MULTI-SCALE COHERENT STRUCTURE EXTRACTION AND VISUALIZATION FOR FLOW ANALYSIS

by
Duong B. Nguyen

A dissertation submitted to the Department of Computer Science,

College of Natural Sciences and Mathematics

in partial fulfillment of the requirements for the degree of

Doctor of Philosophy
in Computer Science

Chair of Committee: Dr. Guoning Chen

Committee Member: Dr. Edgar Gabriel

Committee Member: Dr. Gopal Pandurangan

Committee Member: Dr. Di Yang

University of Houston
December 2020

ACKNOWLEDGMENTS

I would like to express my sincere gratitude to my Ph.D. advisor, Dr. Guoning Chen, for his incredible patience, guidance, kindness, support and instructions over the last four years. His knowledge, insights, and generous help gave me high confidence and courage in my efforts to undertake the challenging research projects. He always appears when I need the help, and without his support, this work would not have been possible.

I am very grateful to Dr. Edgar Gabriel, Dr. Gopal Pandurangan, and Dr. Di Yang for serving on my Ph.D. committee, providing valuable inputs and advice on my dissertation for which I am truly grateful. They have made great contributions in helping me to get through the past Ph.D. evaluations and to think critically about my research projects.

I also want to thank my fellows in DaViM Lab, including Dr. Kaoji Xu, Dr. Lieyu Shi, Dr. Marzieh Berenjkoub, Muhammad Naeem Akram, Kishansingh Rajput, Nguyen Phan, and the past lab member, Dr. Lei Zhang. It was a great privilege to work with them for the past years, and we will keep our friendship forever.

I would especially like to thank my love and family. My wife, Thao Vu, has been very supportive and encouraging throughout my study. My parents have made countless sacrifices and unconditional support to help me get to this point. They have always believed in me and inspired me to reach full potential. I would like to dedicate the dissertation to them.

ABSTRACT

Coherent structures are important features in fluid flows. A better understanding of the physics of coherent structures will help explain a diverse range of physical phenomena and help improve our capability of modeling complex turbulence flows, such as those often seen in combustion, chemical reaction, and heat transfer. However, due to their multi-scale nature and non-unified characterizations, extraction and separation of coherent structures remain a challenging task. This is further complicated by the overly complicated visual representation of these structures, significantly reducing the efficiency of the domain expert workflows for discovering the flow physics when they need to spend a considerable amount of time and effort to read the complex charts/graphs/geometries. In addition, the physical behaviors of flow that experts care about are not reliably conveyed in the visualizations due to the predominant focus on the geometric characteristics of the flow data.

To address the above challenges and support domain experts to analyze various flow behaviors, especially coherent structures in the flow, this work proposes (1) a time activity curve (TAC) based method to encode relevant physics into the geometric representation, (2) a novel pipeline to extract and visualize multi-scale coherent structures for the instantaneous (or time-independent) turbulent Taylor-Couette flow (TCF), and (3) a new framework to extract and visualize large-scale coherent structures in time-dependent shear flows using dynamic mode decomposition (DMD). The TAC-based framework enables us to select pathlines that can effectively represent the physical characteristics of interest and their temporal behavior in the fluid flow, which can be used to study the temporal behaviors of vortices, including their formation, merging, and breakdown. The novel visualization framework for TCFs enables the separation of the large-scale and small-scale coherent structures in the instantaneous TCFs for the first time in 3D, which supports the study of the formation of Taylor rolls – a 3D coherent structures in TCFs, with different simulation parameters. The new DMD based framework for shear flows effectively extracts and visualizes the large-scale coherent structures over time, opening a new door for multi-scale coherent structure extraction and tracking over time for turbulence flow study.

