027	$9.060 \pm 1.10$	$1.284 \pm 0.320$
57641.5	00031 224 028	$12.77 \pm 1.05$	$1.355 \pm 0.141$
57642.5	00031 224 029	$13.37 \pm 0.14$	$1.565 \pm 0.023$
57643.4	00031 224 030	$14.07 \pm 0.16$	$1.620 \pm 0.026$
57644.3	00031 224 031	$14.53 \pm 0.24$	$1.583 \pm 0.036$
57645.9	00031 224 032	$14.35 \pm 0.15$	$1.677 \pm 0.027$
57648.2	00031 224 033	$10.83 \pm 0.48$	$1.349 \pm 0.079$
57654.1	00031 224 035	$14.87 \pm 0.25$	$1.541 \pm 0.037$
57657.9	00031 224 036	$13.85 \pm 0.24$	$1.470 \pm 0.038$
57658.5	00031 224 037	$14.41 \pm 0.41$	$1.662 \pm 0.061$
57660.9	00031 224 038	$13.30 \pm 0.18$	$1.519 \pm 0.028$
57663.6	00031 224 039	$13.58 \pm 0.16$	$1.503 \pm 0.024$
57667.1	00031 224 040	$14.03 \pm 0.17$	$1.650 \pm 0.029$
57669.5	00031 224 041	$13.04 \pm 0.16$	$1.577 \pm 0.027$
57672.4	00031 224 042	$12.40 \pm 0.41$	$1.672 \pm 0.080$
57675.6	00031 224 043	$13.12 \pm 0.85$	$1.452 \pm 0.089$
57678.3	00031 224 044	$14.10 \pm 0.14$	$1.741 \pm 0.027$
57682.5	00031 224 045	$14.37 \pm 0.18$	$1.765 \pm 0.033$

\*Errors are at  $1\sigma$ .

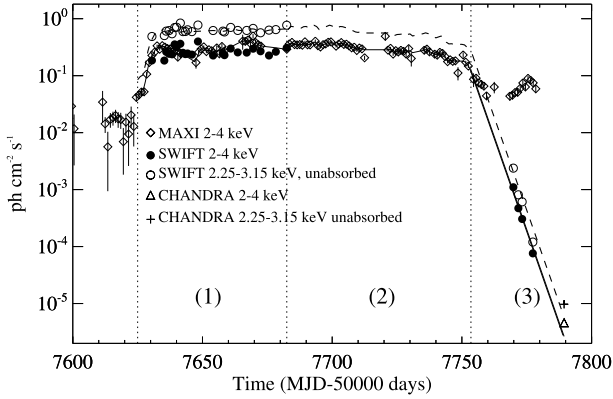
background scaling is applied. The PC spectra for the observations during the decay have already been produced as described in K18.

The unabsorbed light curve estimation is done in three distinct regions: (1) MJD 57625–57682 where both *MAXI* and *Swift* XRT observations are present, (2) MJD 57683–57753 where only *MAXI* data are present, and (3) MJD 57754–57789, which is the decay part with *Swift* XRT observations reaching to the *Chandra* observation. In this region, the *MAXI* light curve is not reliable, possibly due to the nearby bright source GX 340+0 affecting the *MAXI* background.

We first extracted spectra from all *Swift* observations in region (1) and fitted them with an absorbed diskbb model (*tbabs*  $\times$  *diskbb* with *wilm* abundances; J. Wilms, A. Allen & R. McCray 2000). No other component was required in the fits. We then obtained 2–4 keV fluxes. By comparing those with the *MAXI* fluxes in the same energy band, we obtained a cross-calibration factor. We also obtained unabsorbed light curves in 1.5–2.25 keV (E1), 2.25–3.15 keV (E2), and 3.15–5 keV (E3) using the same *Swift* observations. Since there was minimal spectral evolution and no sign of state transition throughout the flat part of the outburst (see Table 1), we obtained a scaling factor by dividing the average unabsorbed *Swift* fluxes by the average *MAXI* 2–4 keV flux.

In region (2), we simply multiplied *MAXI* 2–4 keV fluxes with the scaling and cross-calibration factors obtained in region (1). This region ends at MJD 57753 beyond which the *MAXI* background becomes unreliable.

In region (3), we fitted an exponentially decaying light curve in E1, E2, and E3 that passes through the unabsorbed fluxes of existing *Swift* observations to the *Chandra* unabsorbed flux. For regions (1) and (2), we smoothed out sudden dips and peaks most probably caused by poor exposure and background subtraction in *MAXI*. The absorbed and unabsorbed light curves are shown in Fig. 3. Only E2 is shown for clarity.



**Figure 3.** *MAXI* 2–4 keV and constructed unabsorbed 2.25–3.15 keV (E2) light curves. (1) to (3) denote regions in the light curve separated by vertical dashed lines. The solid line is a smoothed absorbed light curve, and the thick dashed-line is the smoothed unabsorbed 2.25–3.15 keV (E2) light curve.

### 2.3 APEX data and analysis

We performed spectral line mapping observations of the area around 4U 1630–47 using the *APEX* 12 m telescope (R. Güsten et al. 2006) located on Chajnantor at an elevation of 5100 m with the 230 GHz channel of the nFLASH receiver. nFLASH is a dual-sideband receiver, providing  $2 \times 8$  GHz bandwidth with a  $\sim 8$  GHz gap in between. For our observations the receiver was tuned to cover frequencies between  $\sim 229$ – $237$  GHz (in the upper side band–USB) and between  $\sim 213$ – $221$  GHz (in the lower side band–LSB), including the frequencies of the  $J = (2-1)$  main CO isotopologue lines:  $^{12}\text{CO}(2-1)$ ,  $^{13}\text{CO}(2-1)$ , and  $\text{C}^{18}\text{O}(2-1)$  at rest-frequencies of 230.538000, 220.398684, and 219.560358 GHz, respectively. The spectra were recorded using the *APEX* FFT spectrometer at a spectral resolution of  $\sim 244.1$  kHz/ $\sim 0.32$  km s $^{-1}$  in 2021 and at the native spectral resolution of  $\sim 61$  kHz/ $\sim 0.079$  km s $^{-1}$  in 2022.

Observations were taken on 2021, May 08, and in 2022 on July 09/10/11 and on October 11, all in good to very good weather conditions (precipitable water vapour between 0.8 mm and 3.4 mm). In 2021, a  $5 \times 5$  arcmin map centred on the position of 4U 1630–47 was taken, while in 2022 a larger  $6 \times 9$  arcmin region (oriented N–S with a position angle of  $3.4^\circ$  west-of-north) was observed. Mapping was done in the On-The-Fly (OTF) mode, continuously scanning across the map region while dumping spectra at positions spaced by  $\sim 1/3$  of the beam width ( $\sim 30$  arcsec); correspondingly, spacings between scan rows were also set to  $\sim 1/3$  of the beam width. Several map coverages were required to reach the final exposure, where the scanning direction was alternated between east–west and north–south in order to reduce scanning artifacts. Atmospheric calibrations were interspersed every few minutes, while pointing checks were done every 1–2 hours.

Data reduction followed standard procedures within the GILDAS/CLASS software<sup>1</sup>. Low-order polynomial baselines were fitted and subtracted from the online-calibrated spectra, where spectral ranges showing clear line emission were excluded from the fits. Finally the spectra were gridded into cubes, one for each of the three lines, with a spectral resolution of  $0.5$  km s $^{-1}$  and 500 channels spanning the velocity range from  $-175$  to  $+75$  km s $^{-1}$ . The spectra were then fitted with Gaussians to determine the central radial velocities of the molecular clouds (see Fig. 4, top two panels).

<sup>1</sup>see <https://www.iram.fr/IRAMFR/GILDAS>

### 2.4 SGPS analysis to obtain $N_{\text{HI}}$ distribution

We use publicly available data from the Southern Galactic Plane Survey (SGPS) to estimate the total column density corresponding to neutral atomic hydrogen along the line of sight to 4U 1630–47. The spectral data can be decomposed into 11 Gaussian components, spanning radial velocities from  $-150$  to  $50$  km s $^{-1}$  (Fig. 4, bottom panel). To find the total column density corresponding to neutral atomic hydrogen along the line of sight to the source, we can integrate the brightness temperature over the velocity range as follows (B. T. Draine 2011):

$$N_{\text{HI}} = 1.823 \times 10^{18} \int T_{\text{B,HI}}(v_{\text{LSR}}) dv \text{ cm}^{-2} \quad (1)$$

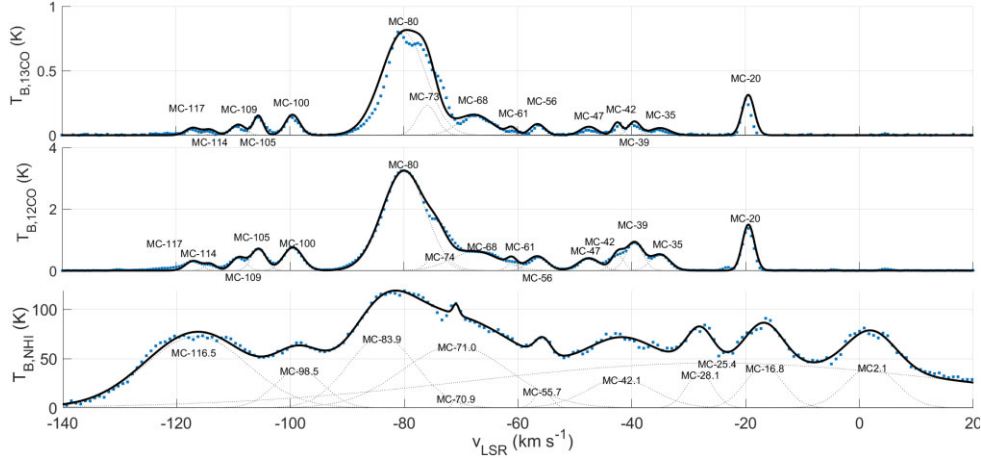
Note that this holds for the assumption of optically thin emission. For denser areas of H I, this provides a lower limit since self-absorption of the H I 21 cm line can occur (HI4PI Collaboration 2016). Integrating over the whole velocity spectrum via equation (1) yields a total neutral atomic hydrogen column density of  $2 \times 10^{22}$  cm $^{-2}$ , which agrees with the lower limit found by T. Augusteijn et al. (2001).

