

The highly variable time evolution of star-forming cores identified with dendrograms

Rachel A. Smullen¹ ,¹★ Kaitlin M. Kratter,¹ Stella S. R. Offner,² Aaron T. Lee³ and Hope How-Huan Chen²

¹Steward Observatory, University of Arizona, Tucson, AZ 85721, USA

²Department of Astronomy, The University of Texas at Austin, Austin, TX 78712, USA

³Department of Physics and Astronomy, Saint Mary's College of California, Moraga, CA 94575, USA

Accepted 2020 July 23. Received 2020 June 26; in original form 2019 November 18

ABSTRACT

We investigate the time evolution of dense cores identified in molecular cloud simulations using dendrograms, which are a common tool to identify hierarchical structure in simulations and observations of star formation. We develop an algorithm to link dendrogram structures through time using the three-dimensional density field from magnetohydrodynamical simulations, thus creating histories for all dense cores in the domain. We find that the population-wide *distributions* of core properties are relatively invariant in time, and quantities like the core mass function match with observations. Despite this consistency, an *individual* core may undergo large (>40 per cent), stochastic variations due to the redefinition of the dendrogram structure between time-steps. This variation occurs independent of environment and stellar content. We identify a population of short-lived (<200 kyr) overdensities masquerading as dense cores that may comprise ~ 20 per cent of any time snapshot. Finally, we note the importance of considering the full history of cores when interpreting the origin of the initial mass function; we find that, especially for systems containing multiple stars, the core mass defined by a dendrogram leaf in a snapshot is typically less than the final system stellar mass. This work reinforces that there is no time-stable density contour that defines a star-forming core. The dendrogram itself can induce significant structure variation between time-steps due to small changes in the density field. Thus, one must use caution when comparing dendrograms of regions with different ages or environment properties because differences in dendrogram structure may not come solely from the physical evolution of dense cores.

Key words: stars: formation – ISM: clouds.

1 INTRODUCTION

Understanding the early progression and end state of star formation is fundamental to many areas of astronomy, from modelling the formation of galaxies to studying the assembly of planetary systems. Stars form in dense molecular cores embedded within gravitationally contracting filamentary structures (André et al. 2010; Arzoumanian et al. 2013; Smith, Glover & Klessen 2014; Arzoumanian et al. 2019). On core scales, gravity sets the dominant dense core properties (Goodman et al. 2009; Lee et al. 2014b; Storm et al. 2016), while turbulence is thought to regulate the star formation efficiency and core formation, including properties like core rotation (Padoan, Haugbølle & Nordlund 2012; Chen & Ostriker 2018). There is also a population of observed, pressure-confined cores that will likely not form stars if left untouched, although these objects may later collapse due to shock interactions (Seo et al. 2015; Keown et al. 2017; Kirk et al. 2017; Chen et al. 2019). Despite an understanding of this broad process of star formation, there are still many open questions. These include the relationship between observed core masses and the initial mass function (IMF), the time evolution of dense core properties, the role of the physical environment in the star formation process, and

the formation mechanisms of bound binary (or higher order multiple) systems, among others.

Previous works have attempted to answer some of these questions by looking at individual snapshots of observed regions or simulations, yet few have ever attempted to correlate the evolution of individual cores with the broad core property distributions reported in the literature. The interplay between the time evolution of individual cores and their contribution to distributions of core properties may be especially important when understanding the connection between the core mass function (CMF) and stellar IMF (Offner et al. 2014, and references therein). There is still debate about whether the IMF directly inherits its shape from the CMF (e.g. Padoan & Nordlund 2002; Hennebelle & Chabrier 2008; Hopkins 2013) or is independent of core masses (e.g. Bonnell et al. 2001; Bate, Bonnell & Bromm 2003; Clark, Klessen & Bonnell 2007). The IMF is frequently fit with the form of a power law at high masses and a lognormal distribution at lower masses as first demonstrated in Chabrier (2003). Subsequent work has suggested that the IMF is mostly independent of star formation physics such as accretion rate and star formation inefficiency (Hennebelle 2012; Cunningham et al. 2018), but may depend on local environmental properties like the global radiation field and local magnetic fields (Bate 2009; Offner et al. 2009; Dib, Schmeja & Hony 2017; Guszejnov, Hopkins & Krumholz 2017; Lee, Hennebelle & Chabrier 2017; Cunningham et al. 2018; Ntormousi & Hennebelle

* E-mail: rsmullen@email.arizona.edu

2019). Thus, it is imperative to know how individual cores may contribute to the interplay between the CMF and IMF evolution.

A fundamental aspect to properly interpreting both simulation snapshots and observations of star-forming regions is understanding *what* overdensities are identified as cores. Core identification in both observed star-forming regions and simulations has been a topic of active investigation for decades. Beginning with the by-eye identification of structure in molecular clouds from Blitz & Stark (1986), the field has expanded in two dominant directions. The first direction is the singular identification of dense clumps, which started from the watershed segmentation algorithm of Williams, de Geus & Blitz (1994). This developed into the CLUMPFIND algorithm that has been utilized extensively. Other core-finding methods that return singular clumps include the gradient-tracing scheme FellWalker (Berry 2015), GaussClumps (Stutzki & Guesten 1990), which fits Gaussians to all peaks in the data, and CUTEX (Molinari et al. 2011), which looks for curvature changes in the data, among others. The other method of core identification is using hierarchical structure methods. Early hierarchical structure methods such as the structure trees from Houlahan & Scalo (1992) then evolved into the commonly adopted dendrogram algorithm first presented in Rosolowsky et al. (2008). Dendrograms connect structures in star-forming regions from filaments to dense cores and allow a better understanding of the hierarchical nature of the star formation process (Goodman et al. 2009).

Each core identification algorithm comes with its own often subtle biases that must be understood in the context of the analysis performed (e.g. Li et al. 2020). For instance, GaussClumps and CUTEX only fit elliptical sources, but GaussClumps can easily handle overlapping sources (Stutzki & Guesten 1990) and CUTEX works well with large background variations (Molinari et al. 2011). CLUMPFIND has been found to be sensitive to input parameters but is widely available (e.g. Berry 2015). FellWalker clumps can sometimes have artificial splitting due to the cleaning process but tends to be more robust to noise (Berry 2015). Dendrograms can be sensitive to the algorithm tuning choices but provide the best understanding of the physical environment surrounding cores (Rosolowsky et al. 2008). The above is not a comprehensive list of the benefits and drawbacks of core identification methods, but it serves to show that every algorithm in use will work better in some situations as compared to others.

Simulations have become a critical tool to interpret the necessarily incomplete window provided by observations in star-forming regions, especially as simulations have grown in resolution and complexity. For instance, Mairs et al. (2014) note the importance of high-resolution observations in recovering the full mass and detailed structure of star-forming cores. Observations at moderate resolution tend to miss mass and structure due to averaging errors (Offner et al. 2012; Mairs et al. 2014). Similarly, Beaumont et al. (2013) reach the important conclusion that position–position–velocity observations carry uncertainties of 40 per cent in computed quantities when compared to a three-dimensional simulation. Effects like gas superposition along the line of sight, line opacity obscuring core structure, and mapping observational (PPV) space to physical (PPP) space contribute confusion to an accurate physical interpretation of cores from observations, because the line-of-sight structure of a core can be easily miscalculated (e.g. Ostriker, Stone & Gammie 2001; Ballesteros-Paredes & Low 2002; Shetty et al. 2010; Beaumont et al. 2013).

In this paper, we explore the time evolution of star-forming cores identified with dendrograms and work to understand the role of the dendrogram algorithm itself in the properties of identified cores. We begin by creating an algorithm to link dendrogram structures through time, which we describe in Section 2. The robustness of this methodology depends on several tunable parameters, and we explore

the effect of the three major parameters in Section 3. We present the time evolution of the identified cores, the distributions of core properties, and other results of note in Section 4, and we then explore the reasons for the variability we find in Section 5. Finally, Section 6 notes the implications of our findings, including the importance of full core histories and the limitations of the dendrogram algorithm. Section 7 summarizes our findings.

2 METHODOLOGY

In this work, we aim to trace the histories of cores in simulations of star formation to test the robustness of core parameters measured throughout a core’s lifetime. Here, we discuss the magnetohydrodynamic simulation used in this work, describe our method of core identification, and present our new core tracking method. We describe the fiducial choices for our core identification and linking methodology in this section and test the impact of variations in these choices in Section 3.

2.1 A note on nomenclature

The meaning of the word ‘core’ is not well defined between works. Observationally based definitions of cores, collected in Chen et al. (2019) and references therein, include ‘dense cores’, which are regions that have a dominant thermal velocity and low virial α (virialized), ‘starless cores’ that do not have a protostar and are not virialized, ‘prestellar cores’ that do not have a protostar but are virialized, and ‘protostellar cores’, which do have protostars and are virialized. Cores can also be defined in simulations as the material that accretes on to a star particle (e.g. Bate et al. 2003), or the region of dense material at a single snapshot (e.g. Ntormousi & Hennebelle 2019).

This paper explores differences between algorithmic, physical, and phenomenological understandings of dense cores in star formation. Towards that goal, we use the term *leaf* for a dense structure inside contours identified with a dendrogram, the term *overdensity* for a physical collection of dense gas in the simulations, and the term *core* for the loosely defined, observationally motivated dense structures that may form stars.

2.2 Simulations

Our simulation initial conditions are identical to those of run W2T2 in Offner & Arce (2015). These conditions are intended to model a piece of a local, Gould Belt star-forming region like the Perseus molecular cloud. For our purposes, the simulation represents a prototypical turbulent molecular cloud that serves as a test-bed for our core identification and tracking method; the properties of the cloud itself have little bearing on our methodology as we are investigating trends in the evolution of core structure. We outline the initialization and parameters of the simulation below.

We run a magnetohydrodynamical (MHD) simulation of a $\sim 3800 M_{\odot}$ (7.5×10^{36} g) gas cloud using the ORION2 code (Klein 1999; Krumholz et al. 2007; Li et al. 2012). ORION2 is a three-dimensional adaptive mesh refinement (AMR) MHD grid code that includes physics such as self-gravity, ideal MHD (Li et al. 2012), and Lagrangian accreting sink particles (Krumholz, McKee & Klein 2004; Lee et al. 2014a). Our simulations are initialized on a 256^3 base grid that corresponds to 5 pc on a side with periodic boundary conditions in all spatial dimensions. We expect little influence on core evolution from our choice of boundary condition as compared to a global molecular cloud simulation.

These simulations refine the spatial resolution based on the Jeans number J such that

$$J \equiv \frac{\Delta x_i}{\lambda_J} < 0.125, \quad (1)$$

where Δx_i is the cell size at the current level i and $\lambda_J = (\pi c_s^2 / G \rho)^{1/2}$ is the Jeans length (Truelove et al. 1997). When $J > 0.125$, finer cells with size Δx_i are added, thus resolving the Jeans wavelength with higher resolution. Our simulations have five refinement levels over the base grid, which defines our minimum resolution as ~ 4034 au per cell for the 256^3 grid at level 0 and the maximum resolution as ~ 126 au per cell for the cells refined to level 5. These sizes are defined based on the mean gas density in the simulation of $\rho_0 = 5 \times 10^{-21} \text{ g cm}^{-3}$ (1300 cm^{-3}) and the mean sound speed of 18800 cm s^{-1} . In regions undergoing gravitational collapse, gas is removed from the grid and replaced with a sink particle if $J > 0.25$ on the finest level (Krumholz et al. 2004). Sinks accrete mass and momentum from gas within a radius of four cells at level 5 as well as interact gravitationally with the surrounding gas.

We generate the cloud initial conditions through a turbulent driving phase that proceeds without gravity, which produces self-consistent turbulent gas density and velocity distributions (e.g. Mac Low 1999; Li et al. 2004; Offner, Klein & McKee 2008). The simulation begins with a uniform density, uniform temperature of 10 K, and a uniform magnetic field in the z -direction, $B_z = 13.5 \mu\text{G}$, which corresponds to an initial thermal pressure to magnetic pressure ratio (plasma β) of $\beta = 8\pi\rho_0 c_s^2 / B_z^2 = 0.1$. Then the gas is stirred for two gas crossing times by perturbing the gas velocities with a random velocity distribution that corresponds to a flat distribution in Fourier space with wave numbers $k = 1 - 2$. At the end of the driving phase, the gas reaches a turbulent steady state with a turbulent power spectrum $P(k) \propto k^{-2}$ and $\beta = 0.02$ (Offner & Arce 2015; Offner & Liu 2018). Finally, self-gravity is turned on, and we evolve the simulation for approximately 70 per cent of a global free-fall time ($t_{\text{ff}} = \sqrt{3\pi/32G\rho_0} \simeq 1.5 \text{ Myr}$).

We adopt a barotropic equation of state of the form $p = \rho c_{\text{iso}}^2 [1.0 + (\rho/\rho_c)^{\gamma-1}]$, where c_{iso} is the sound speed for 10 K gas, ρ_c is the critical density at which the gas transitions from isothermal to adiabatic, and $\gamma = 5/3$ is the adiabatic index. We choose an effective critical density that is comparable to the Jeans density on the maximum AMR level, $\rho_c = 7 \times 10^{-15} \text{ g cm}^{-3}$. This value is smaller than the expected critical density for dense gas, $\rho_c \simeq 10^{-14} \text{ g cm}^{-3}$ (Masunaga, Miyama & Inutsuka 1998), in order to produce some warming when the maximum gas densities are reached. This lower critical density acts to eliminate contiguous small-scale fragmentation, which would otherwise occur within isothermal filaments at high resolution (Kratter, Murray-Clay & Youdin 2010) and roughly approximates the influence of radiative feedback, which is expected to heat the gas once protostars form (Offner et al. 2009).

The time evolution of the dense gas is shown in Fig. 1. The density thresholds used in this plot correspond to the densities of the AMR refinement thresholds (computed from equation 1). The simulations end with about $70 M_\odot$ of mass in 25 sink particles. The fraction of mass in the densest gas shows more variability because of mass accretion on to the sink particles. Most of the dense cores are formed in one large filament that spans the majority of the volume.