TABLE OF CONTENTS

LIST OF TABLES

LIST OF FIGURES

1 Introduction

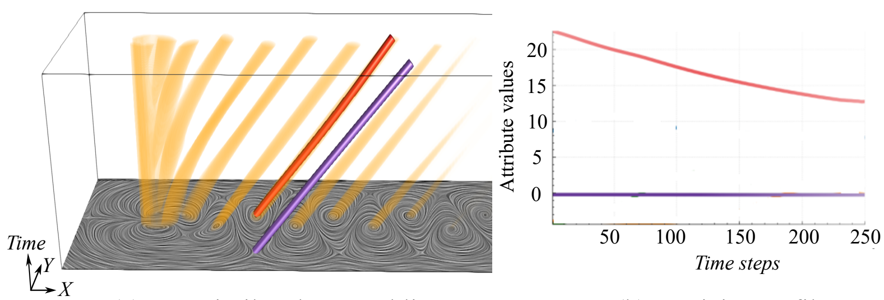
Fluid flows including air and fluids are ubiquitous in nature. Understanding fluid flows is a critical task for the success of many different fields in science and industry, such as automobile and aircraft engineering, climate study, combustion dynamics, earthquake engineering, and medicine. In these fields, researchers seek to unlock the underlying physics laws and try to model the turbulent behavior of fluid flow. Flow visualization comes in as an important tool making these invisible physical processes visible. It employs a wide range of computer graphics, data analysis, and high-performance computation techniques to help domain experts intuitively see and interpret the flow data.

The nature of flow is a turbulent process where energy and other quantities are transferred across scales. The detailed mechanics enabling this transfer are still not well-understood. To adequately separate the time- and length-scales, coherent structures with different geometric and physical characteristics need to be identified. Large-scale coherent structures are often responsible for the lion's share of the transport of mass, heat, and momentum in flows, while small-scale structures are often responsible for energy dissipation. A better understanding of the physics of coherent structures will help explain a diverse range of physical phenomena and help improve our capability of modeling complex turbulence flows, such as those often seen in combustion, chemical reaction and heat transfer.

1.1 Challenges

Challenge 1. Many effective approaches [14, 50, 74, 89] have been developed by the visualization community to assist experts understand coherent structures and flow dynamics. For instance, the geometric-based approaches [59, 14] are commonly applied thanks to their intuitive representation of flow behavior. Examples of geometric-based techniques include various integral curve/surface based representations and integral curve clustering that concentrate on the geometric characteristics of the flow (e.g., the shape or curvature of integral curves).

However, there are two limitations with the existing geometric-based methods. First, important physically relevant features are not always captured. For example, vector field topology, an abstract representation of the geometric characteristics of steady flow, only encodes hyperbolic features in the flow. Other physically relevant information, e.g., vortices, shearing, etc., is not always captured. Similarly, clustering methods that select integral curves to ensure sufficient spatial coverage and to reduce cluttering typically do not consider physical importance. Second, a geometric representation may not intuitively reveal the physical behavior of the flow. In an example shown in Figure 1, the two pathlines (i.e., the red and blue curves) in (a) have similar geometric characteristics, but their corresponding vorticity attribute values measured along them have rather different behavior over time (b), i.e., the vorticity values of the red pathline are much larger than those along the blue pathline. This example demonstrates that the geometric representation of flows is not always reliable for interpreting the underlying physics of the flow.



(a) Two similar shape pathlines (b) Vorticity profiles

Figure 1: Geometric representation may not intuitively reveal the physical behavior of the flow. (a) Two pathlines with similar geometry derived from the flow behind a cylinder and the vorticity field shown by the volume rendering. (b) The corresponding profiles of the vorticity measured along the two pathlines show different physical behaviors.

Challenge 2. The coherent structures exist at all scales in turbulent flows. Due to the multi-scale nature, multi-scale processing is required to disentangle and visualize the structures. However, there are no clear physical boundaries between scales, making the separation of large-scale structures from the small ones extremely difficult. In fluid dynamics, moving ensemble averages [67], two-point correlations [94], and Fourier analysis [66] have been commonly used to detect and analyze coherent structures. The fluid dynamic researchers also employ visualization techniques such as volume rendering and iso-surfacing. The main disadvantage of these methods is that they often include the relatively arbitrary choice of a cut-off to isolate the structures, be it in the shape of a cut-off frequency in Fourier space [66], or in the shape of a threshold value for a certain attribute [1]. Furthermore, these methods are usually tuned for one particular setup of the system; but as the control parameters of the flow change, the shape of the coherent structure also changes. As a result, conventional methods can completely fail to track the structures across certain dimensions of parameter space. Finding an efficient method which can separate coherent structures across scales still remains a challenge.

1.2 My Contributions

To address the above challenges, this dissertation work proposes (1) a time activity curve (TAC) based method to encode relevant physics into the geometric representation, (2) a novel pipeline to extract and visualize multi-scale coherent structures for the instantaneous (or time-independent) turbulent Taylor-Couette flow (TCF), and (3) a new framework to extract and visualize large-scale coherent structures in time-dependent shear flows using dynamic mode decomposition (DMD).