### 2.5 Determining cloud shapes

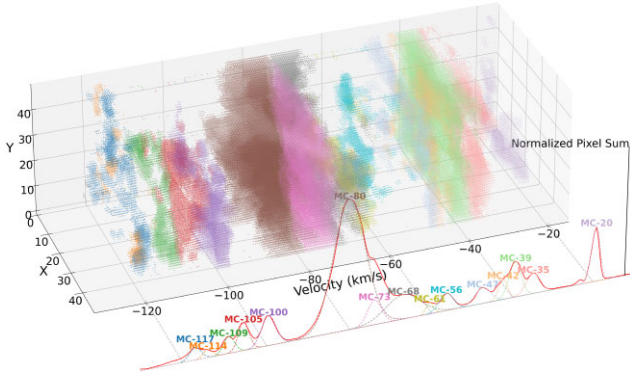
We used millimeter data from 500 high-resolution *APEX* velocity channel maps to determine the 3D representations and densities of molecular clouds along the line of sight. Because of the small field of view and overlapping data, we developed a novel set of algorithms and test functions to determine the cloud shapes as explained in detail in the Appendix, and here we provide a brief summary. The analysis starts with Gaussian decomposition to identify peak radial velocities and velocity dispersions of molecular clouds from the  $^{12}\text{CO}$  brightness spectra. After the images are segmented into velocity groups, thresholding is applied to isolate cloud regions. Finally, a modified Mean-Shift clustering algorithm groups data points into 15 distinct clouds and reconstructs their 3D representations with densities (see Fig. 5). We name the molecular clouds with peak radial velocities such as ‘MC-80’. In Fig. 6, we show two representative *APEX* images, obtained through integrating brightness over the velocity slices of the given clouds to highlight the advantage of having high-resolution images. These images also show the area of the main DSH ring, as well as the low-resolution Bronfman Survey (L. Bronfman et al. 1989) pointing. In K18, MC-68 was particularly discussed as it was utilized to fit the profile for radii smaller than 80 arcsec. There was also some discussion about whether the source is inside this cloud. With the high resolution image, we can see that MC-68, while relatively bright, only covers a small area with a small intersection with the main ring; therefore, the profile at small radii is probably background dominated (see Section 2.2), and there is no strong presence of this cloud at the source position.

Each centroid radial velocity corresponds to a near and far distance for each cloud. These distances are ascertained using the Galactic rotational curve model used by M. J. Reid et al. (2014) similar to K18. Table 2 summarizes the decomposition of  $^{12}\text{CO}$  data into molecular clouds. The values for  $W(\text{CO})$  are found by integrating the average temperature in each slice of each cloud. We note that the  $W(\text{CO})$  values are lower than for those reported in K18 for the same clouds. We found that intrinsically, the *APEX* brightness temperatures are lower than those obtained from the Bronfman data (they are also not covering exactly the same area as shown in Fig. 6). Moreover the thresholding removes an additional 10 per cent–15 per cent.

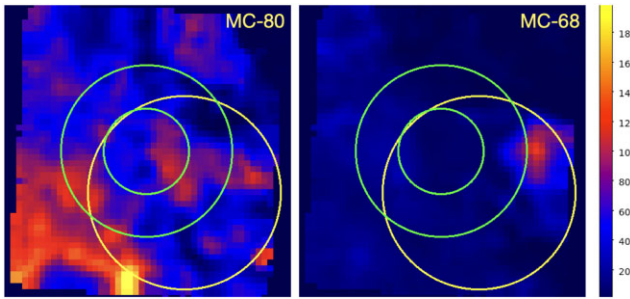




**Figure 4.** Top:  $^{13}\text{CO}$  spectrum from APEX. Middle:  $^{12}\text{CO}$  spectrum from APEX. Bottom:  $N_{\text{HI}}$  spectrum from the Southern Galactic Plane Survey (SGPS) towards 4U 1630–47.  $T_{B,\text{HI}}$ ,  $T_{B,12\text{CO}}$ , and  $T_{B,13\text{CO}}$  are the brightness temperatures of the emission lines. Gaussian decomposition is applied to all spectra.



**Figure 5.** The foreground shows the velocity spectrum with fitted Gaussians, while the background displays the 3D shapes of molecular clouds based on  $x$ ,  $y$ , and velocity coordinates. The colours and the positions of the Gaussians and 3D shapes are matched.



**Figure 6.** APEX integrated cloud images for MC-80 (left) and MC-68 (right). The annulus at the center shows the position of the DSH ring. The circle at the bottom right represents the Bronfman Survey pointing.

## 2.6 DSH image generation

To generate each DSH image for any given source distance, the following algorithm is used:

(i) Generate the permutation of near and far distances for all clouds, then sort them to consider only the ones in front of the source along the line of sight to the observer. This creates a 3D voxel array structure spanning the entire observational path.

**Table 2.** Cloud parameters.

No	Name	Near (kpc)	Far (kpc)	$W(\text{CO})^a$ (K km s $^{-1}$ )	$N_{\text{H}}^b$ $10^{22}$ cm $^{-2}$
1	MC-20	1.82	13.87	3.66	0.15
2	MC-35	2.71	12.91	1.91	0.076
3	MC-39	2.92	12.68	4.39	0.18
4	MC-42	3.07	12.52	1.10	0.044
5	MC-47	3.31	12.26	1.28	0.051
6	MC-56	3.72	11.83	1.46	0.058
7	MC-61	3.93	11.60	1.38	0.055
8	MC-68	4.21	11.31	6.98	0.28
9	MC-73	4.40	11.10	7.07	0.28
10	MC-80	4.67	10.83	29.49	1.18
11	MC-100	5.38	10.09	2.99	0.12
12	MC-105	5.55	9.91	2.49	0.099
13	MC-109	5.69	9.76	1.19	0.048
14	MC-114	5.87	9.58	0.72	0.029
15	MC-117	5.98	9.47	0.94	0.038

<sup>a</sup>Integrated  $^{12}\text{CO}$  emission ( $J = 1-0$ ).

<sup>b</sup>Estimated  $N_{\text{H}}$  from the Bolatto relation (A. D. Bolatto, M. Wolfire & A. K. Leroy 2013).

(ii) Estimate total  $N_{\text{H}}$  for each voxel.

For each voxel, the total hydrogen column density  $N_{\text{H}}$  is calculated by combining the molecular and neutral atomic hydrogen components. For the molecular contribution, we use  $N_{\text{H}_2}(x_p, y_p, z) = X_{\text{CO}} T_{B,12\text{CO}}(x_p, y_p, z) 0.5 \text{ km s}^{-1}$  with the conversion factor  $X_{\text{CO}} = 2 \times 10^{20} \text{ cm}^{-2} (\text{K km s}^{-1})^{-1}$  (A. D. Bolatto et al. 2013).  $x_p$ ,  $y_p$ , and  $z$  denotes the  $x$ ,  $y$  positions of the scattering voxel in the APEX slice  $z$ , respectively.

The neutral atomic hydrogen column density is split into two parts: a voxel-dependent component  $N_{\text{H,W}}(x_p, y_p, z)$  proportional to the  $N_{\text{H}_2}(x_p, y_p, z)$ , representing the atomic hydrogen associated with the molecular clouds, and a uniform component  $N_{\text{H,U}}(z)$  distributed evenly along the line of sight between 2.35–12.9 kpc. The second component corresponds to the neutral atomic hydrogen associated with the spiral arms, and the distances indicate the start and end of the spiral arms in the direction of 4U 1630–47. In summary, the total hydrogen column density for each pixel can be written as

$$N_{\text{H}}(x_p, y_p, z) = 2N_{\text{H}_2}(x_p, y_p, z) + N_{\text{H,W}}(x_p, y_p, z) + N_{\text{H,U}}(z) \quad (2)$$

See Section 4.4 for a discussion of the choice of  $N_H$  distribution.

(iii) Model the extinction of photons in two stages: from the source to the scattering voxel, and from the scattering pixel to the observer. The extinction coefficient for the pixel  $(x_p, y_p)$  at slice  $z$  is:

$$K_z(x_p, y_p) = \left( \sum_{z,as} N_H(x_p, y_p, z) + \sum_{z,bs} N_H(x_a, y_a, z) \right) \times \sigma_{ph,E} \quad (3)$$

where  $\sigma_{ph,E}$  is the energy-dependent absorption cross-section, and  $x_a, y_a$  denote the positions of the absorbing voxels at each slice before scattering. The integrations follow the three-dimensional photon path through the dust and gas distribution. The first sum considers the extinction undergone by the photons from the scattering pixel to the observer ( $z, as$ : sum over slices after scattering). Since, after scattering, the photons propagate along the same direction used to generate the projected image (see Fig. 1), the  $x_p$  and  $y_p$  indices in each slice remain identical to those in the projected image. On the other hand, the second summation accounts for the extinction experienced by the photons along their trajectory from the source to the scattering pixel ( $z, bs$ : sum over slices before scattering). The pixel indices corresponding to each slice along this path are determined using spherical geometry, based on the distance of the slice and the distance to the source. To determine  $\sigma_{ph,E}$ , we utilized X-ray spectral fits and compared absorbed and unabsorbed fluxes in each energy band. There may also be some local absorption at the source. Such absorption will enter the calculation as  $e^{-N_{H,local}\sigma_{ph,E}}$  for all pixels; hence it just changes normalization. See Section 4.4.2 for a discussion of local absorption.

(iv) Determine the time delay of arrival of scattered photons,  $\Delta t$ :

$$\Delta t = \frac{x D \theta^2(x_p, y_p)}{2c(1-x)} \quad (4)$$

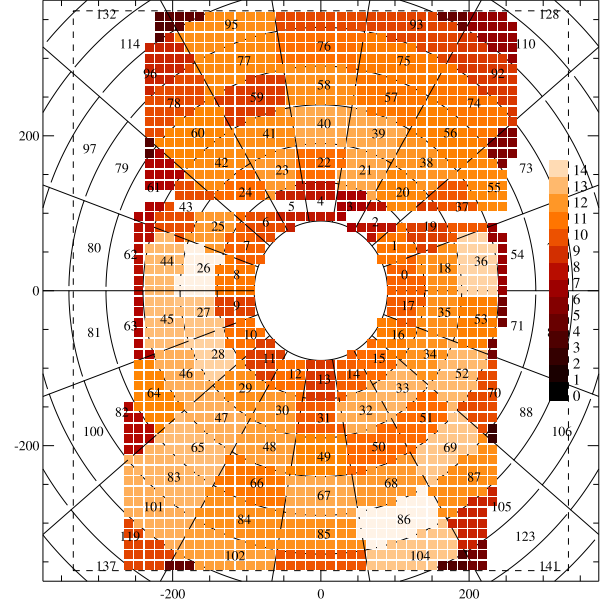
where  $x$  is the ratio between the voxel distance  $d$ , and source distance  $D$  ( $x = d/D$ ),  $c$  is the speed of light (see Fig. 1).

(v) Determine dust scattering halo intensity  $I_v(x_p, y_p)$  by integrating input flux as a function of time and scattering cross section per Hydrogen column density over all slices while taking absorption into account:

$$I_v(x_p, y_p) = \sum_z N_H(x_p, y_p, z) \frac{d\sigma_{sc,E}}{d\Omega} \frac{F_v(t = t_{obs} - \Delta t)}{(1-x)^2} \times \exp[-K_z(x_p, y_p)] \quad (5)$$

The multiplicative term  $N_{H(x_p, y_p, z)}$  is determined via equation (2). The term  $\frac{d\sigma_{sc,E}}{d\Omega} \frac{1}{(1-x)^2}$  is found for each  $x$  and  $\theta$  using the latest version of the NEWDUST code (L. Corrales 2023). This term corresponds to the *norm\_int* in NEWDUST, which is the surface brightness divided by the X-ray flux. NEWDUST uses a custom-defined grid of energy and  $\theta$  as inputs. See Section 4.4.3 for a discussion of the NEWDUST cross sections. We chose Rayleigh-Gans dust model with Drude approximation (L. R. Corrales et al. 2016), and an *md* parameter (dust mass column in  $\text{g cm}^{-2}$ ) of 0.000232 per  $10^{22} \text{ cm}^{-2}$  of  $N_H$  which corresponds to a dust-to-mass ratio of 0.01 (L. Corrales, personal communication). The *norm\_int* values are computed for  $\theta$  ranging from 0 to 600 arcsec with  $x$  values ranging from 0.001 to 0.999 with 0.001 resolution for 3 energy bands (E1, E2, E3).