2.3 Structure identification

Dendrograms are a common tool used to identify dense structures in both simulations and observations of star-forming regions; many

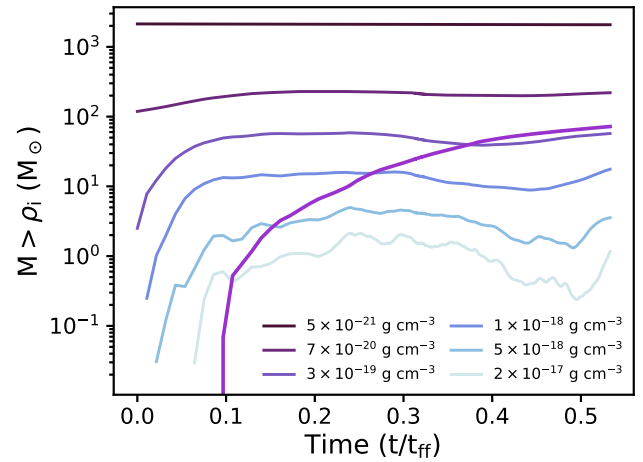


Figure 1. Time evolution of dense gas in our simulation. The blue coloured lines indicate the total mass above a given density across time, while the purple line shows the mass of sink particles across time. The lowest density shown, $5 \times 10^{-21} \text{ g cm}^{-3}$, corresponds to a number density of approximately $1300 \text{ nH}_2 \text{ cm}^{-3}$, while the highest density of $2 \times 10^{-17} \text{ g cm}^{-3}$ corresponds to a molecular hydrogen density of $5 \times 10^6 \text{ cm}^{-3}$.

previous works have used them to find and identify properties of bound clumps and filaments (e.g. Rosolowsky et al. 2008; Goodman et al. 2009; Beaumont et al. 2013; Burkhart et al. 2013; Lee et al. 2014b; Seo et al. 2015; Friesen et al. 2016; Storm et al. 2016; Keown et al. 2017; Wong et al. 2017; Chen et al. 2018; Nayak et al. 2018). Dendrograms also provide a metric to quantify the structure of molecular cloud emission and associated physical properties (e.g. Boyden et al. 2016; Koch et al. 2017; Boyden et al. 2018; Koch et al. 2019). For example, Goodman et al. (2009) demonstrated that dendrograms produce more physically reasonable identifications of cores in 3D spectral line data compared to another previously popular algorithm, CLUMPFIND. As an additional benefit, dendrograms naturally identify nested features and therefore reflect the relationship between structures of different sizes in the data. Thus, we choose dendrograms as our structure identification algorithm to better quantify the interpretation of this widely adopted method. However, because of the fundamental similarities between all core identification methodologies – that cores are identified from peaks in quantities such as density or emission – the results of this work should be generally applicable across algorithms.

A dendrogram is a tree algorithm that identifies hierarchical structures in any input quantity in a two- or three-dimensional grid. A dendrogram contains leaves (the most refined structure), trunks (the lowest level structure that may contain refined substructure), and branches (structures that connect leaves to other branches or trunks). The dendrogram initialization is commonly defined by three parameters: the background level cut-off that defines the base of the tree, the minimum difference (height) between two nested structures in the quantity being dendrogrammed required to create a new branch or leaf, and the minimum size of an identified structure. The dendrogram is built by first identifying the maximum value in the grid. The algorithm then iteratively searches adjoining cells and uses the size and density increase criteria to determine if a new branch or leaf needs to be created. The tree ends when the background cut-off value is reached. Neighbouring leaves can be children of a single branch if they are both contained in the spatial bounds of the branch. A region can contain multiple unconnected trees if an area surrounding two structures is below the background cut-off. Note

that dendrograms are inherently relative structures because they are computed based on the maximum value in a region. This work utilizes the ASTRODENDRO Python package.¹

We carefully consider how we optimize the three parameters (background cut-off, minimum density increase for new structure creation, and minimum structure size) that define how a dendrogram is built.² The background cut-off will set the fraction of gas in the simulation included in the dendrogram and impact the total height of the tree (peak to minimum density, which will likely also impact the number of structures in a tree) and the number of branches that can be created. The minimum density increase to create new structure sets a height at which new branches and leaves are created: a smaller value allows smaller density increases to be considered as new structure, while a larger value makes the creation of new structure much more stringent. Finally, the minimum structure size influences the size and internal complexity of an identified core. Too large of a size means that we might group individual compact structures into one leaf, while too small of a size might overresolve substructure in the star-forming cores we wish to study (e.g. lumps in a disc-like overdensity).

The background density threshold is a major limitation to the complexity of the dendrogram. A low background density threshold connects more of the cloud structure, including filaments, but these structures would not be readily observable in traditional tracers. On the other hand, a high-density threshold might prematurely truncate low-density structure. We use a fiducial density threshold of $\rho = 7 \times 10^{-20} \text{ g cm}^{-3}$ ($1.8 \times 10^4 \text{ cm}^{-3}$), which is the density needed to refine a cell from level 0 to level 1. The threshold density chosen herein roughly corresponds to the minimum density observed in ammonia emission ($n \gtrsim 10^4 \text{ cm}^{-3}$ or $\rho = 4 \times 10^{-20} \text{ g cm}^{-3}$), so the structures identified in our dendrograms would be observed in synthetic observations (Flower, des Forêts & Walmsley 2006). This creates a dendrogram that contains only a few per cent of the data by volume and consistently contains the same dense structures throughout the entire simulation time. As seen in Fig. 1, our fiducial density encompasses a roughly constant mass (around $100 M_{\odot}$) over the length of the simulation. Variations of the background density cut-off are described in Section 3.1.

We next consider the density increase to create a new leaf. This parameter impacts the inclusion of low-density structures in our dendrograms. A small density increment produces many nested structures, and these new structures (typically intermediate branches) do not add to the understanding of either leaf structure or the hierarchy. A large density increment leads to very large leaves and begins to underresolve the dense structures that best resemble dense cores by combining multiple overdensities into one leaf. We therefore chose a factor of 3 increase in density as the fiducial density contrast required to create a new leaf. This choice is further discussed in Section 3.2.

Finally, we set the fiducial minimum size of a structure to be 125 voxels (three-dimensional cells; at our fiducial grid size, one voxel is $(1000 \text{ au})^3$). The minimum size of 125 voxels is large enough to encompass compact structures, such as protostellar discs, without being small enough to allow clumpy substructures to split into multiple leaves. At our fiducial resolution, this leads to a *minimum* leaf size of about $(5000 \text{ au})^3$, although most structures are substantially larger. For comparison, the dendrograms of Friesen et al. (2016), who investigate the size and mass of embedded clumps

in the Serpens South protocluster, have a smallest effective radius of $\sim 0.02 \text{ pc}$ or $\sim 4100 \text{ au}$. Typical observed core sizes from works such as Seo et al. (2015) and Keown et al. (2017) are 10^{-2} – 10^{-1} pc , so our minimum size resolves structures in our simulation that are similar to observed cores. The choice of the other two dendrogram initialization parameters can, in some instances, negate the utility of the minimum size. If the background density threshold is high and the density increase is large, small structures will be not be able to be resolved and every structure will exceed the minimum structure size. With our fiducial parameters, structure can approach the minimum size but the majority of leaves have volumes of a few hundred to a few thousand voxels (core sizes $\gtrsim 0.05 \text{ pc}$).

While dendrograms can be computed for any scalar quantity, we choose to compute dendrograms on the three-dimensional density grid. The large dynamic range of physical density in our simulations means that a logarithmic scaling better traces the physical structures. Therefore, to define structures in our simulations, we compute a dendrogram with periodic (wrapping) boundary conditions on the log of the density field at each simulation snapshot. The dendrogram routine can only search a uniform grid, so we must apply a covering grid to our AMR simulations. Covering grids interpolate the AMR data on to a fixed grid of size 256×2^i in each dimension where 256 is the base size of the simulation and i is the level for which we want to create the grid. We define our dendrogram on a level 2 covering grid with each cell having a side length of $1.5 \times 10^{16} \text{ cm}$, or about $0.005 \text{ pc} \approx 1000 \text{ au}$. Our choice of fiducial resolution is discussed in Section 3.3.

Our choice of parameters leads to dendrograms containing about 80 leaves in all but the earliest time-steps when gravitating structure has barely started to collapse. An example of the dendrogram computed with these fiducial parameters at an intermediate time-step (40 per cent t_{ff}) is shown in Fig. 2. Many of the leaves (>50 per cent) are isolated and not part of a larger structure that contains further refinement due to our choice of background density cut-off; however, these leaves tend to be of relatively low density, and most will likely not form stars as they are temporary structures (see further discussion in Section 4.4). Most leaves (>80 per cent) do not form sink particles by the end of the simulation, and leaves containing one sink particle are about 2–3 times more common than leaves hosting multiple sink particles. Sink-hosting leaves can decrease in peak density after sink formation due to accretion of high-density gas on to the sink particle.

With the dendrogram defined, we output a catalogue of important leaf parameters at each time-step (density, position, velocity, magnetic fields, etc.) using the full AMR grid that falls within the volume of the uniform-grid leaf surface. This catalogue is then used to perform the linking algorithm defined in the next section.

2.4 Linking structures through time

Once the 3D structures are constructed for every time-step, we link them through time. We take a two-step approach by first linking structures between consecutive time-steps and then by reconstructing a structure’s full path through time.

2.4.1 Pair-wise linking

To match structures between time-steps, we use a geometric search that relies upon simulation outputs being frequent enough that structures do not move significantly (more than about half of their size) between outputs. Beginning with an individual leaf (l_a) at time-step t_a , we search for all leaves at time-step t_b where the centre of

¹<http://www.dendrograms.org/>

²In the ASTRODENDRO package, these variables are named `min_value`, `min_delta`, and `min_npix`.

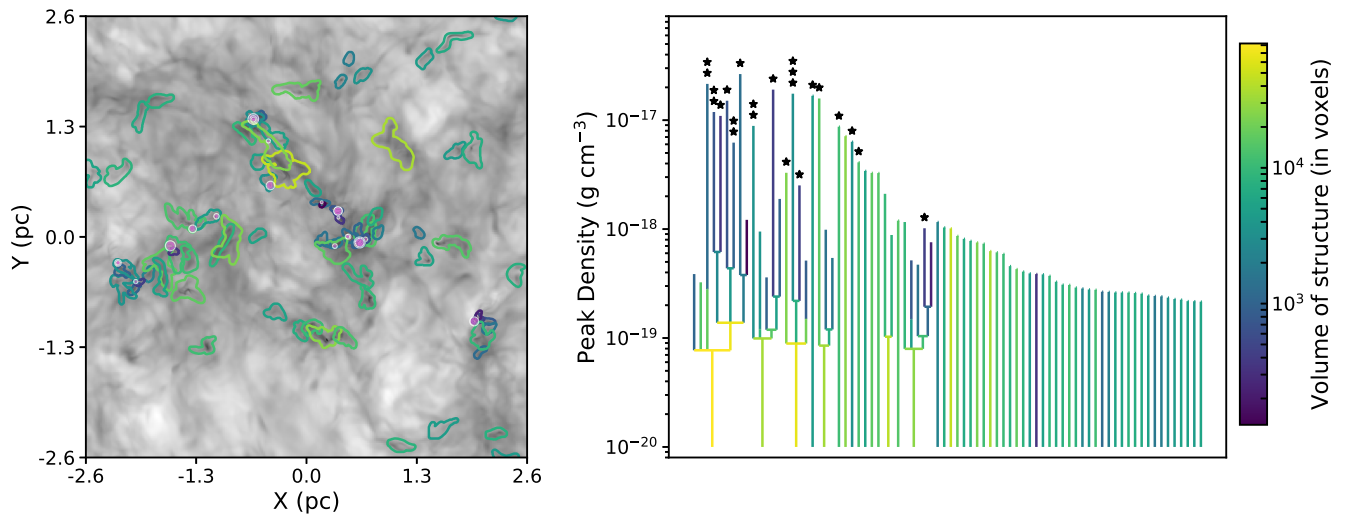


Figure 2. An example of the dendrogram computed with fiducial parameters at an intermediate time-step. The right-hand panel shows the tree coloured by the volume of the leaf. Black stars denote the presence of sink particles in the leaf. The x -axis has no physical meaning; the structures have been roughly sorted by peak density. The left-hand panel shows the $x - y$ projection of the leaves. Pink circles denote the location of sink particles with sizes scaled by the mass of the sink.

mass of l_a is within the surface of a leaf at t_b . We then reverse the search such that we look for the centre of mass of a time t_b leaf to be within the surface of a time t_a leaf. We do allow for an offset of the centre of mass from the boundary of the leaf in two dimensions because of the possibility of dendrogram contours being defined differently between consecutive time-steps as discussed below. The choice of this offset is described in Section 3.4, but our fiducial value is set to 10 grid cells at level 2. A leaf at one time-step can be associated with multiple leaves at the next time-step, and we describe the consequences of this further below. This search is then continued between all consecutive pairs of time-steps (i.e. $t_a \leftrightarrow t_b$, $t_b \leftrightarrow t_c$, etc.).

There are four cases that result from the pair-wise linking as shown in Fig. 3. Leaves can be uniquely identified with a single structure between $t_a \rightarrow t_b$ and $t_b \rightarrow t_a$. This is most common and leads to a single path between time-steps (panels ‘standard’ and ‘offset’). However, multiple leaves can be found at one time-step that map back to a single leaf at the adjacent time-step. If the single leaf is at an earlier time-step and the multiple leaves are at a later time-step, this is a ‘split’. If the single leaf is at a later time-step and the multiple leaves are at an earlier time-step, this is a ‘merger’. In our simulations, splits and mergers are most frequently due to dendrogram leaf boundaries being drawn to encompass multiple nearby overdensities, not actual physical merging or fragmentation. Physical evolution can happen but is difficult to disentangle from the changes in dendrogram contours.