(vi) Add the World Coordinate System (WCS) and save the generated image as a FITS file.



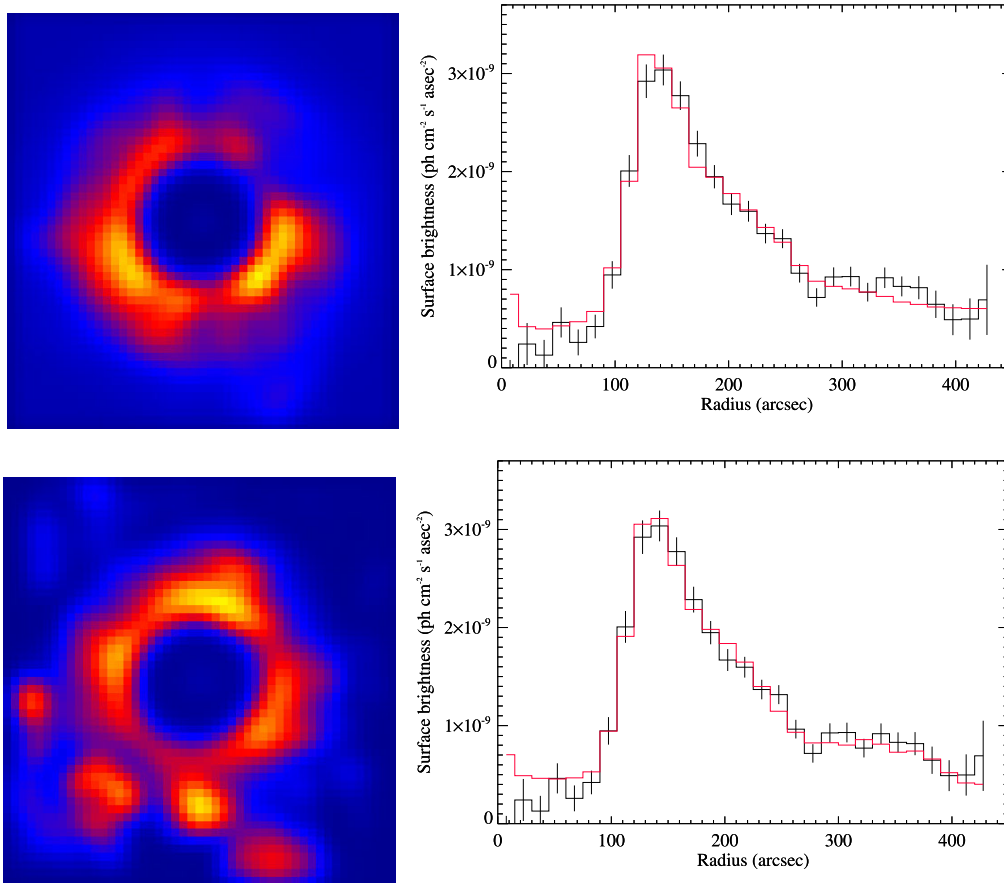
**Figure 7.** The *Chandra* and generated images are segmented with the wedge-shaped bins as shown here. The dashed lines roughly represent the *APEX* field of view. The brightness represents the signal-to-noise ratio in *Chandra* E2 band in the given wedge.

## 2.7 Generated image post-processing

To compare the generated images with the *Chandra* image, we first removed detected point sources from the *Chandra* images in the given energy bands as described in K18. We then rebinned the *Chandra* images to match the resolution of the *APEX* images. We determined pixels with low exposures (<50 percent of the maximum, mostly corresponding to the chip boundaries) using the *Chandra* exposure map, flagged and excluded them from the analysis. Fig. 2 is the resulting image in the E2 band we used for direct comparison with the generated images.

We created radial surface brightness profiles (SBP) to compare with our previous results in K18. We summed photon fluxes inside radial rings of thickness 15 arcsec, and divided them by the angular area of each ring to obtain the SBP. The largest ring ends at 420 arcsec to cover the entire *APEX* field of view. The flagged pixels were removed from both the sum and the areas. The same procedure is applied to the generated images. The errors in the *Chandra* radial profile are calculated assuming Poisson statistics. We then fitted the *Chandra* profiles with the generated image profiles with two parameters: a normalization and a constant. While the normalization was free, the constant was allowed to vary between  $-0.3 \times 10^{-9}$ – $8 \times 10^{-9} \text{ ph cm}^{-2} \text{ s}^{-1} \text{ arcsec}^{-2}$  as it represents the systematic error in determining the background (see Section 2.1). For the fitting process, we used  $\chi^2$  minimization with the *MPFIT* package in IDL (C. B. Markwardt 2009).

To take advantage of the high-resolution *APEX* images, we also produced azimuthal (or wedge) profiles. This is done by first dividing the radial region of interest into 8 rings of size 50 arcsec, and then dividing each ring into 18 azimuthal bins. After removing flagged pixels in *Chandra* images with low exposure, we calculated the signal-to-noise ratio (SNR) in each bin, again assuming Poisson statistics. Finally, we made a cut on the SNR ( $>9$ ) to obtain a ‘wedge profile’ with each wedge number corresponding to one of the bins in the image (see Fig. 7). We constructed the wedge profile for each generated image with the same bins, and again fitted, using



**Figure 8.** Top: The best fit generated image and radial profile fit to the *Chandra* profile in E2 for 11.5 kpc with  $N_{\text{H,U}} = 2.5 \times 10^{22} \text{ cm}^{-2}$  with no  $N_{\text{H,W}}$ . Bottom: Same as top, but for a distance of 13.6 kpc with no  $N_{\text{H,U}}$  or  $N_{\text{H,W}}$ .

normalization and a constant as free parameters (see Section 3.2). Note that the SNR does not always follow the brightness; the outer, larger wedges tend to have a larger SNR as the total counts are higher.

### 3 RESULTS

The main idea of this work is to generate DSH images using the 3D cloud shapes and compare them with the actual *Chandra* image. The images may be created for any source distance,  $N_{\text{H,U}}$  that represents the cumulative  $N_{\text{H}}$  distributed uniformly to all pixels, and  $N_{\text{H,W}}$  that represents the cumulative  $N_{\text{H}}$  distributed proportional to the molecular cloud brightness in each pixel (see Section 2.6). Given that there are 15 clouds, and each cloud can be in the near or far distance estimate, 32 768 images need to be created and analysed for each possible source distance,  $N_{\text{H,U}}$  and  $N_{\text{H,W}}$  combination.

To reduce the number of images created, we used a simpler analytical model that assumes spherical clouds with the amount of dust proportional to  $W(\text{CO})$  (similar to the one used in K18), permuted source distances, and near- and far-choices for each cloud. This pre-analysis indicated that there are three possible distance ranges: 11.2–11.7 kpc, 13.4–13.8 kpc, and 4.7–5.2 kpc which would result in generated images resembling the *Chandra* image. Moreover, we further reduced the number of permutations using the same model; e.g. for the 11.2–11.7 kpc range MC-80 cannot be at the near distance as a bright ring would appear at  $\sim 400''$  from the source, or for 13.4–13.6 kpc MC-80 cannot be at the far distance as we would have observed a double ring structure. We have generated images with

all remaining available cloud permutations, 0.1 kpc distance steps and  $N_{\text{H,U}}$  and  $N_{\text{H,W}}$  at 0, 1, 2, and  $2.5 \times 10^{22} \text{ cm}^{-2}$  and applied fits described in Section 2.7. For the 4.7–5.2 kpc range, we used 0.05 kpc distance steps to obtain a better fit.

#### 3.1 Radial profile results

For the 11.2–11.7 kpc radial profiles, the minimum reduced  $\chi^2$  is obtained for 11.5 kpc with uniform  $N_{\text{H,U}}$  value of  $2.5 \times 10^{22} \text{ cm}^{-2}$  with no  $N_{\text{H,W}}$  (see Fig. 8). The reduced  $\chi^2$  for the fit is 1.29 with a normalization of 0.168. The number of degrees of freedom for all radial fits is 24. The same figure also shows the best-fitting case for the 13.6 kpc distance with a reduced  $\chi^2$  of 1.00. Notice the similar radial fits stemming from very different images. Both  $N_{\text{H,U}}$  and  $N_{\text{H,W}}$  are zero for the best 13.6 kpc case, the addition of any external  $N_{\text{H}}$  makes the fit substantially worse as shown in Table 3. In this table,  $N_{\text{H,s}}$  is the average of total  $N_{\text{H}}$  for the nine pixels encompassing the source position, calculated to compare with the  $N_{\text{H}}$  values obtained from the X-ray spectral fits.

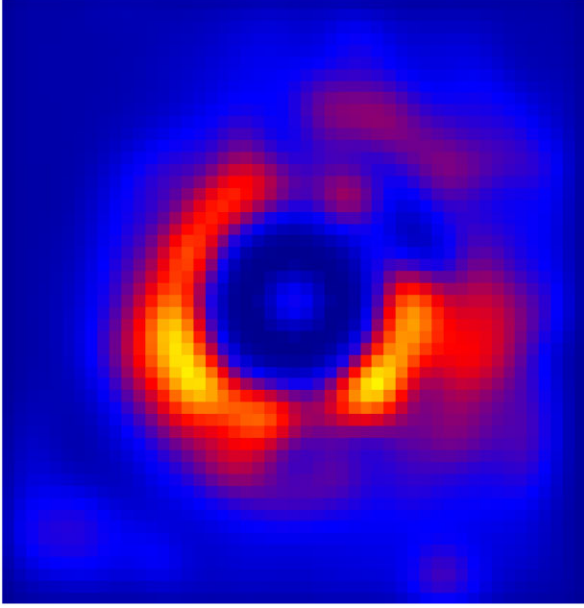
K18 near distance suggestion of 4.85 kpc (Fig. 9) performs worse in radial profile fits compared to 11.5 and 13.6 kpc distances as seen in Table 3. It also indicated a low  $N_{\text{H}}$  incompatible with the X-ray observations.

#### 3.2 Azimuthal profile results

The azimuthal profile fits take full advantage of the high resolution *APEX* images and 3D cloud shapes. For the 11.2–11.7 kpc range, the

**Table 3.** Radial fit results.

Distance (kpc)	$N_{\text{H,U}}$ $10^{22} \text{ cm}^{-2}$	$N_{\text{H,W}}$ $10^{22} \text{ cm}^{-2}$	$\chi^2$	Norm	Bkg ( $\text{ph cm}^{-2} \text{ s}^{-2} \text{ asec}^{-2}$ )	$N_{\text{H,s}}$ ( $10^{22} \text{ cm}^{-2}$ )
11.5	0	0	2.33	0.72	$6.3 \times 10^{-10}$	2.43
11.5	1	0	1.77	0.169	$4.8 \times 10^{-10}$	3.30
11.5	1	2	1.96	0.098	$5.4 \times 10^{-10}$	4.02
11.5	2.5	0	1.29	0.168	$3.3 \times 10^{-10}$	4.61
11.5	2.5	2	1.63	0.099	$4.6 \times 10^{-10}$	6.37
13.6	0	0	1.00	0.767	$4.0 \times 10^{-10}$	2.43
13.6	1	0	5.91	0.517	$-0.3 \times 10^{-11}$	3.43
4.85	2	0	2.11	1.048	$2.8 \times 10^{-10}$	2.51

**Figure 9.** The best-fitting image for 4.85 kpc distance.  $N_{\text{H,U}} = 2 \times 10^{22} \text{ cm}^{-2}$  (left) with no  $N_{\text{H,W}}$ .

best fits are obtained for 11.5 kpc with very little change in reduced  $\chi^2$  (3.06–3.13) when  $N_{\text{H,U}}$  and  $N_{\text{H,W}}$  were varied (see Table 4). The best fit distribution and image are shown in Fig. 10. The 13.6 kpc distance case yields a lower reduced  $\chi^2$  of 2.02, which is also shown in Fig. 10. Placing the MC-80 cloud in the near distance results in the best-fitting distance of 4.85 kpc with a reduced  $\chi^2$  of 3.391.

The azimuthal fits allow us to constrain the placement of clouds in the near- and far-distances. In Fig. 11, we plot the fractional distribution of whether the clouds are in the near- or far- distance for all generated images with reduced  $\chi^2 < 3.2$  for 11.5 kpc distance case. The  $\chi^2$  cut corresponds to around 5 percent of all generated images. The best fits require MC-73 and MC-80 to be at the far distance, and MC-100 and MC-105 at the near distance. The results also indicate that MC-109 and MC-114 are more likely to be at the near distance estimate, and MC-117 to be at the far distance estimate.

### 3.3 Energy dependence

While all the radial and azimuthal fits utilize the E2 band data, we performed an additional analysis using *Chandra* DSH images in the soft and hard bands (E1 and E3) in order to reveal the contributions of individual clouds to absorption and emission. By comparing the intensity distribution across these bands, we try to determine the

order of clouds along the line of sight by correlating regions of higher absorption with cloud images, since clouds that exhibit higher absorption are likely positioned in the foreground. This approach, similar to that demonstrated by S. Heinz et al. (2015), not only clarifies the individual roles of the clouds but also helps in ruling out distance estimates that are inconsistent with the observed absorption and emission patterns.

First, we subtract the normalized E3 band *Chandra* image from that of E1 to highlight the absorbed regions (see Fig. 12). Then, we integrate the 3D cloud velocity slices into one image for every cloud, using  $^{13}\text{CO}$  and  $^{12}\text{CO}$  data. Next, we compare the spatial overlap between the extracted difference image and each integrated cloud image. For this analysis, we generate a mask from the difference image by applying a threshold that isolates the regions with higher absorption. We then compute the overlap between these dark regions and the bright areas of the integrated cloud images to quantify the spatial correlation, thereby linking absorption features to specific molecular clouds.

Secondly, we executed our DSH simulations with the best-fitting parameters while preserving the full set of clouds along the line of sight and isolated the contribution of each cloud one by one to understand the individual contributions of the molecular clouds to the overall dust-scattering structure. This allowed us to identify clouds that are in front of the source and absorb X-rays but do not contribute strongly to the dust scattering image. We stress that our complete image-generation method takes these effects into account, therefore this is not an independent method, but it is much easier to implement without executing a large set of simulations.

For a demonstration, we considered the highly absorbed region on the right side of the difference image. We found that MC-100 and MC-105 are likely to be responsible as they contribute mostly to the absorption in the given regions but not to emission (see Fig. 13). Since both near and far distance estimates of MC-100 and MC-105 are larger than 4.9 kpc, this independently rules out the 4.9 kpc case discussed in K18.

## 4 DISCUSSION

Our methodology to generate DSH images using the high-resolution mm images from *APEX* includes a set of underlying assumptions, as well as limitations from the observing instruments. In this section, we will start by discussing the implied distances from our results and then continue with the impact of our assumptions on our conclusions.

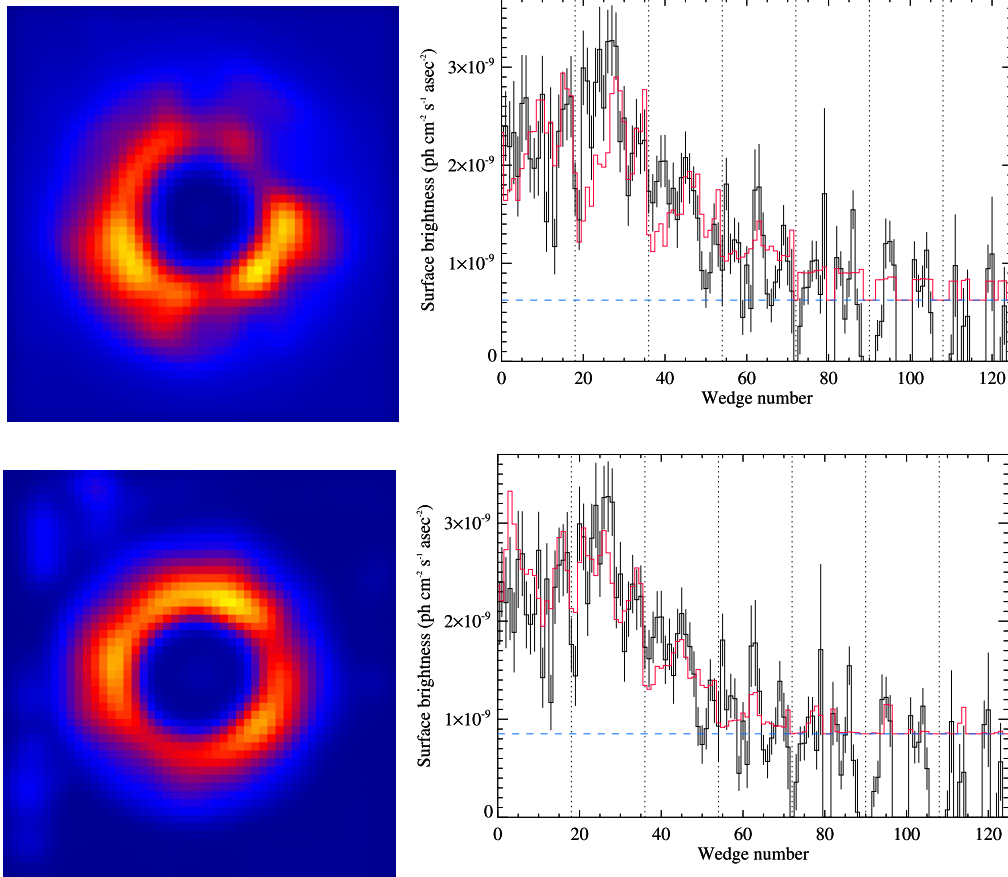
### 4.1 Is the source at 13.6 kpc?

Our method, for both the radial and azimuthal profile fits, indicates a lower  $\chi^2$  value for a source distance of 13.6 kpc than that of 11.5 kpc.



**Table 4.** Azimuthal fit results.

Distance (kpc)	$N_{H,U}$ $10^{22} \text{ cm}^{-2}$	$N_{H,W}$ $10^{22} \text{ cm}^{-2}$	$\chi^2_*$	Norm	Bkg ( $\text{ph cm}^{-2} \text{ s}^{-2} \text{ arcsec}^{-2}$ )	$N_{H,s}$ ( $10^{22} \text{ cm}^{-2}$ )
11.5	0–2.5	0–2	3.06–3.13	0.067–0.133	$6.5\text{--}9.0 \times 10^{-10}$	2.43–6.37
13.6	0	0	2.022	0.736	$8.5 \times 10^{-10}$	2.43
4.85	2	0	3.391	0.743	$8.5 \times 10^{-10}$	2.511

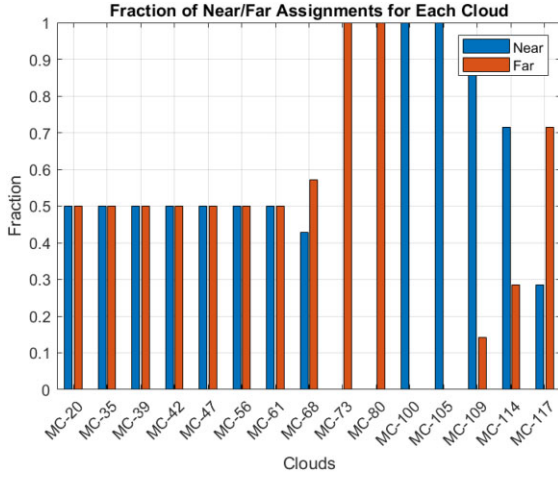
\*Reduced  $\chi^2$  with 67 deg of freedom**Figure 10.** Top: the best-fitting generated image and azimuthal profile fit to the *Chandra* profile in E2 for 11.5 kpc with  $N_{H,U} = 2.5 \times 10^{22} \text{ cm}^{-2}$ . Bottom: same as top, but for a distance of 13.6 kpc and zero  $N_{H,U}$  and  $N_{H,W}$ . See Fig. 7 to match the wedge number to a region in the image. The vertical dotted lines limit radial boundaries, and the blue horizontal dashed line shows the best-fit background.

However, there are a couple of major problems with this distance estimate. The brightest molecular cloud, MC-80 cannot be at the far distance estimate for a source distance of 13.6 kpc, because it should produce a very bright second ring at 200–400 arcsec that we do not observe (Fig. 14).

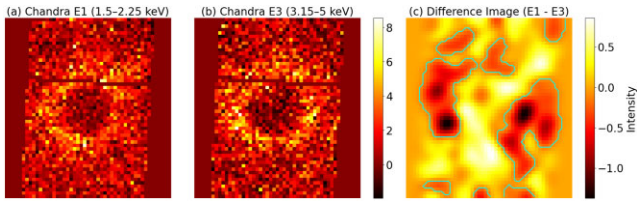
The best fits indicate MC-80 to be at the near distance; however, our analysis does not extend beyond the field of view of *APEX*. From Bronfman Survey (L. Bronfman et al. 1989), we know that MC-80 covers a much larger area than the *APEX* (see also K18 with a plot of  $-79 \text{ km s}^{-1}$  cloud sizes). Assuming a similar and uniform  $W(\text{CO})$  beyond the *APEX* coverage, we generated a larger image encompassing the *Chandra* field of view. In Fig. 15, one can see the *Chandra* E2 image and the generated image together. Bronfman survey pointings and the corresponding  $^{12}\text{CO}$  spectra are also overlaid. The generated image indicates a second ring between 600 and 750 arcsec. The average brightness in the outer ring of the

generated image,  $2.3 \times 10^{-9} \text{ ph cm}^{-2} \text{ s}^{-1}$  is almost the same as the average brightness between 100 and 200 arcsec where the observed ring is the brightest. The *Chandra* image does not show any such bright outer ring. We note that there may be regions with less dust in MC-80 coinciding with the *Chandra* pointing, and indeed, the Bronfman Survey towards the North indicates a smaller peak at  $-79 \text{ km s}^{-1}$ . However, the southern pointings are as bright as the central pointing, and yet there is no excess in the *Chandra* image.