In the 170 output time-steps of our simulation, we link 11 000 leaf pairs: 10 500 of those are securely linked, meaning that we identify the same linked pair looking forward and backward. We find ~ 200 splits and mergers. About 60–70 percent of linked pairs have no offset, and the average offset of the remaining linked pairs is about 10 cells (~ 0.05 pc).

We initially incorporated the velocity information, specifically the leaf centre-of-mass velocity (which includes the contribution from any sink particles that may be present in the leaf), into our algorithm to help inform the direction of motion to uniquely track a leaf through time and reduce the number of nearby, unassociated leaves linked. However, because of variability in the computation of dendrogram structures between time-steps, the leaf centre of mass

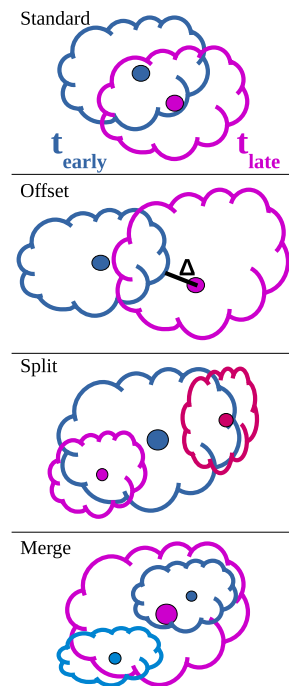


Figure 3. A cartoon of the cases that result from the pair-wise linking. Blue colours indicate earlier times and the purple colours indicate later times. The top is the ‘standard’ linking where the centre of mass at one time-step (filled circle) is found within the volume at the other time-step (open contour). The ‘offset’ linking allows there to be a small offset (Δ) between the leaf centre of mass and the leaf volume at consecutive time-steps, which typically arises from dendrogram contours being redrawn to include more material. ‘Splits’ and ‘mergers’ are cases in which a leaf at one time-step can be associated with more than one leaf at a consecutive time. Note that, while this cartoon is shown in 2D, the linking in our data is done in 3D.

does not always move in predictable patterns. Therefore, the addition of velocity information does not improve our linking. To demonstrate this issue, we present two examples of leaf behaviour in Fig. 4. The error bars in the upper panels have been doubled in length to be

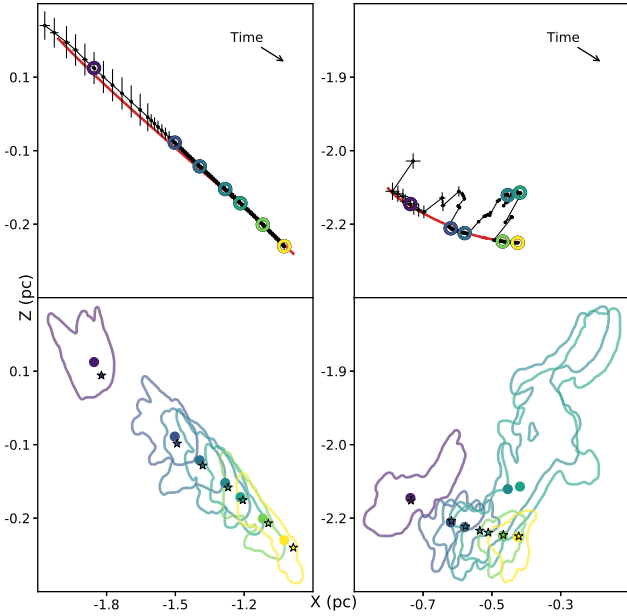


Figure 4. Two examples of leaf behaviour illustrated by sink-hosting leaves. The left column shows an ideal behaviour where the leaf structure evolves smoothly, while the right column shows a leaf that undergoes significant dendrogram structure variations in time. The upper panels show the $x - z$ positions of a leaf through time. The black line shows the centre of mass of the leaf through time, while the red line shows the position of the sink particle in the leaf. The error bars on the black line show the ‘expected’ position of the leaf at the next time-step given the velocity of the centre of mass of the leaf. The bottom panels show a selection of the projected leaf contours at the times indicated by coloured points in the top panel. The solid circles denote the centre of mass of the leaf and the star indicates the position of the sink particle. The lower right panel demonstrates a common barrier to velocity-based linking: the dendrogram leaf contours can change significantly between time-steps, offsetting the centre of mass of the leaf from the expected position.

more visible. The left-hand panels shows what a well-behaved leaf looks like: the leaf centre of mass at the next time is within the position expected from the velocity. However, a significant number ($\gtrsim 25$ per cent) of our leaves have a history that look more like the right-hand panels, where at certain times, the dendrogram contours are redrawn to include more material. This then changes the centre of mass of the leaf and the expected position of the leaf centre is wildly offset from the computed location of the centre of mass. The leaf centres at consecutive time-steps are typically within the leaf contours, meaning that the less complicated geometric search discussed above is more reliable for our data. We do encounter a few pathological cases where a reconstruction cannot be performed in an automated way, such as when the leaf is shaped like a banana – the centre of mass lies outside the leaf contour and is therefore computed to have a large offset to the leaf boundary at the next time-step.

2.4.2 Path reconstruction

The last step to fully trace the histories of overdensities in our simulations is to transform the pair-wise linking into a coherent path through time. We use the terminology ‘path’ to denote a single set of related leaves through time and ‘path family’ to denote a group of paths that were found to be related to a single starting leaf. For the analysis presented herein, we work backward in time (from the end of the simulation to the start) because the most relevant structures

to compare to observations are the compact overdensities found in well-evolved regions at later times in the integration. Because we use a fixed starting point in time, cores may be traced at different evolutionary stages.

We start by selecting a leaf (l_0) from the cohort of leaves at the final time-step (t_0). We search through the linked pairs $t_0 \leftrightarrow t_1$ to find the leaf at t_1 linked to l_0 . This found leaf is added to the path. We then check if the leaf at t_1 is associated with any other leaves at t_0 . We then iteratively repeat this process to search for matches to the earliest leaf in the path going backward in time.

For the cases where there are mergers (two or more leaves at an earlier time being associated with only one leaf at a later time), we add one of the leaves to the current path and then add new paths to the path family by copying the current path and appending the other merged leaf. Each path in the path family is then reconstructed independently.

For the cases where there are splits (two or more leaves at a later time being associated with only one leaf at an earlier time), we create a new path and recursively search in the opposite direction (from early times to late times) to find the path(s) associated with the new leaf.

Path families can have many component paths because each new split or merger effectively doubles the number of paths in a path family. While not always indicative of physical interactions, a large-number path family does indicate that the structure lives in a crowded area of the simulation volume.

3 PARAMETER VARIATIONS

All core identification algorithms include tunable parameters, and the dendrogram algorithm we adopt here is no exception. In observational studies, the parameters are chosen based on the noise, sensitivity, and resolution of the data. When analysing simulation data there is more flexibility in parameter choice. Consequently, we explore a variety of parameter values to assess the physical impact of our parameter choices, including background cut-off density, density increment to create a new structure, grid resolution, and linking distance Δ . Thus, in this section, we explore how variations in these parameters impact our reconstructed path families. While the minimum leaf volume at constant resolution is also a tunable parameter, we find negligible impact on the final dendrogram structure when varying this quantity within reasonable limits.

3.1 Background cut-off

The background cut-off influences the tree complexity, leaf structure, and computational requirements of a dendrogram. Values near the mean density in the simulation ($5 \times 10^{-21} \text{ g cm}^{-3}$) include too much gas that never participates in the star formation process. Very large, low-density structures affiliated with the filamentary structure are commonly identified as leaves. Values at high density (level 2 refinement density or higher, or around $3 \times 10^{-19} \text{ g cm}^{-3} \sim 8 \times 10^4 \text{ cm}^{-3}$) exclude an extremely large portion of the gas ($\gtrsim 99$ per cent), including gas at early times that will eventually fall in to a dense core. The level 2 refinement density corresponds to a Jeans length of 0.08 pc, which is smaller than typically observed cores. The dendrogram would be less likely to resolve any structure larger than this, which includes most of the objects that resemble observed cores; instead, the algorithm would only identify small peaks in larger overdensities. For these reasons, and the physical arguments described in the previous section, we use the fiducial density of $7 \times 10^{-20} \text{ g cm}^{-3}$.

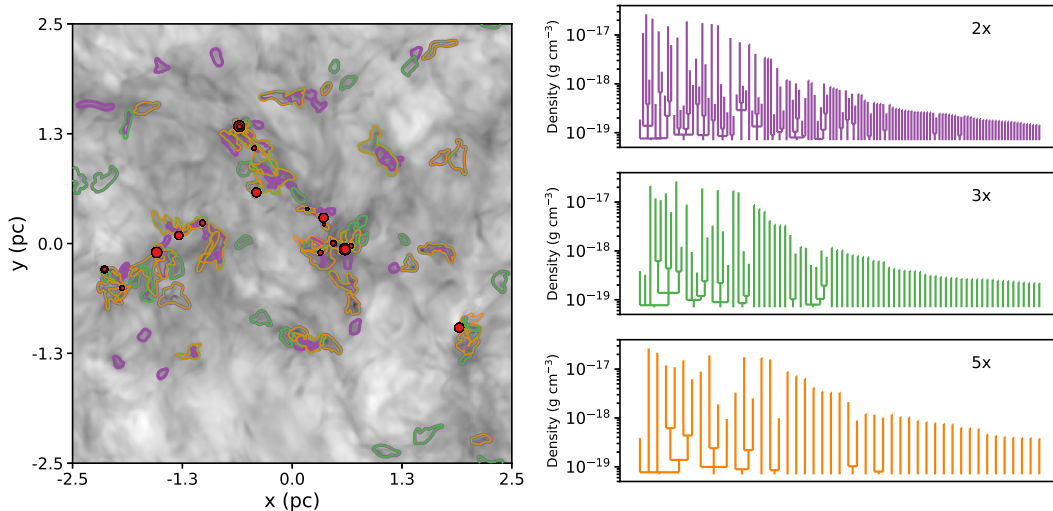


Figure 5. A comparison of dendrogram leaves given different density increments. On the left, the purple, green, and orange contours demarcate leaves from dendrograms with a $2\times$, $3\times$, and $5\times$ density increase (contrast) required to create a new leaf, respectively. Contours are shown over a density projection of the simulation with darker colours indicating denser regions. Red circles show the locations of sink particles and are scaled by sink mass. The right-hand panels show the full dendrograms for each density increment. The different trees trace out the same dense material, indicating that our results are relatively invariant of the choice of contrast.

3.2 Density increment

The contrast required to create a new structure in the dendrogram mainly impacts the low-density structure identified in the tree. We compare density increases of factors of 2, 3 (our fiducial choice), 4, 5, and 10 and find little difference between the leaves identified, although the trees themselves are quite different. We show the comparison of factors of 2, 3, and 5 in Fig. 5. All high-density structure is contained in all trees; the major differences arise in the low-density structures. Every sink particle lives in a nearly identical leaf, meaning that the important structures for star formation are not impacted greatly by our choices of density contrast parameter.

The trees computed with large density contrasts identify much less structure because overdensities must be much more significant to be added to the tree. This means that, in dense regions especially, two neighbouring overdensities may be enclosed in one leaf. Small density contrasts can lead to very low density leaves being added to the tree. These leaves are insignificant temporary perturbations above the background cut-off and add a level of unnecessary ‘noise’ to the linking process.

Burkhart et al. (2013) perform a similar analysis by varying the density increase required to create a new structure (δ in that work) and comparing the resulting dendrograms across a suite of MHD simulations. They find that dendrogram structure varies significantly with δ and can provide information about the relative importance of shocks, self-gravity, and super-Alfvénic turbulence.

3.3 Resolution

The size and shape of structures in the dendrogram are highly correlated with the resolution of the uniform grid used to compute the dendrogram. When all other parameters are kept fixed, an increase in resolution, unsurprisingly, allows for both more refined structures and physically smaller structures. The algorithm identifies more structures because each increase in the level of the uniform grid provides a factor of 8 increase in the number of cells, meaning that there is more flexibility to define compact structure. Physically smaller structures are identified because the minimum size of a structure

is fixed at 125 cells; therefore, each increase in level decreases the minimum required physical volume of a structure by a factor of 8.

We compute our fiducial dendrogram at level 2. 1024^3 cells in the volume gives a 1000 au per cell resolution. We also tested level 3 (2048^3 cells; 500 au per cell) and level 4 (4096^3 cells; 250 au per cell) resolution. The memory required for producing a dendrogram at the full level 4 volume was prohibitive. We therefore use a subset of the volume of size [2048, 2048, 1536] at level 4 (about 10 per cent of the volume) that contains 15 of our 24 sink particles and many of the structures identified at level 2. Other than resolution, we compute the dendrograms for each case using the same fiducial parameters as described above in Section 2.3. The comparison volume contains 40 level 2 leaves, 52 level 3 leaves, and 63 level 4 leaves.

The contours of leaves at the three levels are shown in Fig. 6. As seen in the left-hand panel, the contours at all levels broadly agree. Only in the densest regions do the leaf volumes differ significantly. The right-hand panel shows one of these dense regions: a triple system in a complicated overdensity illustrates how differences in the resolution can change the leaf structure. The level 2 dendrogram encloses all triple members in one leaf. The level 3 dendrogram draws the central binary in one contour but excludes the tertiary component along the z -direction. The tertiary’s local overdensity is not large enough to create an independent leaf. The level 4 dendrogram assigns all the small overdensities in the greater disc-like overdensity to their own leaves.

Fig. 6 suggests that level 4 is too sensitive to substructure: the overdensity of one physically bound system is often split into substructures such that we lose information about the bound core. It is not possible to compute important system quantities such as gravitational potential or virial parameter without having the full bound structure contained in a single leaf.