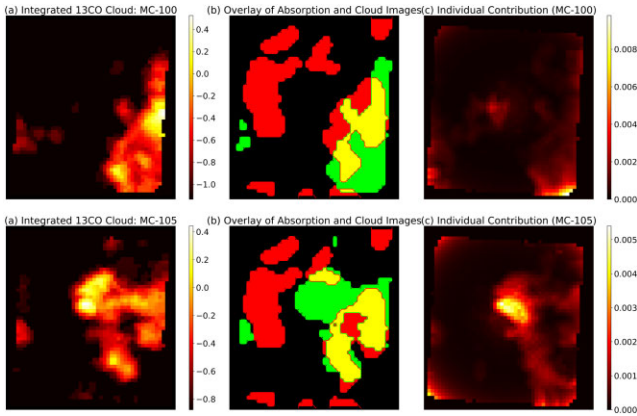
The second reason for 13.6 kpc not being the source distance is the low total  $N_H$  of  $2.4 \times 10^{22} \text{ cm}^{-2}$  that the best fit indicates (see Table 4). For the 13.6 kpc case, the main scatterers are MC-35 and MC-39 clouds, which possess much lower amounts of dust than MC-80 (see Table 2). The fits do not allow adding  $N_{H,W}$  or  $N_{H,U}$ . While we certainly are underestimating  $N_H$  in general in the generated images (see Section 4.4.2), it is not possible to make up for the missing column density for the case of 13.6 kpc source distance.



**Figure 11.** The fractions of clouds in near (N) and far (F) distances for 11.5 kpc with  $N_{H,U} = 2.0 \times 10^{22} \text{cm}^{-2}$  and no  $N_{H,u}$  under reduced  $\chi^2$  of 3.2 for azimuthal distribution fits.



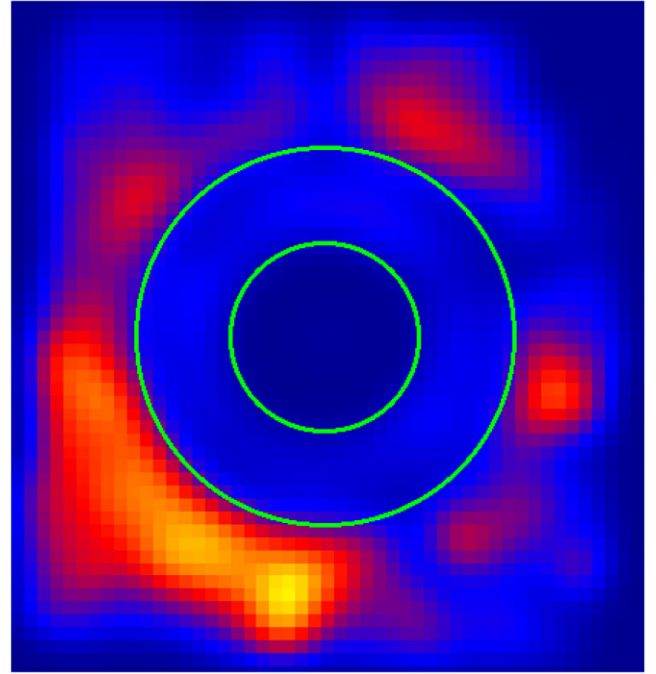
**Figure 12.** Normalized (with average brightness) *Chandra* images in E1 (a), E3 (b), and the smoothed difference (c). Highly absorbed regions are highlighted with contours.



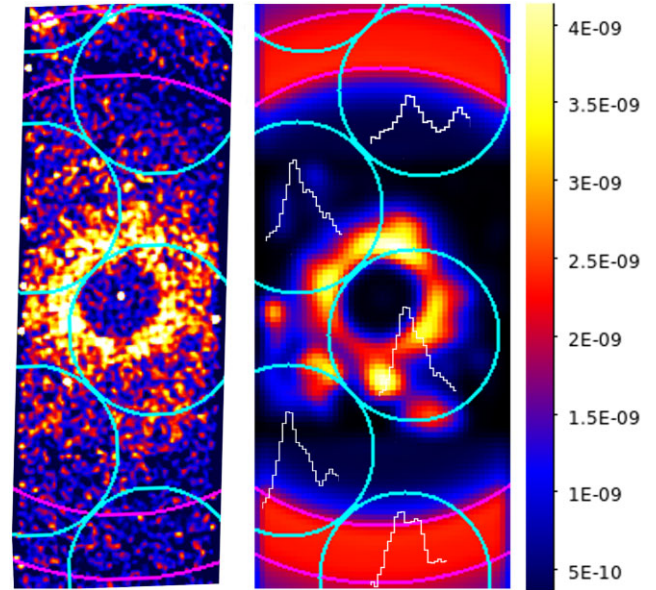
**Figure 13.** (a) The integrated  $^{13}\text{CO}$  image for the target clouds (MC100 and MC105); (b) an overlay where red indicates absorption regions, green indicates cloud regions, and yellow denotes their overlap, and (c) generated image isolating the individual contribution of the selected cloud (MC100 and MC105). The scale is with respect to the average of the full image shown in Fig. 8.

#### 4.2 Is the source at 4.85 kpc?

Both radial and azimuthal fits result in a worse  $\chi^2$  value for the 4.85 kpc case. We also showed in Section 3.3 that the mm images indicate MC-105 and MC-100 as intervening clouds, yet both their near and far distances are larger than 4.85 kpc. Similarly to the 13.6 kpc case, the implied  $N_H$  is significantly lower than values obtained from X-ray spectral fits.



**Figure 14.** Generated image for the source in the 13.6 kpc case, and MC-80 being at the far distance. The annulus encloses the main ring observed in *Chandra*.



**Figure 15.** Left: *Chandra* E2 image with Bronfman Survey pointings overlaid with circular regions. Right: generated image with source at 13.6 kpc and best radial fit parameters. The  $^{12}\text{CO}$  spectra from the Bronfman Survey are also shown inside circles. The scale on the right is only for the generated image. The magenta annulus for both images shows the region between 600 and 750 arcsec.

In K18, the uniform  $N_H$  was distributed between  $x$  values of 0.1 and 0.99, even for the near distance case. This is likely to be unphysical for the near distance case, because all of the of the uniform H I would be concentrated in less that 4.85 kpc from the solar system, and not in spiral arms. In this current work, we specifically distributed uniform H I between 2.35 and 12.9 kpc and thereby include the spiral

arms in the line of sight. We rule out the 4.85 kpc estimate more strongly than K18. It is also ruled out strongly through the analysis of X-ray polarization data by *IXPE* (H. Krawczynski et al. 2024).

### 4.3 Is the source at 11.5 kpc?

In K18, we selected MC-79 (MC-80 in this work) as the primary scattering cloud responsible for the main halo ring, since the Bronfman Survey identified it as the brightest CO emitter. This led to the conclusion that the source could be located either at 11.5 or 4.9 kpc, depending on whether MC-79 lies at its far or near distance. However, a purely spatial analysis within the limited *APEX* field of view indicates that the radial and azimuthal halo distributions are better explained by scattering primarily in MC-35 and MC-39, which would instead place the source at 13.6 kpc. The extended *Chandra* image, however, rules out this scenario, as a strong secondary ring from scattering in MC-80 should have been present but is not observed.

By systematically excluding the 13.6 and 4.85 kpc distances (Sections 4.1 and 4.2, respectively), we conclude that 11.5 kpc is the most probable distance to the source.

### 4.4 Validity of the approach and the distance error

The main hypothesis behind this study (and other similar works) is the presence of high-concentration dusty regions that scatter X-rays and produce the observed distinct rings. The most natural source of concentrated dust is molecular clouds. However, only  $\sim 20$  per cent–30 per cent of the ISM in mass is in molecular clouds (B. T. Draine 2011), and the neutral H I regions also carry large amounts of dust. We do not possess high-resolution images of the region in 21 cm, and we relied only on low-resolution surveys to estimate the atomic contribution to the total column density. As seen in Fig. 4, for most molecular clouds, there is an associated peak in the H I spectrum, certainly for the most important clouds MC-73/MC-80 and MC-35/MC-39 complex. Since dust should also exist in regions other than the molecular clouds, we added two types of contributions from  $N_{HI}$  to overall  $N_H$ , a contribution that is proportional to the molecular column ( $N_{H,W}$ ), as well as a uniform distribution added to all pixels ( $N_{H,U}$ , see Section 2.6). An azimuthal, or imaging-based analysis as conducted in this work would be less relevant if the dust concentration in atomic hydrogen clouds and dust concentration in molecular clouds were highly uncorrelated. Perhaps, the relatively poor reduced  $\chi^2$  we obtained in the azimuthal fits can be explained by the presence of additional dust in specific positions due to atomic H clouds (e.g. in the first quadrant North-East in Fig. 10, corresponding to wedges 20–22 and 38–40 shown in Fig. 7).

Another variable that one needs to take into account is the variation in the dust-to-gas ratio, both within the clouds and between the clouds (A. Giannetti et al. 2017). In our case, neither is allowed to vary. In K18, allowing some variation in the  $N_H$  of clouds resulted in better radial fits, however, the variations in K18 may also be caused by the shape of the clouds that we try to take into account in this work. Allowing such a variation will certainly reduce the  $\chi^2$  in the azimuthal fits, but independently varying all normalization with 15 clouds will possibly lead to overfitting, not to mention the prohibitively long time to generate all images to check all variations.