Next, we assess the utility of the level 3 dendrogram. Ideally, we want to minimize large changes in the density contours while including all important structure (or sink particles) in a given overdensity. To this end, we compare the derived leaf parameters between leaves tracing the same physical structure at level 2 and level 3 in Fig. 7. The correlation in mass, peak density, and velocity dispersion is very good (typically within a factor of 3) despite the

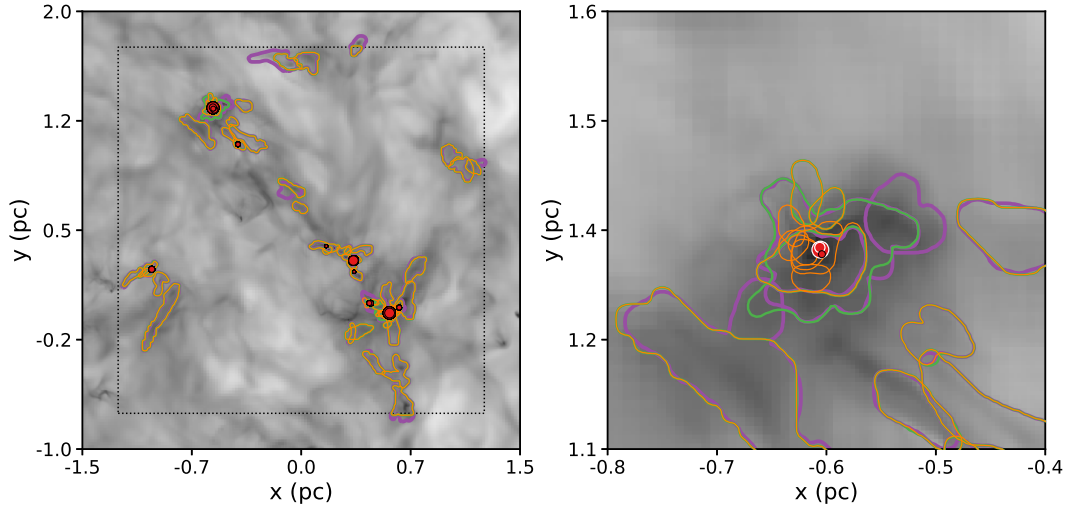


Figure 6. Contours of leaves from trees computed on uniform grids at AMR levels 2 (the fiducial choice in this work), 3, and 4 in purple, green, and orange, respectively. The contours are shown over a density projection of the simulation with darker colours indicating denser regions. Red circles show the locations of sink particles and are scaled in size relative to their mass. The black dashed line in the left-hand panel indicates the extent of the comparison volume. As is seen in the left-hand panel, the contours at all levels broadly agree. The right-hand panel reveals that only in the densest regions do the leaf volumes differ significantly. This region contains an overdensity that surrounds a bound triple system. The sink particles outlined in white comprise the central binary (5 and $2 M_{\odot}$), while the small circle outlined in black is the tertiary companion ($0.8 M_{\odot}$). The tertiary is separated from the central binary by a few thousand au.

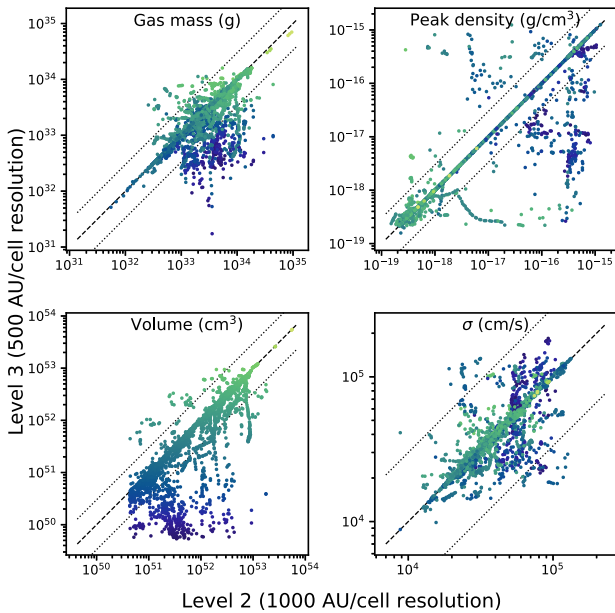


Figure 7. A comparison of derived leaf parameters between correlated leaves at level 2 and level 3. The panels show gas mass, peak density, volume, and velocity dispersion. Points are coloured by the volume at level 3. The horizontal axis shows the values of leaves computed on level 2 and the vertical axis shows the value for the corresponding leaf computed at level 3. The black dashed line shows the one-to-one correlation, while the dotted lines show a factor of 3 difference. The correlation in mass, peak density, and velocity dispersion is very good despite the fact that there is a factor of 8 difference in the minimum leaf volume.

fact that there is a factor of 8 difference in the minimum volume. Outliers below the one-to-one correlation in all panels except velocity dispersion arise from the population of small, low-density leaves that are identified as independent structures at level 3. These small overdensities are typically included as part of a larger level 2 leaf

at the lower resolution, but the low density of the structures means that the mass-weighted parameters such as velocity dispersion are mostly agnostic about their inclusion. Leaves above the one-to-one line in the same panels come from ambiguities in the correlation of leaves across resolution, but they constitute a small fraction of the total number of leaves shown. Patterns of points forming lines in any of the panels are indicative of time evolution. Thus, we conclude that the difference in leaves identified at level 2 and level 3 does not impact our understanding of dense core evolution.

3.4 Linking distance

While the aforementioned parameters control the construction of the dendrogram at each time snapshot, the linking distance is the crucial parameter that controls the history of structures (Δ in the ‘offset’ panel in Fig. 3). The linking distance is the distance between the centre of mass of a leaf at one time-step and the surface of the leaf to which it has been linked. Linking distance will simultaneously impact the number of paths in a family and the number of time-steps traced in an individual path. Typical leaf sizes are of order 0.2 pc, so we test linking distances of 0 cells (no offset; a leaf centre is within the contour at the neighbouring time), 10 cells ($\sim 10^4$ au; about a typical leaf radius), 100 cells ($\sim 10^5$ au; about 10 leaf radii), and 200 cells ($\sim 2 \times 10^5$ au). Our goal is to robustly identify leaves with common histories without permitting too many uncertain connections while, at the same time, allowing for variations in dendrogram leaf contours.

We present the results of our investigation in Fig. 8. There are minimal variations between the 100 cell and 200 cell linking distances, so we only present the 100 cell results in the figure. For most path families, specifically those of isolated leaves, linking distance does not make a difference in the number of paths reconstructed. The smaller two linking distances, on average, create smaller numbers of paths in the family. Some of this is due to large variations in dendrogram contours between time-steps; the changes in the leaf boundaries can be larger than the linking distance. Linking distance does have a much stronger impact on the length of a path history, however. Larger linking distances typically lead to longer paths,

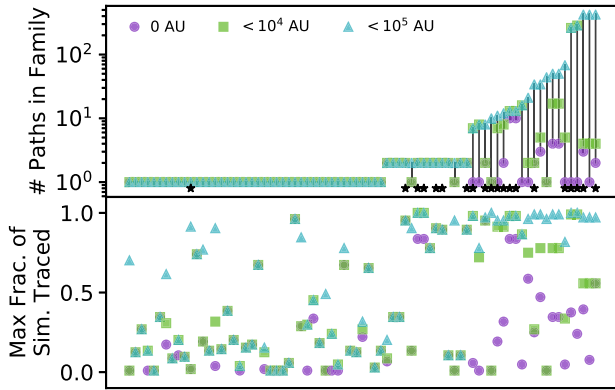


Figure 8. Comparison of different linking distances. The coloured shapes represent different linking distances. The top panel shows the number of paths in a family (the unique combinations of leaf histories identified for a single starting leaf). The horizontal axis is arbitrary and simply serves to order the leaves. Vertical lines connect the path families for a single leaf at different linking distances. Black stars along the bottom indicate the presence of at least one sink particle in a leaf. The bottom panel shows the maximum fraction of the total simulation time traced by a path family; ordering matches the top panel. The choice of linking distance is most significant in dense environments where there are many leaves in close proximity.

while a linking distance of 0 can sometimes artificially truncate a path.

The linking distance becomes an important consideration for leaves in dense environments, which is also where the majority of sink particles reside, including the many bound multiple systems. In these environments, leaves can exhibit significant variation in structure between time-steps, and therefore a very small linking distance will result in frequent premature truncation of paths. However, because of the proximity to many dense structures, it is easy to link two nearby, but not physically interacting, structures, leading to a large increase in the number of paths in a family.

We adopt 10 grid cells as our fiducial linking distance because it allows some variation in the dendrogram leaf contours without leading to linking with many nearby, unassociated leaves. We are still able to identify paths through a substantial fraction of the simulation time, but we do not reach the extremely numerous, and less physically meaningful, path families found with larger linking distances.

4 RESULTS: CORE PROPERTIES AND EVOLUTION

We have identified dense core analogues using dendrograms at multiple simulation outputs and reconstructed the time evolution of these leaves. We now study the broad distributions of leaf properties (such as is frequently done in other work, both observational and computational) and the time evolution of individual leaves in our simulations. We summarize our findings below.

4.1 Core property distributions versus individual core history

We study both the distribution and individual evolution of the large sample of leaves in our simulation. Every parameter distribution we investigate is relatively constant in time. However, the individual evolution of a leaf can be quite variable. We present an example of this dichotomy in Fig. 9. The mass distribution does not vary significantly in time; major variations are only seen at the earliest times when structure is beginning to collapse due to self-gravity.

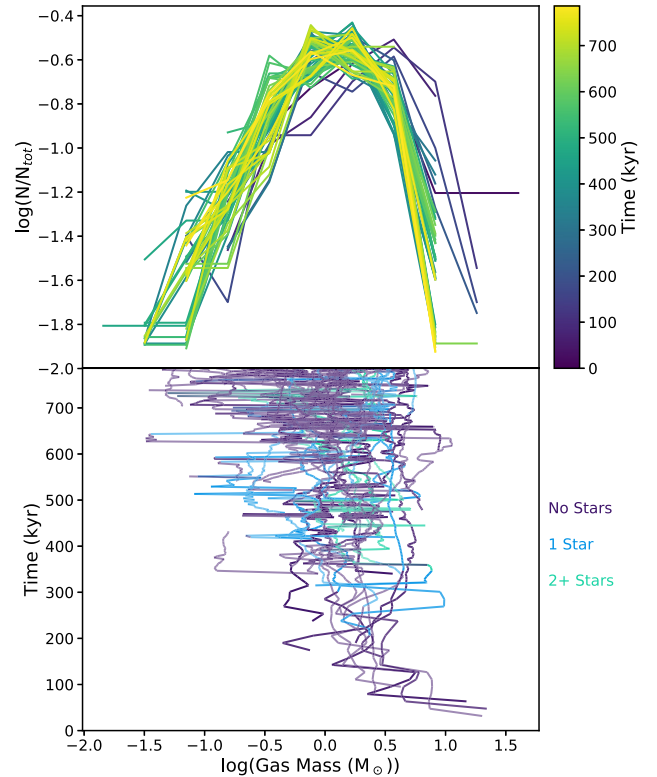


Figure 9. Core mass distribution versus individual core evolution. The top panel shows the distribution of gas mass in the cores across time. The bottom panel shows the gas mass evolution of a subset of reconstructed paths through time; dark purple lines show leaves without sink particles, blue lines show leaves with a single sink particle, and green lines show leaves with multiple sink particles. While the broad distribution of gas mass is nearly invariant in time, any individual leaf may have large variations in its reconstructed history.

The bottom panel reveals that a leaf may have a computed gas mass that can span upwards of an order of magnitude in time, and the typical evolution does not smoothly vary from time to time. Note that in these (and all future figures), time is measured relative to the beginning of the simulation.

To better understand the relative variability in core evolution, we use a parameter called the coefficient of variation (CV), which is defined as the standard deviation of a parameter (σ) divided by the mean (μ) of that parameter. This quantity allows us to directly compare leaf properties of varied units and physical scales and has units of per cent. For our analysis, we consider paths that are tracked for more than 15 kyr and compute the standard deviation of the total path. Because of the rapid evolution in both volume and density of leaves at early times due to the introduction of gravitationally collapsing structure, we exclude the earliest ~ 30 per cent of the simulation from our computation of the CV.

Table 1 presents the minimum, mean, and maximum variation of 16 different parameters: total mass, gas mass, leaf volume, leaf size, mean density, oblateness, virial parameter, the Mach number of gas in the core, the Mach number of the core in the simulation volume, the Alfvén Mach number of the core, the angular momentum magnitude, the variation in the angular momentum orientation, the magnetic field magnitude, the variation in the magnetic field orientation, and plasma β . The definitions of these parameters are presented in Appendix A. While there are few substantial trends to remark upon for individual quantities, for completeness, we report CVs for the entire ensemble of parameters studied in our analysis. We have separated the paths into

Table 1. Coefficient of variance range for different core properties. The CV is defined as the standard deviation divided by the mean of a quantity.