#### 4.4.1 Ambiguity in cloud distance estimation and the error in source distance estimation

In this work, and in K18, we used the kinematic distance estimation method which assumes that clouds move on circular velocities around

the Galactic centre. The calculation using the relation between the cloud’s heliocentric distance and the line-of-sight velocity  $v_{lsr}$  results in two solutions, with the so-called near- and far-distance estimates (M. Schmidt 1957; M. A. Kolpak et al. 2003). Additional analysis is required to resolve this kinematic distance ambiguity (KDA), such as comparing apparent radii with radii estimated through molecular line widths (P. M. Solomon et al. 1987), comparing apparent radii with radii estimated using molecular line width and surface density (M.-A. Miville-Deschênes, N. Murray & E. J. Lee 2017), using H I absorption (M. Riener et al. 2020, and references therein), reddening of foreground and background stars (E. F. Schlafly et al. 2014), association with spiral arms (M. J. Reid et al. 2016), or trigonometric parallaxes (M. J. Reid et al. 2019). In this work, rather than trying to resolve the ambiguity with additional methods, we let being in the near- or far-distance estimate as a free parameter and let the azimuthal and radial fitting algorithms pick the best estimate. For our best distance estimate, the results are encouraging, we were able to constrain the placement of 7 out of 15 clouds. Many of the clouds that we could not constrain are weak emitters (see Table 2 and Figs 11). They also are low  $v_{lsr}$  clouds, indicating that they are either behind the source (far-distance estimate), or having such a low  $x$  value that (near-distance estimate) at the time of observation, the input flux from the source is low (see equation 4). More success could have been achieved in determining the distances of clouds with a series of sensitive X-ray observations during the decay that would allow for picking contributions of low  $x$  values with brighter input flux. Finally, if we compare our cloud placements with K18, MC-73, MC-80, MC-100, and MC-105 match, and are consistent with the estimation from the  $\sigma$ – $R$  relation (P. M. Solomon et al. 1987).

Apart from the KDA, the presence of spiral density waves introduces non-circular motions in the gas, causing deviations from the simple rotation curve base estimation (F. G. Ramón-Fox & I. A. Bonnell 2017). These radial and azimuthal peculiar velocity components introduce dispersions typically less than  $10 \text{ km s}^{-1}$ , however, they can be as high as  $20 \text{ km s}^{-1}$  in specific regions within 5 kpc of the Galactic centre and within the Norma arm (M. J. Reid et al. 2019). Therefore a systematic uncertainty of 0.5 kpc exists due to the peculiar velocities of molecular clouds.

The second source of uncertainty is the Galactic rotation curve used in determining kinematic distance estimates. While we used M. J. Reid et al. (2014) A5 model to be consistent with K18, using other models such as those given in J. Brand & L. Blitz (1993) and M. J. Reid et al. (2009, 2019) resulted in variations of 0.2–0.5 kpc for the far distance estimate of the main cloud, MC-80.

Finally, there is the uncertainty of the order of 0.1 kpc arising from the radial fitting. The  $\chi^2$  changes steeply as the source distance is varied with respect to the given cloud distances. Overall, the sum of all systematic and statistical errors is 1 kpc dominated by uncertainties in determining the actual cloud distances.

#### 4.4.2 Extinction and local absorption

For sources in the Galactic plane, it is difficult to estimate the total column density. One approach relies on X-ray spectral fits, but the  $N_H$  obtained through the fits depends on the model chosen, the properties of the instrument, the abundance, and the cross-sections. As already mentioned in Section 1.1, the column density is known to vary in this source. Interestingly, *Swift* observations used here (Table 1) show significant variations despite using the same



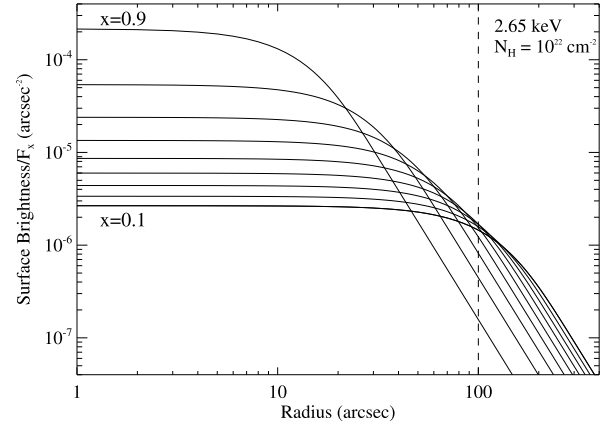
instrument, abundances, and cross sections, leaving changes local to the source as the most likely cause for the variability within a given data set. Such changes could then be attributed to absorption from the variable accretion disc outflow known to be present in 4U 1630–47 (see below), or alternatively, to changes in the source continuum during the outburst that are not well captured by the model.

One can also try to calculate the column density by adding the contributions from the molecular and neutral H clouds as we used in our formulation, however, both contributions will underestimate the  $N_H$  as the emission will be optically thick. Our study suffers from a discrepancy between the two methods. We used a *diskbb* model which is relevant for 4U 1630–47 in the soft state, and also used *wilm* abundances (J. Wilms et al. 2000) and *vern* cross-sections (D. A. Verner et al. 1996) in the fits, which resulted in typical  $N_H$  of  $\sim 13 \times 10^{22} \text{ cm}^{-2}$  (see Table 1). However, X-ray spectral fitting throughout the literature has resulted in a relatively large range of  $N_H$  values ( $\sim 5\text{--}15 \times 10^{22} \text{ cm}^{-2}$ , e.g. M. Díaz Trigo et al. 2014; J. Neilsen et al. 2014; J. A. Tomsick et al. 2014; E. Gatzuz et al. 2019). While the lowest values from the range in the literature are roughly consistent with the maximum value we obtain from molecular clouds and atomic H I ( $6 \times 10^{22} \text{ cm}^{-2}$ , see Table 3), the  $N_H$  value used in this paper is rather at the highest end of the range.

As a result, our generated images are much brighter than observed as evidenced by the low normalization factors in the case of 11.5 kpc. To put this discrepancy in another way; when we calculate unabsorbed fluxes, we use a much higher  $N_H$  compared to that used in the absorption factor in equation (3). Our aim is not accurately determining the extinction; therefore, all the discrepancies here are adjusted by the normalization factor. However, this choice may have some effect on the energy-dependent azimuthal distribution, as extinction by the intervening and post-scattering clouds is calculated separately.

For the model used in this paper, there will be less discrepancy if metal abundances are increased. In fact, if we use *angr* abundances (E. Anders & N. Grevesse 1989), while the unabsorbed fluxes remain the same, the  $N_H$  from the spectral fits reduce to  $\sim 10 \times 10^{22} \text{ cm}^{-2}$  and the generated images become dimmer because of the increased absorption cross section, such that the normalizations are now closer to 1. Another interesting possibility is that the different values observed in the fits to X-ray spectra are partially due to variable local absorption at the accretion environment of the source. 4U 1630–47 is a high inclination system showing strong, highly photoionized absorption from a disc wind with column densities above  $\sim 10 \times 10^{22} \text{ cm}^{-2}$  (e.g. A. Kubota et al. 2007; M. Díaz Trigo et al. 2014; E. Gatzuz et al. 2019). While so far the presence of a colder component of this wind could not be demonstrated due to the high extinction in the line of sight, the presence of cold winds has been established already in a few X-ray binaries and the co-existence of the hot and cold components is a subject of intense research (T. Muñoz-Darias et al. 2019; T. Muñoz-Darias & G. Ponti 2022).

Recent work by A. Kushwaha et al. (2023) shows that some *NICER* spectrum of this source could only be fitted if partial absorption (*pcfabs*) is added to the model, with almost equal amounts of  $N_H$  in *pcfabs* and *tbags*. When we tried adding *pcfabs* to the *Swift* observations discussed in Section 2.2, our fits improved with substantial  $N_H$  in *pcfabs*, but the improvement was not statistically significant. While we cannot confirm the presence of local absorption with the quality of our spectral data, it is an intriguing possibility that would reduce our discrepancy in the normalization factors and it is also supported by other observations.



**Figure 16.** The surface brightness divided by X-ray flux (*norm\_int*) obtained from *NEWDUST* fits for 10  $x$  values, from 0.1 to 0.9. The values are calculated at an average energy of 2.65 keV (E2) and for an  $N_H$  of  $1 \times 10^{22} \text{ cm}^{-2}$ . The dashed line is at  $100''$ , the radius at which the ring profile is rising quickly.

#### 4.4.3 Scattering cross-section

In K18, we used a generic power-law to represent the cross section as a function of scattering angle. However, while a power law is a good fit for the scattering cross section for large angles, for smaller angles, the cross-section flattens. This is evident in *NEWDUST* calculations of the normalized intensity (*norm\_int*) as shown in Fig. 16 as the cross-section flattens for  $x$  values less than 0.5 close to 100 arcsec, the radius at which the halo brightens significantly. We note that for the main halo, between 100 and 250 arcsec, *NEWDUST* cross sections as a function of scattering angle can be represented by a power law, and therefore K18 results are still validated. However, fits extending to lower  $\theta$  are affected more significantly.

The other advantage of using *NEWDUST* instead of a generic power-law is obtaining the exact cross-section, where in K18 an unknown normalization factor was used. We did not consider dust models other than the default in *NEWDUST*.

We stress that the calculations in *NEWDUST* are for an optically thin medium, whereas our column densities indicate that this may not be a valid approximation and there may be a significant number of double scatterings. We utilized the *REFLEX* ray-tracing code (C. Ricci & S. Paltani 2023) to check the ratio of double and single scatterings as a function of column density. We first simulated a geometrically thin (0.2 pc) sheet of gas and dust. We confirmed that the *REFLEX* generated image’s surface brightness profile matches that of *NEWDUST* for the same energy distribution and  $x$  value. We then set up the *REFLEX* parameters with a monochromatic source at 2.7 keV, a cylindrical cloud of thickness 200 pc at a distance of 650 pc away from the source which is at 11.5 kpc. The camera size and resolution mimicked the *APEX* field of view and resolution. This configuration very roughly represents MC-80 for our best distance case. We varied the column density between  $0.5\text{--}10 \times 10^{22} \text{ cm}^{-2}$  and obtained the ratio of double scatterings to all dust scatterings as 1.4 per cent, 2.8 per cent, 11.7 per cent, and 18.5 per cent at 0.5, 1., 5., and  $10 \times 10^{22} \text{ cm}^{-2}$ , respectively. Given that the actual column density is very likely to be larger than  $5 \times 10^{22} \text{ cm}^{-2}$ , our approach does not take into account double scatterings that are between 10–15 per cent of all interactions recorded. The net effect of double scatterings would be loss of flux if the double scattering removes the photon from the field of view, and the smoothing of the halo image if the photon stays, which could be another reason that our fits require a strong uniform dust distribution.



## 5 SUMMARY AND FUTURE WORK

Using high-resolution mm and X-ray band images from *APEX*, and *Chandra*, we developed a method to determine the distance to the peculiar black hole binary source 4U 1630–47. The method includes a machine learning formalism to determine 3D representations of molecular clouds and uses the X-ray flux history and dust scattering basics to generate synthetic images of the dust scattering halo for all combinations of near and far distance estimates of molecular clouds in the line of sight.

The study confirmed the previous distance estimate of 11.5 kpc for the source, and provided strong constraints on the near and far placement for 7 out of 15 molecular clouds. Other distance estimates, while providing a low  $\chi^2$  in our fitting algorithm, were ruled out when the full *Chandra* image,  $^{12}\text{CO}$  spectra from the past surveys, and the total extinction were taken into account.

While molecular cloud images and their radial and azimuthal distributions were taken into account in earlier work (e.g. S. Heinz et al. 2015), this is the first time that very high-resolution mm images are utilized to produce synthetic images for direct comparison with the observed X-ray halo. The results are promising and can be improved further in future studies. High-resolution maps and spectra of neutral hydrogen gas could highly complement the method as the results indicate as much contribution from neutral hydrogen clouds as the molecular clouds. This may also help with obtaining a better grip on the total  $N_{\text{H}}$  in the line of sight. Our method keeps the overall normalization free because we cannot accurately determine the total hydrogen density, whereas this parameter comes into our equations both in determining dust scattering intensity and absorption. A future study could tackle finding the right combination of H density and absorption cross sections to generate consistent images in more than one energy band.