Quantity	Number of paths in a family								
	Min (per cent)	$n = 1$ (34)	Max (per cent)	$1 < n < 10$ (89)			Min (per cent)	$n \geq 10$ (790)	Max (per cent)
		Mean (per cent)		Min (per cent)	Mean (per cent)	Max (per cent)		Mean (per cent)	
Total mass	9	27	53	12	55	213	9	42	153
Gas mass	9	27	53	10	57	279	13	52	144
Volume	1	17	61	10	68	266	18	67	152
Size	1	31	277	4	52	246	6	27	243
Mean density	2	17	96	3	56	196	12	67	244
Oblateness	1	26	89	16	43	92	15	38	74
Virial α	3	27	75	8	49	214	14	41	254
Internal $\mathcal{M}$	1	10	34	4	24	65	4	22	49
Total $\mathcal{M}$	0	6	20	2	11	40	5	13	28
Alfvén $\mathcal{M}$	2	11	30	6	32	104	11	30	72
$ \vec{j} $	3	38	154	18	77	242	10	80	189
$(\text{Max}(\vec{j}) - \text{Min}(\vec{j})) / \text{Mean}(\vec{j})$	0	25	64	5	27	53	11	31	61
$ \vec{B} $	1	12	56	1	27	54	4	24	60
$(\text{Max}(\vec{B}) - \text{Min}(\vec{B})) / \text{Mean}(\vec{B})$	1	17	45	6	29	71	10	31	48
Plasma β	2	16	67	6	51	156	14	43	140
Δ	19	59	272	31	110	270	36	94	262
Quantity	Number of sinks at final time-step								
	Min (per cent)	$n = 0$ (520)	Max (per cent)	Min (per cent)	$n = 1$ (85)	Max (per cent)	Min (per cent)	$n > 1$ (308)	Max (per cent)
		Mean (per cent)			Mean (per cent)			Mean (per cent)	
Total mass	9	54	213	9	44	213	21	25	110
Gas mass	9	54	279	11	51	207	24	48	157
Volume	1	61	266	7	53	190	28	77	196
Size	1	39	277	6	29	246	13	14	31
Mean density	2	53	185	17	61	196	18	83	244
Oblateness	1	36	89	29	51	92	15	38	59
Virial α	3	49	141	9	44	96	15	28	254
Internal $\mathcal{M}$	1	26	55	5	19	61	4	15	65
Total $\mathcal{M}$	0	14	25	3	11	28	6	11	40
Alfvén $\mathcal{M}$	2	19	104	8	28	72	15	47	57
$ \vec{j} $	3	64	189	10	56	242	21	109	185
$(\text{Max}(\vec{j}) - \text{Min}(\vec{j})) / \text{Mean}(\vec{j})$	0	28	64	8	30	61	22	35	53
$ \vec{B} $	1	20	56	11	22	43	8	33	60
$(\text{Max}(\vec{B}) - \text{Min}(\vec{B})) / \text{Mean}(\vec{B})$	1	26	71	12	32	50	13	37	45
Plasma β	2	37	79	3	38	154	14	53	156
Δ	19	90	272	45	85	241	53	105	262
Quantity	Length of path history								
	Min (per cent)	$t < 75$ kyr (35)	Max (per cent)	$75 \text{ kyr} < t < 250 \text{ kyr}$ (21)			Min (per cent)	$t > 250$ kyr (857)	Max (per cent)
		Mean (per cent)		Min (per cent)	Mean (per cent)	Max (per cent)		Mean (per cent)	
Total mass	9	54	213	10	42	81	9	43	153
Gas mass	9	62	279	10	41	81	11	51	144
Volume	1	60	266	2	36	102	5	67	152
Size	1	29	217	6	44	171	6	29	277
Mean density	2	35	196	3	37	110	5	65	244
Oblateness	1	31	62	9	41	92	10	38	76
Virial α	3	35	254	14	47	141	8	42	144
Internal $\mathcal{M}$	1	14	50	5	14	32	2	22	65
Total $\mathcal{M}$	0	6	18	1	9	17	2	13	40
Alfvén $\mathcal{M}$	2	27	104	4	17	38	7	30	72
$ \vec{j} $	3	52	232	5	59	154	10	80	242
$(\text{Max}(\vec{j}) - \text{Min}(\vec{j})) / \text{Mean}(\vec{j})$	0	24	52	8	22	64	7	31	61

Table 1 – continued

Quantity	Number of paths in a family								
	Min (per cent)	$n = 1$ (34) Mean (per cent)	Max (per cent)	Min (per cent)	$1 < n < 10$ (89) Mean (per cent)	Max (per cent)	Min (per cent)	$n \geq 10$ (790) Mean (per cent)	Max (per cent)
$ \vec{B} $	1	17	48	5	15	45	4	25	60
$(\text{Max}(\vec{B}) - \text{Min}(\vec{B})) / \text{Mean}(\vec{B})$	1	25	71	7	22	43	8	30	50
Plasma β	2	36	154	3	26	67	15	43	156
Δ	19	85	270	31	80	166	34	95	272

three bins in each section. Under each bin is the number of individual paths that fall into the bin. The final line in each table section is Δ , which is the spread in CV (maximum CV minus minimum CV) for the collection of quantities and is designed to show the variation in variability for each bin.

4.1.1 Size of path family

We first split our full contingent of paths by the number of paths in a family. Paths with $n = 1$ are isolated; they typically show the least variation. However, these paths are, on average, shorter than paths in other bins that can lead to suppressed variation. Paths with $n \geq 10$ are typically in very dense environments and are therefore most susceptible to being linked to multiple nearby leaves. This can cause variations to be artificially high as physically unassociated cores (overdensities that do not physically interact in space) are linked in the same path; the large CV of volume in the non-isolated paths hints that leaf contour changes (arising from structures bouncing above and below the structure refinement threshold due to minute changes in the local density field) may cause the large CVs in other parameters.

4.1.2 Number of sinks

We then group paths by the number of sinks in the leaf at the final time-step. Paths with $n = 0$ are starless overdensities. The starless paths with low CVs are typically short-lived, low-density leaves. The paths containing multiple sink particles are frequently part of large path families in dense regions where physically independent overdensities are identified as related due to a temporary co-location or dendrogram contour changes that cause multiply linked leaves.

4.1.3 Length of path

Finally, we divide the ensemble of paths by their length. The shortest paths (< 75 kyr) are frequently temporary overdensities and therefore have little physical evolution over the time they are traced as indicated by the low CVs. However, there are also short paths with high variability that belong to a large path family. The longest paths (lifetimes greater than 250 kyr) have little correlation with the size of the path family or the presence of stars, so the CVs in the final bin span a large range.

4.1.4 General trends

The path histories we trace have significant variation – frequently upwards of 40 per cent in CV. In all three methods of dividing paths presented in the table, the average CV increases from left to right, meaning that shorter lived, isolated, starless cores tend to have less

Table 2. Coefficient of variance range for isolated, long-lived, starless cores.

Quantity	Min (per cent)	Mean (per cent)	Max (per cent)
Total mass	10	30	53
Gas mass	10	30	53
Volume	2	18	61
Size	6	45	277
Mean density	3	17	54
Oblateness	9	33	89
Virial α	14	35	75
Internal $\mathcal{M}$	2	12	27
Total $\mathcal{M}$	1	8	20
Alfvén $\mathcal{M}$	4	15	30
$ \vec{j} $	5	48	154
$(\text{Max}(\vec{j}) - \text{Min}(\vec{j})) / \text{Mean}(\vec{j})$	8	28	64
$ \vec{B} $	5	14	56
$(\text{Max}(\vec{B}) - \text{Min}(\vec{B})) / \text{Mean}(\vec{B})$	7	21	45
Plasma β	6	21	67

variability. However, the maximum CV does not show the same trend, indicating that any given path can vary significantly.

It is also important to note that a low variability in one parameter does not indicate low variability in all parameters. This is demonstrated with the parameter Δ at the bottom of each section. This quantity shows the maximum difference in the CV of the 16 parameters for each leaf, or the maximum variation in the variation of our computed properties. The average Δ in all cases is over 50 per cent, meaning that the majority of paths have little correlation in the amount of variability in different quantities. Thus, the computed properties of observed overdensities identified with dendrograms may not correlate well with the physical evolution of the bound core itself.

4.2 Isolated, starless cores

Naively, we might expect the long-lived, isolated, starless cores in our simulations to show the least variation. Observed lower density cores can have lifetimes of upwards of 1 Myr, which suggests that these cores should vary slowly over their lifetime if they are free from external influence (André et al. 2014). Thus, we analyse these isolated cores separately; Table 2 shows the coefficient of variance for the 18 isolated (one path in the family), long-lived ($t > 75$ kyr), starless leaves. While some vary by only a few per cent, other paths have variability of CV > 40 per cent for the 15 parameters in the table. This indicates that these leaves have large computed variability, in contrast to our naive expectation. The individual leaf evolution tracks are shown in Fig. 10, where we plot mean density, mass, volume, and virial parameter. Most leaves show fairly large stochastic variations

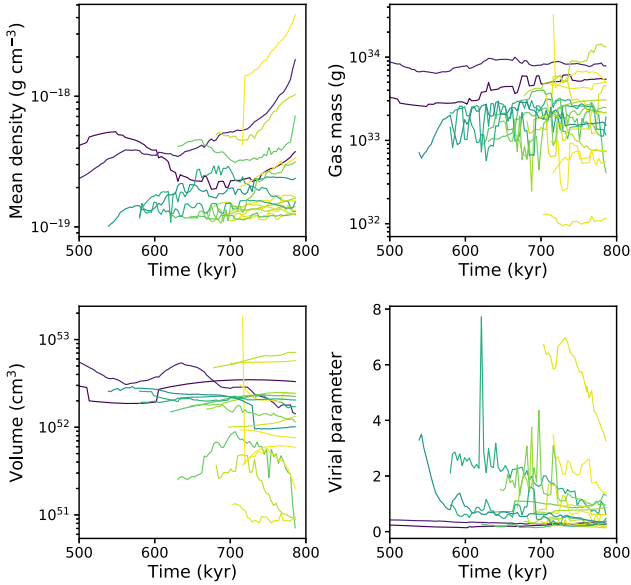


Figure 10. Individual core evolution for the long-lived, isolated, starless cores in our simulation. We show mean density, gas mass, volume, and virial parameter in the four panels. Most cores show stochastic variation of the order of a factor of a few over the course of their lifetimes.

in individual quantities on short time-scales. These variations are commonly due to changes in the physical structures included in the dendrogram leaf rather than significant physical evolution. However, a few of our leaves (namely, those plotted in purple), show relatively quiescent evolution over their full lifetime, which are akin to the structures identified in Chen et al. (2019). These quiescent cores will contribute to the statistics of core property distributions while not participating in the star formation process, thereby confusing the mapping of the CMF to the IMF of stars (e.g. Offner et al. 2014).

4.3 Virial evolution of cores

Despite the wide variability in the time evolution of other core properties identified with dendrograms, the virial evolution of leaves does trend in the expected direction of lower virial numbers with time as seen in the lower panel of Fig. 11. This fits the classic view of star formation where a magnetized core undergoing global gravitational contraction will eventually become supercritical and collapse to form a star (Mouschovias & Spitzer 1976). However, our core tracking algorithm does not find strong evidence that a leaf with low virial α will form a star as shown in Fig. 11. Note that we are using the simplified gravitational α (which is frequently used in observations, e.g. Kirk et al. 2017) and not computing the full virial α that includes boundary terms. Most leaves (>70 per cent), especially those hosting sink particles, do finish the simulations with $\alpha < 2$. However, a substantial population of the long-lived, starless leaves have $\alpha < 2$ as well. Many of these low- α leaves persist for longer than a local free-fall time (a few hundred kyr) without forming a star. Thus, virial α is not necessarily the best predictor of future star formation; other physics, such as pressure or magnetic support, are important factors in the global evolution of a core.

4.4 Short-lived overdensities

We observe a population of short-lived, low-density peaks arising from turbulent flows that contribute a level of ‘noise’ to the in-

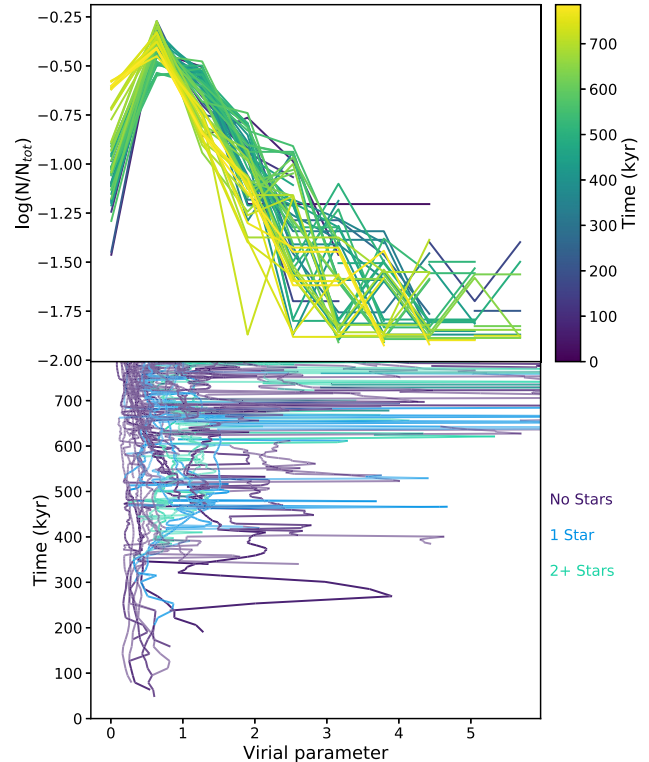


Figure 11. Core property distribution versus individual core evolution of the virial parameter. The top panel shows the distribution of the virial parameter of the cores across time. The bottom panel shows the virial evolution of a subset of reconstructed paths through time. Colouring is the same as Fig. 9.

terpretation of long-term core evolution, since they do not go on to collapse and form protostars. These overdensities account for about 25 percent of path families identified when tracing paths forward from an intermediate time-step. We identify this temporary population of ‘imposter cores’ as having lifetimes less than 200 kyr and densities less than $1 \times 10^{-18} \text{ g cm}^{-3}$ ($2.5 \times 10^5 \text{ cm}^{-3}$), as can be seen in Fig. 12. The majority of isolated paths occupy a much lower density and shorter lifetime than the general path population. Most of the paths in the low-density and short lifetime region also show the trend that the maximum density (which is typically also less than $10^{-18} \text{ g cm}^{-3}$) is higher than the last identified density (the ending density of the path), suggesting that these leaves are physically temporary overdensities that decay below the threshold density required to be identified in the dendrogram. The free-fall time of these overdensities is about 100 kyr; because the free-fall time is roughly equivalent to the overdensity lifetime, these objects are not dominated by gravitational collapse.

The presence of this substantial population of imposter cores could introduce a bias in the instantaneous CMF. These overdensities have gas masses of order one solar mass and sizes of roughly a tenth of a parsec, which is similar to masses and sizes of observed cores and may therefore masquerade as pre-star-forming cores. Thus, at any given time, roughly 15–25 percent of cores identified in a region may be from this temporary population. We performed a two-sample Kolmogorov–Smirnov (K–S) test on the computed core properties of imposter cores compared to all other cores at the same time: we could not distinguish differences in the distributions of any parameter except density. For instance, gas mass and virial parameter both have a p -value of 3 per cent and K–S statistic of 0.4; if the p -value is high (our preferred cut-off is more than 1 per cent) and the K–S statistic is

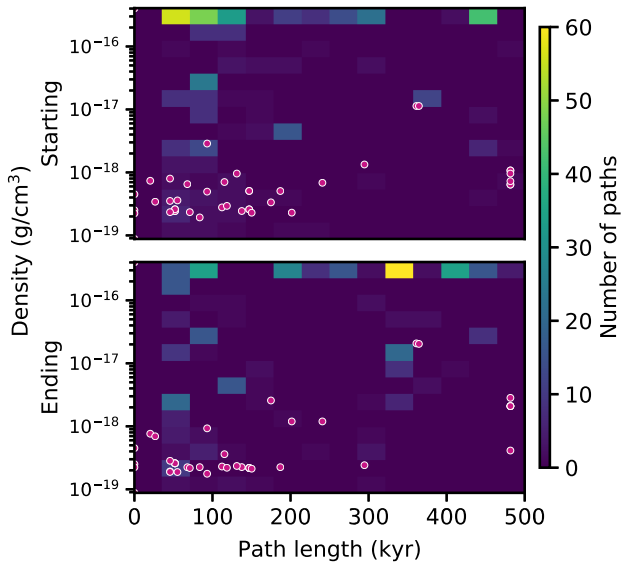


Figure 12. Starting and ending densities of isolated objects. The background histogram shows the starting and ending peak densities of all paths (top and bottom, respectively) plotted against the length of the path. The purple points denote isolated paths. There is a substantial population of low-density leaves with lifetimes less than 150 kyr (25 percent of all path families) that are temporary, non-gravitating overdensities that disperse and fall below the dendrogram floor.

low (our preferred cut-off is less than 0.6), we cannot claim that the two samples are drawn from different populations. Indeed, p -values for our computed parameters except for density are typically above 3–5 percent and K–S statistics are typically less than 0.5. Thus, imposter cores are not easily separated from any other population of cores, so they will complicate the correlation of core and stellar properties. These temporary overdensities are explored in more detail in Chen et al. (2020).

4.5 Core mass function

We measure the CMF of our simulation. We show our CMF through time in Fig. 13 together with mass functions from the literature. We compute the total core mass, which includes mass from both gas and sink particles. We compare our CMF to the fiducial CMF from Guszejnov & Hopkins (2015), the observed CMF from Alves, Lombardi & Lada (2007), the IMF inferred from observations from Chabrier (2003), and a lognormal distribution. The CMF in our simulation agrees well with the observed CMF of the Pipe dark cloud from Alves et al. (2007) and with other observational CMFs not plotted here such as those observed in Perseus, Serpens, and Ophiuchus from Enoch et al. (2008), Aquila from André et al. (2010) and Könyves et al. (2015), and Vela C from Massi et al. (2019). We also find a good agreement with a Chabrier (2003) system IMF scaled by a factor of 6. We do not, however, create the population of low-mass cores of the Guszejnov & Hopkins (2015) model. We also do not create the population of low-mass sinks particles seen by other simulations such as Bate (2012) despite our ability to create sink particles with masses much less than $1 M_{\odot}$ (although this is expected due to the coarse spatial resolution).

As is seen in the figure, our CMF is relatively invariant through time. Our peak mass is constant at around $1.4 M_{\odot}$ with a range from about 0.3 – $10 M_{\odot}$. There is a small apparent bias towards higher masses at earlier times, which is a by-product of low number statistics

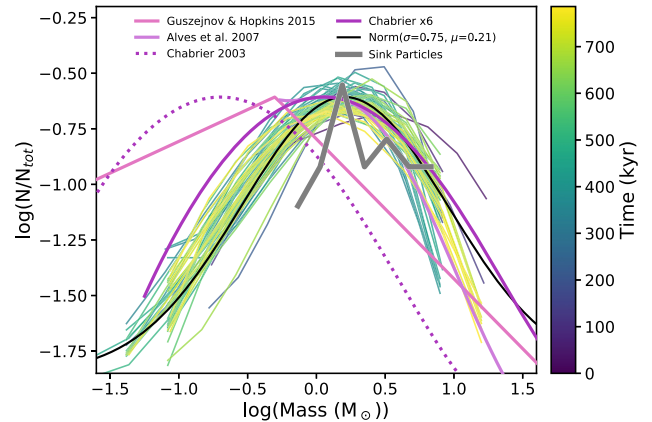


Figure 13. The CMF across time in our simulations. The purple to yellow colour scale shows our normalized core masses for a selection of time-steps. We plot the total core mass, which includes mass from both gas and sink particles. The mass function of our sink particles (which are equivalent to a protostar and compact disc) is shown in the thick grey line. We also show the fiducial CMF from Guszejnov & Hopkins (2015), the observed CMF from Alves et al. (2007), and the IMF inferred from observations from Chabrier (2003). We plot a lognormal in black.

and the lack of significantly refined structure in our simulations shortly after gravity is turned on. The constant nature of the CMF is likely due to two effects. The dendrogram introduces new leaves when temporary overdensities are significant enough to warrant leaf creation, leading to the transient population of low-mass ‘cores’ discussed above that balances the small physical growth of persistent cores and the algorithmic fragmentation of more massive leaves into smaller structures. The trend of nearly constant CMF across time in a singular environment is seen in other simulations, such as Cunningham et al. (2018), where core mass distributions do not show significant mass evolution after formation. As described above, any given leaf can occupy a wide range of the total mass space as we track it through time, but the ensemble of leaves maintains a constant distribution in time. Thus, the CMF derived from a dendrogram population does not necessarily correlate with the final IMF of the region; the stochastic nature of leaf mass evolution makes it very difficult to compute a relationship between a core mass at any snapshot and the resultant stellar mass.

We compare the CMF derived from the leaf gas mass to the CMF derived from more observationally motivated core definitions including the leaves that eventually form stars (equivalent to prestellar and protostellar cores), leaves with $\alpha < 2$, and leaves that are Jeans unstable for their mean density. All of these different populations produce quantitatively similar results, as shown in Fig. 14. The CMFs all have peaks slightly higher than $1 M_{\odot}$, a spread of about two orders of magnitude, and are invariant in time. We therefore conclude that any structures not involved in the traditional star formation process (e.g. transient overdensities) have little impact on the derived CMF. Computing the CMF based on different properties in observations also produce similarly invariant CMFs (e.g. Sokol et al. 2019).

The highly variable masses of identified leaves through time means that we cannot infer the IMF by looking at a population of cores identified with dendrograms at a given time snapshot. While there may be an underlying physical evolution of star-forming cores, the instantaneous properties of a region identified by a dendrogram cannot be assured to correlate with that evolution. We reiterate that our goal in this work is not to measure CMFs in detail but to (i) show that our results are comparable to observations and (ii) illustrate that

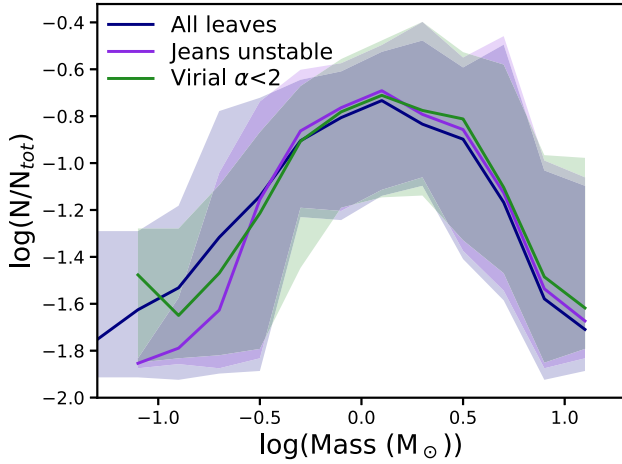


Figure 14. The core mass function for different core definitions. Solid lines show an average CMF across time, while the shaded regions show the minimum and maximum bins across time. Blue shows the CMF of all leaves in the simulation (which is what is shown in Fig. 13). The other colours show core selections that might be more physically motivated: purple shows cores that have masses greater than the local Jeans mass, and green shows cores that have virial $\alpha < 2$. All mean CMFs overlap except at the lowest mass end, where some of the low-mass cores do not satisfy the stricter Jeans or Virial criteria.

the apparent similarities between observed CMFs and IMFs may not derive from a simplistic evolution of the former to the latter.

5 RESULTS: INSIGHTS FROM METHODOLOGY

It is imperative to understand the impact of the dendrogram algorithm on cores identified in both simulations and observations due to the algorithm’s wide popularity in the literature. In this section, we discuss the insights into the use of dendrograms gained from this work.

We have used dendrograms to identify dense structures in our simulations of a star-forming region, but we have also demonstrated a limitation of dendrograms: because dendrograms identify *relative* variations in structure, leaf structure may vary significantly between time-steps due to small variations in the local density structures. Dendrograms are built beginning from the maximum value, so any variations in that maximum may cascade into substantial changes in the resultant dendrogram architecture.

An example of this phenomenon is shown in Fig. 15. The left and right columns depict neighbouring time-steps. Despite very little physical evolution between time-steps (a ~ 5 per cent change in the peak density), the dendrogram identifies leaf structure quite differently. This translates to a nearly order of magnitude variation in the volume of the sink-hosting leaf and substantial variation in the computed properties of the leaf.

The two leaves in the right-hand panel that are part of the sink-hosting subtree (the two leftmost leaves in the dendrogram) are not physically interacting over the course of the simulation. They are simply nearby overdensities. However, because of the variations in the dendrogram structure, our algorithm identifies these two leaves as belonging to the same path family. Thus, one of the major failings of tracking overdensities identified via dendrogram through a simulation to study core evolution is that it becomes difficult to disentangle physical evolution from ‘algorithmic’ artefacts. In other words, there is no easy, automated way to differentiate between physical structure change and dendrogram structure change.

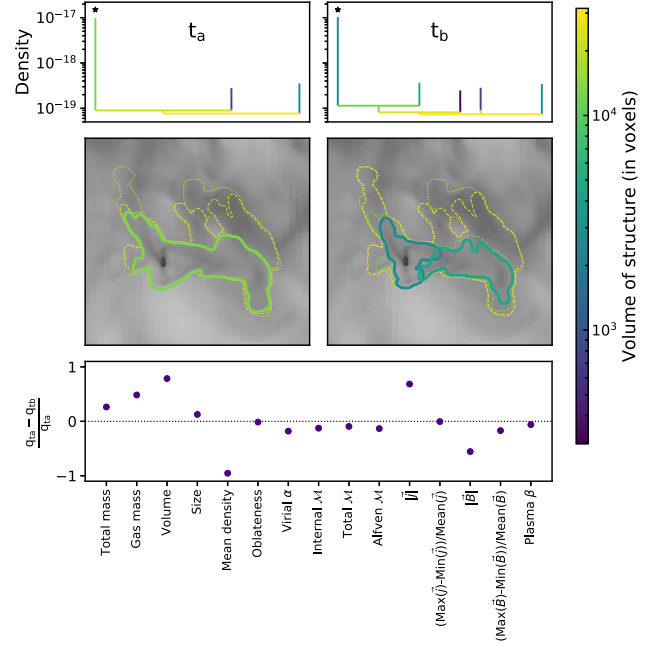


Figure 15. The dendrograms of a small volume in consecutive time-steps. The left and right columns depict different time-steps. The upper panels show the dendrogram structure coloured by leaf volume. The starred leaf in each panel is the leaf containing the dominant overdensity in the middle panels. The middle panels show the leaf contours over a grey-scale density projection of the simulation. Dotted contours show the trunk, dashed contours show branches, and solid contours show leaves. Despite very little physical evolution between time-steps, the dendrogram identifies different tree structure, leading to significantly different leaf morphologies. The bottom panel shows the impact of the different leaf structure on computed leaf properties for the leaf starred in the upper panels. Critical quantities such as mass show significant differences between the two times that can only be attributed to the redefinition of the leaf contours.

The change in consecutive dendrograms arises because of small variation in the relative properties of structures (typically intermediate density structure). Fig. 16 aims to illustrate the issue. Structures 1, 2, and 3 are shown at two consecutive times. The physical properties of the structures (peak and width, in this cartoon) do not change between the top (earlier time) and bottom (later time). However, their relative locations with respect to one another do change. The structures have moved closer to one another and therefore the saddle point between them has become shallower. This causes the nodes (horizontal lines) to be at different heights at the two different times. At the earlier time, the node is low enough that both structures 2 and 3 exceed the density increase criterion (indicated by the pink vertical lines), while at the later times, the node is at a high enough density that the individual density peaks are not significant enough to allow substructure to be identified.

To further explain the example presented from this work, the second leaf in the right-hand panel of Fig. 15 is just above the density refinement criterion at its physical location in that time-step. However, the peak density in that region drops by 4 per cent in the left-hand panel, which then leads to the overdensity not being quite ‘peaky’ enough to satisfy the density refinement when compared with the maximum peak. This is not a problem unique to our density refinement criterion: any density refinement chosen will exhibit these artefacts to some degree due to the relative nature of dendrogram structure identification. Even observationally, these issues may be seen: differences in resolution or noise levels in observations of the

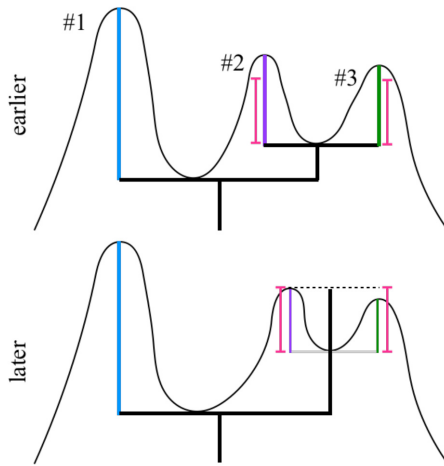


Figure 16. Cartoon explanation of the origin of algorithmic structure variation in time. Structures 1 (blue), 2 (purple), and 3 (green) are shown at two consecutive times. The physical properties of the structures (peak and width) do not change between the top (earlier time) and bottom (later time). However, their relative locations with respect to one another do change: structures 2 and 3 move closer to one another, thereby increasing the density of the saddle point between them. This causes the nodes (horizontal black lines) to be at different heights at the two different times. At the earlier time, both structures 2 and 3 exceed the density increase criterion (indicated by the pink vertical lines), while at the later times, the individual density peaks of 2 and 3 are not significant enough to allow for substructure to be identified.

same region may lead to changes in the computed hierarchy. Any variation between consecutive observations in the region around a peak dendrogram can lead to variations in the contour drawn by the dendrogram.