This analysis may be extended to other sources with existing halo images. The X-ray imaging resolution does not have to be super-high, as the beam width of *APEX* is 30 arcsec. However, the timing of observations and the sensitivity are important, when the source is bright, the X-ray telescope PSF and the halo image may be hard to disentangle.

With future high-resolution and high-statistics X-ray images (e.g. using *Lynx* or *AXIS*; J. A. Gaskin et al. 2019; C. S. Reynolds et al. 2023), and known source distance, the problem can be reversed, and one can determine the distances to the molecular clouds with good accuracy to test Galactic rotation curve models, as well as how much the molecular clouds stick to the distances given by their radial velocities.

## ACKNOWLEDGEMENTS

This publication is based on data acquired with the Atacama Pathfinder Experiment (APEX) under programme IDs 0107.F-9323 and 0110.F-9305. APEX is a collaboration between the Max-Planck-Institut für Radioastronomie, the European Southern Observatory, and the Onsala Space Observatory. Swedish observations on APEX are supported through Swedish Research Council grant no. 2017–00648. This publication has made use of MAXI data provided by RIKEN, JAXA, and the MAXI team. The authors thank Sebastian Heinz, Lia Corrales, and Stephane Paltani for useful discussions. The authors thank the anonymous reviewer for their valuable suggestions and corrections that have improved the clarity and quality of this publication.

## DATA AVAILABILITY

The *Chandra* and *Swift* data underlying this article are available in HEASARC <https://heasarc.gsfc.nasa.gov/docs/archive.html>, and the *MAXI* data are available at [maxi.riken.jp](http://maxi.riken.jp). The *APEX* data are available at ESO archives <https://archive.eso.org/wdb/wdb/eso/apex/form>. The generated images in this article will be shared on reasonable request to the corresponding author.

## REFERENCES

- Abdulghani Y., Lohfink A. M., Chauhan J., 2024, *MNRAS*, 530, 424
- Abe Y., Fukazawa Y., Kubota A., Kasama D., Makishima K., 2005, *PASJ*, 57, 629
- Anders E., Grevesse N., 1989, *Geochimica et Cosmochimica Acta*, 53, 197
- Arnason R. M., Papei H., Barmby P., Bahramian A., Gorski M. D., 2021, *MNRAS*, 502, 5455
- Atri P. et al., 2020, *MNRAS*, 493, L81
- Audley M. D., Nagase F., Mitsuda K., Angelini L., Kelley R. L., 2006, *MNRAS*, 367, 1147
- Augusteijn T., Kuulkers E., van Kerkwijk M. H., 2001, *A&A*, 375, 447
- Bachetti M. et al., 2014, *Nature*, 514, 202
- Barret D., McClintock J. E., Grindlay J. E., 1996, *ApJ*, 473, 963
- Berry D. S., 2015, *Astronomy and Computing*, 10, 22
- Berry D. S., Reinhold K., Jenness T., Economou F., 2007, in Richard A. Shaw, Frank Hill, David J. Bell., eds., ASP Conf. Ser. Vol. 376 Astronomical Data Analysis Software and Systems XVI. Astron. Soc. Pac., San Francisco, p. 425
- Bolatto A. D., Wolfire M., Leroy A. K., 2013, *ARA&A*, 51, 207
- Brand J., Blitz L., 1993, *A&A*, 275, 67
- Brightman M. et al., 2019, *BAAS*, 51, 352
- Bronfman L., Alvarez H., Cohen R. S., Thaddeus P., 1989, *ApJS*, 71, 481
- Burridge B. J. et al., 2025, *ApJ*, preprint ([arXiv:2502.06448](https://arxiv.org/abs/2502.06448))
- Chauhan J. et al., 2021, *MNRAS*, 501, L60
- Cheng Y., 1995, *IEEE Trans. Pattern Anal. Mach. Intell.*, 17, 790
- Comaniciu D., Meer P., 1999, in Proceedings of the seventh IEEE international conference on computer vision. IEEE, p. 1197
- Corral-Santana J. M., Casares J., Muñoz-Darias T., Bauer F. E., Martínez-Pais I. G., Russell D. M., 2016, *A&A*, 587, A61
- Corrales L. R., García J., Wilms J., Baganoff F., 2016, *MNRAS*, 458, 1345
- Corrales L. R., Paerels F., 2015, *MNRAS*, 453, 1121
- Corrales L., 2023, *ebur/newdust: Python extinction and scattering halo calculations for astrophysical particulates*. Zenodo
- Costantini E., Corrales L., 2022, in Bambi C., Sanganello A., eds, Handbook of X-ray and Gamma-ray Astrophysics. Springer, Berlin, p. 40
- Díaz Trigo M., Migliari S., Miller-Jones J. C. A., Guainazzi M., 2014, *A&A*, 571, A76
- Díaz Trigo M., Miller-Jones J. C. A., Migliari S., Broderick J. W., Tzioumis T., 2013, *Nature*, 504, 260
- Draine B. T., 2011, Physics of the Interstellar and Intergalactic Medium. Princeton Univ. Press, Princeton
- Fukunaga K., Hostettler L., 1975, *IEEE Trans. Inf. Theory*, 21, 32
- Gaskin J. A. et al., 2019, *J. Astron. Telesc. Instrum. Syst.*, 5, 021001
- Gatuzz E., Díaz Trigo M., Miller-Jones J. C. A., Migliari S., 2019, *MNRAS*, 482, 2597
- Giannetti A. et al., 2017, *A&A*, 606, L12
- Güsten R., Nyman L. Å., Schilke P., Menten K., Cesarsky C., Booth R., 2006, *A&A*, 454, L13
- Heinz S. et al., 2015, *ApJ*, 806, 265
- Heinz S., Corrales L., Smith R., Brandt W. N., Jonker P. G., Plotkin R. M., Neilsen J., 2016, *ApJ*, 825, 15
- HI4PI Collaboration et al., 2016, *A&A*, 594, A116
- Jon P. et al., 2024, *MNRAS*, 530, 929
- Jones C., Forman W., Tananbaum H., Turner M. J. L., 1976, *ApJ*, 210, L9
- Kalemci E., Kara E., Tomsick J. A., 2022, in Bambi C., Sanganello A., eds, Handbook of X-ray and Gamma-ray Astrophysics. Springer, Berlin, p. 9
- Kalemci E., Maccarone T. J., Tomsick J. A., 2018, *ApJ*, 859, 88

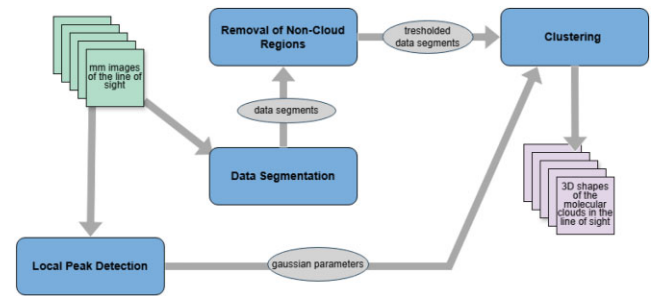
- Kolpak M. A., Jackson J. M., Bania T. M., Clemens D. P., Dickey J. M., 2003, *ApJ*, 582, 756
- Krawczynski H. et al., 2024, *ApJ*, 977, L10
- Kubota A. et al., 2007, *PASJ*, 59, 185
- Kushwaha A., Jayasurya K. M., Agrawal V. K., Nandi A., 2023, *MNRAS*, 524, L15
- Maccarone T. J., Coppi P. S., 2003, *MNRAS*, 338, 189
- Markwardt C. B., 2009, in Bohlender D. A., Durand D., Dowler P., eds, ASP Conf. Ser. Vol. 411, *Astronomical Data Analysis Software and Systems XVIII*. Astron. Soc. Pac., San Francisco, p. 251
- Masias M., Freixenet J., Lladó X., Peracaula M., 2012, *MNRAS*, 422, 1674
- Mathis J. S., Lee C.-W., 1991, *ApJ*, 376, 490
- McClure-Griffiths N. M., Dickey J. M., Gaensler B. M., Green A. J., Haverkorn M., Strasser S., 2005, *ApJS*, 158, 178
- Miller-Jones J. C. A., Jonker P. G., Dhawan V., Briskin W., Rupen M. P., Nelemans G., Gallo E., 2009, *ApJ*, 706, L230
- Miville-Deschênes M.-A., Murray N., Lee E. J., 2017, *ApJ*, 834, 57
- Mohan S. S., Leela S., 2013, *IOSR J. Comput. Eng. (IOSR-JCE)*, 14, 80
- Muno M. P., Morgan E. H., Remillard R. A., 1999, *ApJ*, 527, 321
- Muñoz-Darias T. et al., 2019, *ApJ*, 879, L4
- Muñoz-Darias T., Ponti G., 2022, *A&A*, 664, A104
- Neilsen J., Coriat M., Fender R., Lee J. C., Ponti G., Tzioumis A. K., Edwards P. G., Broderick J. W., 2014, *ApJ*, 784, L5
- Otsu N., et al., 1975, *Automatica*, 11, 23
- Overbeck J. W., 1965, *ApJ*, 141, 864
- Predehl P., Burwitz V., Paerels F., Trümper J., 2000, *A&A*, 357, L25
- Predehl P., Schmitt J. H. M. M., 1995, *A&A*, 293, 889
- Ramón-Fox F. G., Bonnell I. A., 2017, *MNRAS*, 474, 2028
- Reid M. J. et al., 2009, *ApJ*, 700, 137
- Reid M. J. et al., 2014, *ApJ*, 783, 130
- Reid M. J. et al., 2019, *ApJ*, 885, 131
- Reid M. J., Dame T. M., Menten K. M., Brunthaler A., 2016, *ApJ*, 823, 77
- Revnivtsev M., Gilfanov M., Churazov E., Sunyaev R., 2002, *A&A*, 391, 1013
- Reynolds C. S. et al., 2023, in Siegmund O. H., Hoadley K., eds, Proc. SPIE Conf. Ser. Vol. 12678, *UV, X-Ray, and Gamma-Ray Space Instrumentation for Astronomy XXIII*. SPIE, Bellingham, p. 126781E
- Ricci C., Paltani S., 2023, *ApJ*, 945, 55
- Riener M., Kainulainen J., Henshaw J. D., Beuther H., 2020, *A&A*, 640, A72
- Rosolowsky E., Pineda J., Kauffmann J., Goodman A., 2008, *ApJ*, 679, 1338
- Rousseuw P. J., 1987, *J. Comput. Appl. Math.*, 20, 53
- Schlaflly E. F. et al., 2014, *ApJ*, 786, 29
- Schmidt M., 1957, *Bull. Astron. Inst. Neth.*, 13, 247
- Solomon P. M., Rivolo A. R., Barrett J., Yahil A., 1987, *ApJ*, 319, 730
- Stutzki J., Guesten R., 1990, *ApJ*, 356, 513
- Thompson T. W. J., Rothschild R. E., 2009, *ApJ*, 691, 1744
- Tomsick J. A., Corbel S., Goldwurm A., Kaaret P., 2005, *ApJ*, 630, 413
- Tomsick J. A., Yamaoka K., Corbel S., Kalemci E., Migliari S., Kaaret P., 2014, *ApJ*, 791, 70
- Trümper J., Schönfelder V., 1973, *A&A*, 25, 445
- Uttley P., Klein-Wolt M., 2015, *MNRAS*, 451, 475
- Vahdat Motlagh A., Kalemci E., Maccarone T. J., 2019, *MNRAS*, 485, 2744
- Verner D. A., Ferland G. J., Korista K. T., Yakovlev D. G., 1996, *ApJ*, 465, 487
- Wang Y., Méndez M., 2016, *MNRAS*, 456, 1579
- Williams J. P., De Geus E. J., Blitz L., 1994, *ApJ*, 428, 693
- Wilms J., Allen A., McCray R., 2000, *ApJ*, 542, 914
- Xiang J., Lee J. C., Nowak M. A., Wilms J., 2011, *ApJ*, 738, 78
- Xu D., Zhu Y., 2024, *Astron. Comput.*, 48, 100838
- Zhou H., Wang X., Schaefer G., 2011, *Innovations in intelligent image analysis*, Vol 309, Springer, Berlin, Heidelberg, p. 291.