6 DISCUSSION

6.1 Interpreting the IMF from the CMF

One natural question we can ask in this work is how the instantaneous core masses correlate with the stars they form, which can provocatively be described as translating a CMF into an IMF. We plot this in Fig. 17, where we show the total sink mass at the end against the initial masses of leaves that merge into the final overdensity. First, there is a wide array of scatter in the initial leaf masses that does not correlate well with the final sink mass; some of this scatter may result from the lack of feedback in our simulations, which is expected to create overmassive sink particles, and some may arise from the fact that a protostar can accrete from outside the initially bound gas core. This observation is consistent with other works for low- to intermediate-mass stars such as Smith, Clark & Bonnell (2009) and Mairs et al. (2014). Second, the *sum* of all component leaf masses seems to be a very important consideration, especially when considering the growth of systems containing multiples.

The young, typically singular, sink particles fall above the 1:1 correlation in Fig. 17, indicating that the nascent overdensity is the primary reservoir of gas that accretes on to the protostar by the final snapshot of the simulation. Older sinks and systems containing multiple sinks frequently have individual leaf masses below the 1:1 line. Because protostellar outflow feedback is expected to reduce stellar masses by at least a factor of 3 compared to the case without feedback (Offner et al. 2014; Offner & Chaban 2017), we also indicate the 3:1 sink:leaf mass ratio with a dot-dashed line. For the most massive multiple systems, the component leaves (which

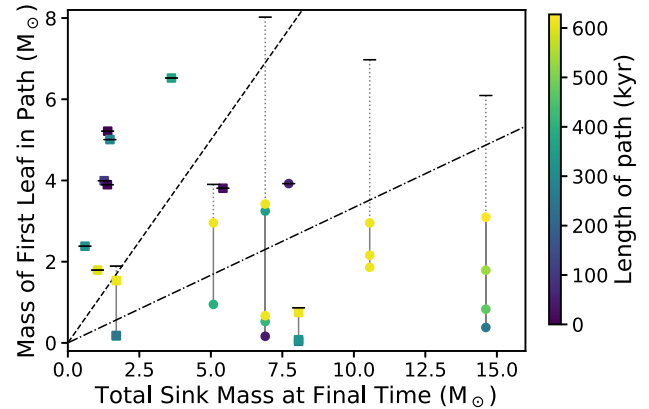


Figure 17. Sink mass at the final time plotted against leaf mass. Squares show leaves containing one star at the final time, while circles show leaves containing multiple stars. Points are coloured by the length of the path, and we have truncated the paths to have an earliest age of 250 kyr after the gravity turned on. Systems that consist of multiple paths are connected by a vertical line. The sum of the leaf masses is indicated by the horizontal marker. The dashed line shows the 1:1 correlation, while the dot-dashed line shows the trend if the sink mass is reduced by the factor of ~ 3 arising from the lack of protostellar feedback in the simulations. For instance, the rightmost system (currently at $15 M_{\odot}$) would likely have a mass of about $5 M_{\odot}$ if feedback was included in these simulations. Additionally, the inherent variability in leaf snapshots can cause a clearly unphysical mass budget for a protostellar system.

constitute the idealistic ‘gas reservoir’ for the protostars) still do not lie above the aforementioned feedback relation, but the sum of the leaf masses puts these systems into a comparable space as other systems. For instance, if the expected reduction in mass due to feedback was included, the rightmost system in Fig. 17 would likely have a total sink mass of $\sim 5 M_{\odot}$ and a combined leaf (core/reservoir) mass of $\sim 6 M_{\odot}$; however, all component leaves for that system that would be measured in a CMF have masses $\lesssim 3 M_{\odot}$. The leaves would then appear to contain sufficient mass to form the stars, but we caution that this, and even the simple 3:1 relationship, is an oversimplification of the nature of protostellar feedback and its cumulative effect on the surrounding gas.

It is important to note that the systems containing multiple stars do not form from multiple leaves containing single stars merging together; rather, the bound multiple forms in one leaf, and the subsequent accretion of gas overdensities may help trigger the formation of new stars. This result is similar to that seen in Padoan et al. (2019), who see little correlation between core mass (or even extended mass around a core) and stellar mass for high-mass stars. This observation supports binary formation models such as turbulent fragmentation in a single core (e.g. Offner et al. 2010) or disc fragmentation (e.g. Kratter et al. 2010) as opposed to dynamical capture (e.g. Bate et al. 2003).

6.2 Other ways to identify cores

There are a few ways one may attempt to overcome the limitations of core identification and comparison using density dendrograms. We discuss the advantages and limitations of these ideas.

One could try to overcome the impositions of the dendrogram algorithm itself. Custom merging strategies are possible in ASTRO-DENDRO; a ‘pruning’ strategy will allow peaks near the density refinement criterion to remain more stable and thereby overcome some of the relative structure variation shown in Section 5. However,

this solution is subject to human bias due to the addition of another tunable parameter. It is unclear how to create a custom merger strategy in a way that is agnostic to the human-desired structure without introducing more bias.

In principle, one could also create contours at absolute density levels instead of relying on a relative measure. By using an absolute density contour, the leaf structure should slowly vary from time-step to time-step and may therefore better identify bound cores. One could then create a hierarchical structure tree that is similar to a dendrogram, but the nature of hierarchy would be more difficult to determine due to the fact that many density peaks would be broken into a single nested hierarchy. Additionally, this type of hierarchy would destroy the physical utility of dendrograms in studying the relation of physical structures in a region and again relies on an arbitrary density threshold, which we advocate against.

Core identification in simulations might also be better done in 2D synthetic observation space instead of 3D density grids because there will be fewer variations in integrated intensity between time-steps. However, this method is best suited to isolated cores and may suffer from false overdensities in emission created by chance alignments (e.g. Beaumont et al. 2014).

Finally, density may not even be the best tracer of star-forming cores as cores are highly dynamic and will not be defined by the same density contour across time. It is likely that a more physically motivated property such as virial parameter, velocity dispersion, or gravitational potential could be a better quantity with which to build hierarchical structures (see, for instance, Mao, Ostriker & Kim 2020). These properties should be less variable across time and should therefore provide a more stable core identification. However, these quantities are more difficult to measure observationally and will make comparisons between simulations and observations harder.

In simulations, one can also include tracer particles that will trace the evolution of gas in a core identification-independent way. However, interpreting that evolution is non-trivial. Smith et al. (2009) and others find that a non-negligible fraction of tracer particles in a bound gas clump will accrete on to a sink particle outside of that bound clump. Indeed, they find that most of the mass in a sink particle can be accreted from outside its nascent core. Thus, the meaning of a core in this context becomes even less apparent, as the star may contain gas from all around the molecular cloud.

One could employ alternate core identification algorithms used in the field. All of these other methods (CLUMPFIND, FellWalker, GaussClumps, etc.), would suffer similar issues because they all fundamentally rely on the relative positions and heights of peaks to determine structure. Different algorithms might have different sensitivity to the less dense material surrounding density peaks, but all algorithms have some way to combine peaks that are thought to not be independent. In a simple test in which we identified structures in two consecutive time-steps with dendrograms, CLUMPFIND, FellWalker, and Reinhold, all core identification algorithms produced structures that broadly followed the 1:1 trend in volume, but all algorithms had $\sim$ order of magnitude deviations in volume for at least a few structures. The dendrogram and FellWalker algorithms had the least dispersion along the 1:1 relation in this test, but the substantial redefinition of ‘core’ boundaries was still observed.

Each of the methods discussed above would likely identify the same dense gas structures, but the variations in core identification would still likely lead to changes in computed core properties between methodologies. Further work is needed to explore the full impact and limitations of these different structure identification methods in the time domain. However, this work suggests that there

is no unique way to define a core in both simulations and observations using existing methods.

6.3 Implications of core identification

This work shows that there is no time-stable density contour with which to define cores. Because of the dynamic nature of core evolution, a single set of dendrogram parameters will not trace unique core parameters across the entire lifetime of core formation. Additionally, we show that a substantial change in the cloud properties (due to time evolution in this case) are required to see changes in the observed CMF: over >70 per cent of our simulation snapshots show the same CMF, despite order unity variations in individual cores. Changes in the distribution occur at early times. In the context of our simulations, this is because gravity has had less time to overcome the turbulence in the gas. In real systems, this would correspond to the time when the cloud itself was only weakly bound. This trend suggests that variations in the CMF only coarsely trace the time evolution of a star-forming region. Thus variations in the CMF from one star-forming region to should not be attributed solely to differences in age. Finally, computing a dendrogram in density or intensity on an observed region introduces an inherent uncertainty in the physical importance of structures identified. Dendrograms have many tunable parameters, so disentangling physical structure from algorithmically imposed structure in an automated fashion is a non-trivial endeavour.

The large variability in the computed core boundaries will likely be less dramatic in observational space due to the integration of the signal along a line of sight. The lower density material around the edges of our identified leaves will not contribute as much signal, so structures will appear more compact around only the densest part of the core. However, as Beaumont et al. (2013) and citations therein show, simulation projections and observation are highly subject to projection effects, such as non-physical cores being identified due to a large column of low-density material. Thus, neither physical nor observational spaces have cores that can be robustly and uniquely defined across all time.

7 CONCLUSIONS

In this work, we have presented an algorithm that links dendrogram leaves through time in order to study the evolution of dense cores in MHD simulations. We aim to understand not only the evolution of the star-forming gas reservoir in our simulations, but also the manner in which the use of the dendrogram algorithm may bias interpretation of core properties and evolution. Ideally, the parameters used for identifying and linking cores are set by the underlying physics. As is shown in this work, we ultimately conclude that there is no robust set of density-based parameters that can trace coherent cores through time. Additionally, we find the following:

(i) The distributions of core properties, such as mass, are relatively invariant in time. The CMF matches well with observed CMF distributions such as Alves et al. (2007) and shifted IMFs such as Chabrier (2003). Most property distributions do not show significant trends over long time-scales.

(ii) Individual core histories show large variability (>40 per cent) on short time-scales (<100 kyr) that arise from changes in the leaf boundaries. This non-monotonic variability persists across environment (isolated or crowded) and stellar content. Additionally, a leaf history that shows low variability in one parameter will not necessarily show low variability in all parameters. There are no obvious regular trends in time with the exception of virial parameter

(which tends to decrease to $\alpha < 2$ as the cores reach the end of the simulation). There is some evidence for long-term evolution in of individual paths traced in other properties that may correspond to physical evolution, but the shorter stochastic variability makes these trends difficult to quantify.

(iii) The variability exhibited in our analysis of individual core evolution is at least partially attributable to the dendrogram algorithm itself. Small changes in the relative structure of the density between time-steps can propagate to incredibly large changes in the computed boundaries of structures. In extreme cases, volumes can change by an order of magnitude between time-steps, leading to nearly 100 per cent variability in computed core properties. The sensitivity of the dendrogram to small changes in physical conditions raises concerns about hierarchies identified in both simulations and observations. For instance, changes in noise or resolution may lead to different hierarchies in the same region.

(iv) We find a population of short-lived overdensities in each time-step that may serve as a substantial source of ‘noise’ for core property distributions in observations. The overdensities tend to have lower density ($< 10^{-18} \text{ g cm}^{-3}$) and lifetimes less than 200 kyr, and they account for 15–25 per cent of identified leaves every time-step. These overdensities have other properties, such as mass and size, that are comparable to other cores in the simulation that go on to form stars.

(v) Assessing the full history of cores (including events like mergers) may be important for interpreting the IMF. We find that, especially for massive multiple star systems, the sum of all initial leaves associated with the multiple is typically required to agree with CMF–IMF scaling assumptions even when inefficiency produced by feedback is taken into account.

(vi) There is no time-stable density contour that defines a star-forming core. The dynamic nature of core formation and evolution means that dendrograms will not trace the same structures across time in a reliable way. Thus, we urge caution when comparing dendrograms of different ages or environments because differences in the dendrogram may come from the algorithm itself instead of physical changes.

In summary, cores identified with dendrograms are subject to algorithmic limitations that impact the physical interpretation of ‘observed’ core boundaries. And yet, understanding the full time evolution of star-forming cores is critical to understanding the end results of star formation, such as interpreting the relationship (or lack thereof) between the CMF and IMF. We have shown the need for caution when extrapolating instantaneous observations of star-forming cores either forward or backward in time, as cores can have substantial variability both intrinsically and observationally.

ACKNOWLEDGEMENTS

We are grateful for the helpful suggestions of our anonymous referee. We also thank others, including Jaime Pineda and Alyssa Goodman, for insightful discussions during the course of this work. RAS acknowledges support from the National Science Foundation under Grant No. DGE-1143953. ATL, SSRO, and KMK acknowledge support from NASA grant NNX15AT05G. KMK additionally acknowledges support from NASA grant 80NSSC18K0726 and RCSA award ID 26077. SSRO and HHC acknowledge support from a Cottrell Scholar Award from Research Corporation and NSF Career Grant AST-1748571. The analysis presented herein utilized HPC resources supported by the University of Arizona TRIF, UITS, and RDI and maintained by the UA Research Technologies department. This research made use of ASTRODENDRO, a Python

package to compute dendrograms of astronomical data, and the software package yt from Turk et al. (2010).

DATA AVAILABILITY

The data underlying this article will be shared on reasonable request to the corresponding author.

REFERENCES

- Alves J., Lombardi M., Lada C. J., 2007, *A&A*, 462, L17
 André P. et al., 2010, *A&A*, 518, L102
 André P., Di Francesco J., Ward-Thompson D., Inutsuka S.-I., Pudritz R. E., Pineda J. E., 2014, *Protostars and Planets VI*, The University of Arizona Press, Tucson. p. 27
 Arzoumanian D., André P., Peretto N., Könyves V., 2013, *A&A*, 553, A119
 Arzoumanian D. et al., 2019, *A&A*, 621, A42
 Ballesteros-Paredes J., Low M.-M. M., 2002, *ApJ*, 570, 734
 Bate M. R., 2009, *MNRAS*, 392, 1363
 Bate M. R., 2012, *MNRAS*, 419, 3115
 Bate M. R., Bonnell I. A., Bromm V., 2003, *MNRAS*, 339, 577
 Beaumont C. N., Offner S. S. R., Shetty R., Glover S. C. O., Goodman A. A., 2013, *ApJ*, 777, 173
 Beaumont C. N., Goodman A. A., Kendrew S., Williams J. P., Simpson R., 2014, *ApJS*, 214, 3
 Berry D. S., 2015, *Astron. Comput.*, 10, 22
 Blitz L., Stark A. A., 1986, *ApJ*, 300, L89
 Bonnell I. A., Bate M. R., Clarke C. J., Pringle J. E., 2001, *MNRAS*, 323, 785
 Boyden R. D., Koch E. W., Rosolowsky E. W., Offner S. S. R., 2016, *ApJ*, 833, 233
 Boyden R. D., Offner S. S. R., Koch E. W., Rosolowsky E. W., 2018, *ApJ*, 860, 157
 Burkhardt B., Lazarian A., Goodman A., Rosolowsky E., 2013, *ApJ*, 770, 141
 Chabrier G., 2003, *PASP*, 115, 763
 Chen C.-Y., Ostriker E. C., 2018, *ApJ*, 865, 34
 Chen H. H.-H., Burkhardt B., Goodman A., Collins D. C., 2018, *ApJ*, 859, 162
 Chen H. H.-H. et al., 2019, *ApJ*, 877, 93
 Chen H. H.-H., Offner S. S. R., Pineda J. E., Goodman A. A., Burkert A., Ginsburg A., Choudhury S., 2020, preprint ([arXiv:2006.07325](https://arxiv.org/abs/2006.07325))
 Clark P. C., Klessen R. S., Bonnell I. A., 2007, *MNRAS*, 379, 57
 Cunningham A. J., Krumholz M. R., McKee C. F., Klein R. I., 2018, *MNRAS*, 476, 771
 Dib S., Schmeja S., Hony S., 2017, *MNRAS*, 464, 1738
 Enoch M. L., Li N. J. E., Sargent A. I., Glenn J., Rosolowsky E., Myers P., 2008, *ApJ*, 684, 1240
 Flower D. R., des Forêts G. P., Walmsley C. M., 2006, *A&A*, 456, 215
 Friesen R. K., Bourke T. L., Francesco J. D., Gutermuth R., Myers P. C., 2016, *ApJ*, 833, 204
 Goodman A. A., Rosolowsky E. W., Borkin M. A., Foster J. B., Halle M., Kauffmann J., Pineda J. E., 2009, *Nature*, 457, 63
 Guszejnov D., Hopkins P. F., 2015, *MNRAS*, 450, 4137
 Guszejnov D., Hopkins P. F., Krumholz M. R., 2017, *MNRAS*, 468, 4093
 Hennebelle P., 2012, *A&A*, 545, A147
 Hennebelle P., Chabrier G., 2008, *ApJ*, 684, 395
 Hopkins P. F., 2013, *MNRAS*, 433, 170
 Houlahan P., Scalo J., 1992, *ApJ*, 393, 172
 Keown J. et al., 2017, *ApJ*, 850, 3
 Kirk H. et al., 2017, *ApJ*, 846, 144
 Klein R. I., 1999, *J. Comput. Appl. Math.*, 109, 123
 Koch E. W., Ward C. G., Offner S., Loeppky J. L., Rosolowsky E. W., 2017, *MNRAS*, 471, 1506
 Koch E. W., Rosolowsky E. W., Boyden R. D., Burkhardt B., Ginsburg A., Loeppky J. L., Offner S. S. R., 2019, *AJ*, 158, 1
 Könyves V. et al., 2015, *A&A*, 584, A91
 Kratter K. M., Murray-Clay R. A., Youdin A. N., 2010, *ApJ*, 710, 1375
 Krumholz M. R., McKee C. F., Klein R. I., 2004, *ApJ*, 611, 399

- Krumholz M. R., Klein R. I., McKee C. F., Bolstad J., 2007, *ApJ*, 667, 626
- Lee A. T., Cunningham A. J., McKee C. F., Klein R. I., 2014a, *ApJ*, 783, 50
- Lee K. I. et al., 2014b, *ApJ*, 797, 76
- Lee Y.-N., Hennebelle P., Chabrier G., 2017, *ApJ*, 847, 114
- Li P. S., Norman M. L., Mac Low M.-M., Heitsch F., 2004, *ApJ*, 605, 800
- Li P. S., Martin D. F., Klein R. I., McKee C. F., 2012, *ApJ*, 745, 139
- Li C., Wang H.-c., Wu Y.-w., Ma Y.-h., Lin L.-h., 2020, *RA&A*, 20, 031
- Mac Low M.-M., 1999, *ApJ*, 524, 169
- Mairs S., Johnstone D., Offner S. S. R., Schnee S., 2014, *ApJ*, 783, 60
- Mao S. A., Ostriker E. C., Kim C.-G., 2020, *ApJ*, 898, 52
- Massi F. et al., 2019, *A&A*, 628, A110
- Masunaga H., Miyama S. M., Inutsuka S.-I., 1998, *ApJ*, 495, 346
- Molinari S., Schisano E., Faustini F., Pestalozzi M., di Giorgio A. M., Liu S., 2011, *A&A*, 530, A133
- Mouschovias T. C., Spitzer L., 1976, *ApJ*, 210, 326
- Nayak O., Meixner M., Fukui Y., Tachihara K., Onishi T., Saigo K., Tokuda K., Harada R., 2018, *ApJ*, 854, 154
- Ntormousi E., Hennebelle P., 2019, *A&A*, 625, A82
- Offner S. S. R., Arce H. G., 2015, *ApJ*, 811, 146
- Offner S. S. R., Chaban J., 2017, *ApJ*, 847, 104
- Offner S. S. R., Liu Y., 2018, *Nat. Astron.*, 2, 896
- Offner S. S. R., Klein R. I., McKee C. F., 2008, *ApJ*, 686, 1174
- Offner S. S. R., Klein R. I., McKee C. F., Krumholz M. R., 2009, *ApJ*, 703, 131
- Offner S. S. R., Kratter K. M., Matzner C. D., Krumholz M. R., Klein R. I., 2010, *ApJ*, 725, 1485
- Offner S. S. R., Capodilupo J., Schnee S., Goodman A. A., 2012, *MNRAS*, 420, L53
- Offner S. S. R., Clark P. C., Hennebelle P., Bastian N., Moraux E., Whitworth A. P., 2014, preprint (arXiv:1312.5326)
- Ostriker E. C., Stone J. M., Gammie C. F., 2001, *ApJ*, 546, 980
- Padoan P., Nordlund Å., 2002, *ApJ*, 576, 870
- Padoan P., Haugbølle T., Nordlund t., 2012, *ApJ*, 759, L27
- Padoan P., Pan L., Juvela M., Haugbølle T., Nordlund Å., 2019, preprint (arXiv:1911.04465)
- Rosolowsky E. W., Pineda J. E., Kauffmann J., Goodman A. A., 2008, *ApJ*, 679, 1338
- Seo Y. M. et al., 2015, *ApJ*, 805, 185
- Shetty R., Collins D. C., Kauffmann J., Goodman A. A., Rosolowsky E. W., Norman M. L., 2010, *ApJ*, 712, 1049
- Smith R. J., Clark P. C., Bonnell I. A., 2009, *MNRAS*, 396, 830
- Smith R. J., Glover S. C. O., Klessen R. S., 2014, *MNRAS*, 445, 2900
- Sokol A. D. et al., 2019, *MNRAS*, 483, 407
- Storm S. et al., 2016, *ApJ*, 830, 127
- Stutzki J., Guesten R., 1990, *ApJ*, 356, 513
- Truelove J. K., Klein R. I., McKee C. F., Holliman J. H. II, Howell L. H., Greenough J. A., 1997, *ApJ*, 489, L179
- Turk M. J., Smith B. D., Oishi J. S., Skory S., Skillman S. W., Abel T., Norman M. L., 2010, *ApJS*, 192, 9
- Williams J. P., de Geus E. J., Blitz L., 1994, *ApJ*, 428, 693
- Wong T. et al., 2017, *ApJ*, 850, 139

APPENDIX A: COMPUTED PROPERTIES

We use typical quantities common in star formation studies. However, for transparency, we define their algorithmic definitions used in this paper. Volumes have been calculated using the volume of the leaf on the uniform grid. Quantities are computed on the cell-wise level using the AMR cells identified within the leaf and then summed to a single quantity where indicated.

The centre of mass, which is repeatedly used, is defined as $\mu = \sum m_{\text{gas}} v_{\text{gas}} / \sum m_{\text{gas}}$. Then, iterating over all sinks in a leaf, it is modified as $\mu = (\mu m + v_{\text{sink}} m_{\text{sink}}) / (m + m_{\text{sink}})$.

- (i) Mean density: $\sum m_{\text{cell}} \rho_{\text{cell}} / \sum m_{\text{cell}}$
- (ii) Total mass: $M_{\text{gas}} + M_{\text{sink}}$

(iii) Gas mass: M_{gas}

(iv) Volume: $N_{\text{uniform cells}} \times V_{\text{uniform cells}}$

(v) Size:

$$(a) \text{ Size} = \sqrt{(x_{\text{max}} - x_{\text{min}})^2 + (y_{\text{max}} - y_{\text{min}})^2 + (z_{\text{max}} - z_{\text{min}})^2}$$

(vi) Oblateness:

$$(a) \Delta x = (x_{\text{max}} - x_{\text{min}})$$

$$(b) \text{mag} = \sqrt{\Delta x^2 + \Delta y^2 + \Delta z^2}$$

$$(c) \text{Oblateness} = (\max[\Delta x, \Delta y, \Delta z] - \min[\Delta x, \Delta y, \Delta z]) / \text{mag}$$

(vii) Virial parameter:

$$(a) \sigma = \sqrt{(\sum m((v_x - \mu_x)^2 + (v_y - \mu_y)^2 + (v_z - \mu_z)^2)) / \sum m}$$

$$(b) R = (\sum m / \rho)^{1/3}$$

$$(c) \alpha = 5(\sigma / \sqrt{3})^2 R / 3GM_{\text{tot}}$$

(viii) Internal Mach number:

$$(a) c_s = \sqrt{\gamma P / \rho}$$

$$(b) \mathcal{M} = \sqrt{(\sum m(((v_x - \mu_x)^2 + (v_y - \mu_y)^2 + (v_z - \mu_z)^2) / c_s^2)) / \sum m}$$

(ix) Core Mach number:

$$(a) c_s = \sqrt{\gamma P / \rho}$$

$$(b) \mathcal{M} = \sum m(\sqrt{v_x^2 + v_y^2 + v_z^2} / c_s) / \sum m$$

(x) Alfvén Mach number:

$$(a) v_A = \sqrt{B^2 / 4\pi\rho}$$

$$(b) \mathcal{M}_A = \sqrt{(\sum m(((v_x - \mu_x)^2 + (v_y - \mu_y)^2 + (v_z - \mu_z)^2) / v_A^2)) / \sum m}$$

(xi) Pressure: $\sum m_{\text{cell}} P_{\text{cell}} / \sum m_{\text{cell}}$

(xii) Angular momentum magnitude:

$$(a) r_{\text{cor}} = r - \mu_r$$

$$(b) v_{\text{cor}} = v - \mu_v$$

$$(c) j = \sum (m v_{\text{cor}} \times r_{\text{cor}}) / \sum m$$

$$(d) \text{Magnitude} = \sqrt{j_x^2 + j_y^2 + j_z^2}$$

(xiii) Angular momentum orientation:

$$(a) r_{\text{cor}} = r - \mu_r$$

$$(b) v_{\text{cor}} = v - \mu_v$$

$$(c) j = \sum (m v_{\text{cor}} \times r_{\text{cor}}) / \sum m$$

$$(d) \text{mag} = \sqrt{j_x^2 + j_y^2 + j_z^2}$$

$$(e) \text{Orientation} = (\max(j) - \min(j)) / \text{mag}$$

(xiv) B-field magnitude:

$$(a) B = \left[\frac{\sum m B_x}{\sum m}, \frac{\sum m B_y}{\sum m}, \frac{\sum m B_z}{\sum m} \right]$$

$$(b) \text{Magnitude} = \sqrt{B_x^2 + B_y^2 + B_z^2}$$

(xv) B-field orientation:

$$(a) B = \left[\frac{\sum m B_x}{\sum m}, \frac{\sum m B_y}{\sum m}, \frac{\sum m B_z}{\sum m} \right]$$

$$(b) \text{mag} = \sqrt{B_x^2 + B_y^2 + B_z^2}$$

$$(c) \text{Orientation} = (\max(B) - \min(B)) / \text{mag}$$

(xvi) Plasma β

$$(a) \beta_{\text{cell}} = 8\pi P / B^2$$

$$(b) \beta = \sum m_{\text{cell}} \beta_{\text{cell}} / \sum m_{\text{cell}}$$

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