## APPENDIX A: 3D CLOUD REPRESENTATION ALGORITHM

In this section, we explain the algorithm used to determine the 3D shapes of interstellar molecular clouds along the line of sight to

the source object. The goal is to reconstruct the 3D morphology of molecular clouds using high-resolution maps from *APEX*. These data provide detailed insight into the cloud distribution along the line of sight. The main challenges are the overlap of clouds in velocity space and the narrow field of view, which limits the capture of full cloud structures. Standard clump identification algorithms such as ClumpFind (J. P. Williams, E. J. De Geus & L. Blitz 1994), GaussClumps (J. Stutzki & R. Guesten 1990), FellWalker (D. S. Berry 2015), Reinhold (D. S. Berry et al. 2007), and Dendrograms (E. Rosolowsky et al. 2008) were tested but were unsuitable for our data, as they either overestimated or failed to identify meaningful clumps due to these constraints. Hence, an alternative approach was adopted.

The approach consists of four core stages: local peak detection through Gaussian decomposition, segmentation of the data, threshold to identify cloud regions, and clustering data points into clouds. (see Fig. A1)



**Figure A1.** Illustration of the data processing pipeline for molecular cloud detection, involving local peak detection, segmentation, removal of non-cloud regions, and clustering.

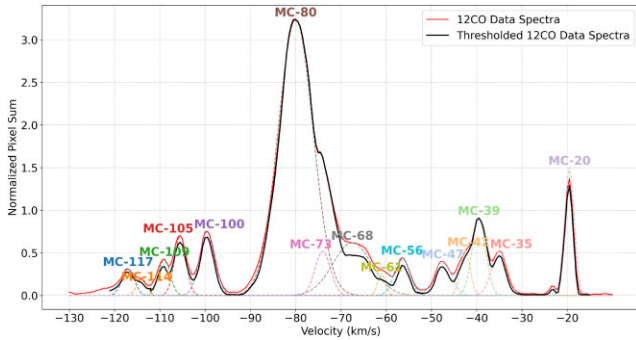
### A1 Initial Gaussian fits and cloud identification

In the first step, spectra along the line of sight are analysed by integrating pixel values across all velocity channels. Integrated spectra are fitted with Gaussian functions, under the assumption that each Gaussian component corresponds to one molecular cloud. A reduced chi-square minimization is employed to determine the minimum number of Gaussians required to fit the data ( $\chi^2 \approx 1$ ) as 15, which is recorded as the number of clouds (see Fig. A2).

For each fitted Gaussian, the key parameters—amplitude, central velocity ( $\mu$ ), and velocity dispersion ( $\sigma$ )—are extracted. The velocity interval for each cloud is defined as the range from  $\mu - 3\sigma$  to  $\mu + 3\sigma$ , capturing approximately 99.73 per cent of the cloud’s intensity. These boundaries are then rounded to the nearest 0.5 or whole number velocity unit to maintain consistency with the data’s resolution. This approach results in some overlap between adjacent clouds, which is addressed in later steps. The peak, or centre, of each cloud is identified by selecting the velocity value at the Gaussian peak ( $\mu$ ) as the cloud’s central velocity. In the spatial dimensions ( $x$ ,  $y$ ), the cloud’s centre is determined by recording the coordinates of the pixel with the maximum intensity at this velocity value.

### A2 Data segmentation

In the second step of the algorithm, the data are divided into five distinct segments using the velocity intervals identified in the previous step. Clouds with overlapping velocity intervals are grouped. For most cloud groups, no overlapping velocities were



**Figure A2.** Velocity spectrum created by summing pixel intensities across velocity channels ( $-130$  to  $-10 \text{ km s}^{-1}$ ). The red solid line shows the normalized original data, the blue solid line beneath it represents Otsu-thresholded data, and the dashed lines are Gaussian fits representing 15 molecular clouds along the line of sight.

present. However, in one case where a minor overlap was observed, the overlapping region was divided evenly between the two groups. This segmentation not only reduces the data size, enhancing the efficiency of both the thresholding and clustering processes but also allows for more precise application of thresholds in each segment. By minimizing overlap and grouping similar velocity regions, the segmentation ensures that subsequent clustering steps are performed more effectively.

### A3 Thresholding

In this step, Otsu’s thresholding method is applied to each data segment to separate cloud-occupied and non-cloud regions based on pixel intensity. Otsu’s algorithm automatically determines an optimal threshold by minimizing within-class variance (N. Otsu & others 1975). This method is widely used in astronomical image segmentation, including radio astronomy for thresholding (M. Masias et al. 2012) (D. Xu & Y. Zhu 2024). Each segment is thresholded separately, allowing for region-specific thresholds that improve accuracy. The non-cloud regions, which fall below the computed threshold, are discarded, and the cloud regions are passed to the clustering step (see Fig. A2).

### A4 Clustering

In the final step, clustering is performed on the thresholded data segments, where each cluster corresponds to a molecular cloud. A modified version of the Mean Shift algorithm is applied after normalizing the data. The number of clusters and their centres, determined in earlier steps, are used as inputs for the clustering. Hyperparameters are fine-tuned using error functions based on the cloud properties. After obtaining the optimal clustering parameters, the results are post-processed to generate the 3D shapes of the clouds.

#### A4.1 Data pre-processing

Each pixel in the thresholded segments is treated as a data point with attributes (x, y, velocity, intensity), and all features are normalized to the  $[0, 1]$  range using Min-Max scaling to prepare for the Mean Shift clustering.

#### A4.2 Clustering algorithm

Each data segment is clustered with Mean-Shift algorithm into a predefined number of clouds, as determined during Gaussian decomposition. The Mean-Shift algorithm, which normally starts with random cluster centres, is applied here using the predetermined peaks as initial cluster centres. The algorithm iteratively shifts these centres toward regions of higher data density, converging at local maxima, where clusters form around the densest regions (Y. Cheng 1995) (K. Fukunaga & L. Hostetler 1975). Mean-Shift has been widely used in image segmentation tasks and also in remote sensing (D. Comaniciu & P. Meer 1999) (H. Zhou, X. Wang & G. Schaefer 2011) (S. S. Mohan & S. Leela 2013). While Mean-Shift typically performs hard clustering, assigning each data point to a single cluster, we adapted it to soft clustering using a Gaussian kernel to handle overlapping clouds. This allows data points to belong to multiple clusters with associated probabilities. To refine these probabilities, the alpha and beta parameters are introduced. The alpha parameter sets a lower threshold, redistributing small probabilities into significant cluster memberships, while the beta parameter caps higher probabilities, treating values above the threshold as fully assigned to a cluster. These modifications help manage overlapping regions and prevent the formation of insignificant clusters.

#### A4.3 Hyperparameter tuning and error functions

Several hyperparameters require tuning in the clustering algorithm, including bandwidths for each cluster, feature scales, and the alpha and beta parameters. Hyperparameter optimization is performed using a combination of the Silhouette index (P. J. Rousseeuw 1987) and two custom error functions. The tuning process begins with gradient descent, starting from default initial values, and refines the parameters based on the Silhouette index. Final fine-tuning is conducted using the two error functions, which are explained below. The final parameters are manually selected from the best results for each data segment.

### A5 Error functions

Two error functions were applied in the hyperparameter tuning process to evaluate the spatial properties of the cloud shapes. These functions assess shape consistency and smoothness to optimize clustering results.

#### A5.1 Shape smoothness error

This error function evaluates the smoothness of cloud shapes by comparing consecutive velocity segments. The difference between each segment and its neighboring segments is computed, and the sum of these differences is normalized by the number of velocity segments per cloud. The total error is then averaged across all clouds. The formulation of the error function is given below:

$$\text{error} = \frac{1}{N} \sum_c \left[ \frac{1}{S_c} \sum_v (2 \cdot I_v - I_{v+0.5} - I_{v-0.5}) \right] \quad (\text{A1})$$

Where  $N$  is the total number of clouds,  $S_c$  is the size of cloud  $c$ , and  $I_v$  is the cloud segment at velocity  $v$ .

#### A5.2 Eccentricity consistency error

This error function measures the consistency of cloud shapes by comparing the eccentricity of each velocity segment to the central

segment. The eccentricity is determined using the eigen-ratio of the elliptic shapes. The squared difference between the central and other segments is summed, normalized by the number of segments, and averaged across all clouds. The formulation of the error function is given below:

$$\text{error} = \frac{1}{N} \sum_c \left[ \frac{1}{S_c} \sum_v (R_c - R_v)^2 \right] \quad (\text{A2})$$

Where  $N$  is the total number of clouds,  $S_c$  is the size of cloud  $c$ ,  $R_c$  is the centre eigen-ratio of cloud  $c$ , and  $R_v$  is the eigen-ratio at velocity  $v$ .

#### A6 Data post-processing

For each pixel with a soft cluster assignment, the pixel is divided into separate data points for each cluster. Each pixel is split into new data

points, with their intensities scaled by the probability of belonging to

each cluster. Once all pixels are processed, the final cloud structures are formed (see Fig. 5). Finally, a Gaussian smoothing function is applied along the velocity axis of each pixel of each cloud. This redistributes the pixel brightnesses around the centroid slice of each cloud while maintaining the total summed intensity for each  $(x_p, y_p)$ , ensuring the cloud remains within a specific thickness (in pc) whilst maintaining the overall intensity. The thicknesses of clouds are determined from the near and far distance size estimates given in M.-A. Miville-Deschênes et al. (2017), similar to what is done in K18.

This paper has been typeset from a  $\text{\LaTeX}$  file prepared by the author.