1.2.1 Physics-based Pathline Selection and Exploration

To encode physics into the geometric representation of flow (**Challenge 1**), Zhang et al. [128] [1] introduced a Lagrangian accumulation framework that assigns each integral curve a value by accumulating the values of an attribute of interest along the curve. This value encodes the overall (or average) physical attributes along the integral curve that can be used to characterize the physical relevance of this integral curve. In the meantime, Lee et al. proposed a visualization framework to analyze time-varying data sets with a time activity curve (TAC) based distance field [52], which is used to highlight features. All these previous works look at the overall characteristics (or the

[1] I also contributed to this work.

sum/average) of local attributes, which may not capture all defining characteristics, e.g., the diffusion behavior of Q concentration [39] during the advection of vortices over time (Figure 14), due to the suppression of local information.

To address the above limitations and incorporate more detailed physics into the analysis and visualization of unsteady flow, a novel visual analysis framework is proposed based on the temporal behavior of local, physical attributes of interest measured along individual pathlines. Similar to Lee et al. [52], the temporal profile of the selected attribute along a pathline is referred to as a *time activity curve* (TAC).

The benefits of analyzing the flow behavior based on TACs are two-fold. First, they are 1D plots that are independent of the flow dimension (i.e., applicable to both 2D and 3D data); hence processing them is easier. Second, the geometric characteristics of TACs (e.g., ascending, descending, peaks, valleys, etc.) reveal the interaction of flow particles with physical features over time. This enables us to explain the geometric characteristics of the corresponding pathlines and vice versa. In addition, the attribute profile provides additional information that the geometry of pathlines cannot convey, such as the decaying of the Q concentration over time, indicating the loss of the rotation momentum of vortices during transport. Such an in-depth coupling of pathline characteristics and physical attributes has not been studied previously.

Based on TACs and the benefits of their analysis, the pathline characterization can be coupled with the attribute profiles measured along the pathlines to provide a more physics-aware exploration of unsteady flow. Different from previous works, all the sampled attribute values along each pathline are considered during their characterization. A number of analysis and exploration techniques based on TACs are introduced, including a new spatio-temporal, hierarchical clustering of pathlines based on their respective TACs and a TAC-based pathline selection and exploration. A visual exploration system is developed that integrates the aforementioned TAC-based analysis and exploration techniques with a number of novel visualizations to support an effective user exploration of the pathline behaviors based on their respective TACs, including a modified edge-bundling visualization of TAC clusters and 2D stack plot for TAC behavior summarization and exploration.

The TAC-based method can reveal multi-scale vortex structures in many flow simulations with relatively simple configurations. It is impractical to apply the framework to the turbulent flows due to the performance constraint of the AHC clustering. It also does not separate structures in different scales.

1.2.2 Multi-scale Coherent Structures Extraction for Taylor-Couette Turbulence

To enable the separation of structures with different scales in turbulent flows (**Challenge 2**), my second contribution focuses on the analysis of the Taylor-Couette flow (TCF). Taylor-Couette flow (TCF) is the fluid motion between two coaxial, independently rotating cylinders. TCF has become an important model system in fluid dynamics, because it helps to understand the development of hydrodynamic stabilities and pattern formation. TCF is linearly unstable when angular momentum decreases with radius, and different configurations of inner and outer cylinder rotation lead to diverse dynamics. At low Reynolds numbers, TCF is dominated by a stacking of large-scale structures called Taylor vortices, which arise due to centrifugal instabilities and fill the entire gap between the cylinders. As the Reynolds number increases, and the flow becomes turbulent, structures with increasingly smaller length- and time-scales appear. The large-scale vortices remain, and are relatively stationary in time (e.g., Figure 2(b)).

In addition to the theory that large structures depend on the Reynolds numbers, the recent statistic-based method [37] indicates that the turbulent rolls may re-appear with different cylinder rotation ratios R_Ω, as also shown in our later results. Nonetheless, the large-scale structures (e.g., Taylor vortices or other coherent structures) modulate the appearance of small-scale structures, and can act as transport barriers; thus, they are particularly interesting to the domain experts.

As illustrated in Figure 3, the traditional thresholding technique based on a selected attribute (e.g., vorticity or Q-criterion) can output occluded, undesired structures with different parameter settings, making the identification and tracking of the desired structures difficult. Making this worse is that attributes such as vorticity and Q-criterion in small structures have different value ranges from those in large structures as reported previously [110]. Often, a high positive Q-criterion

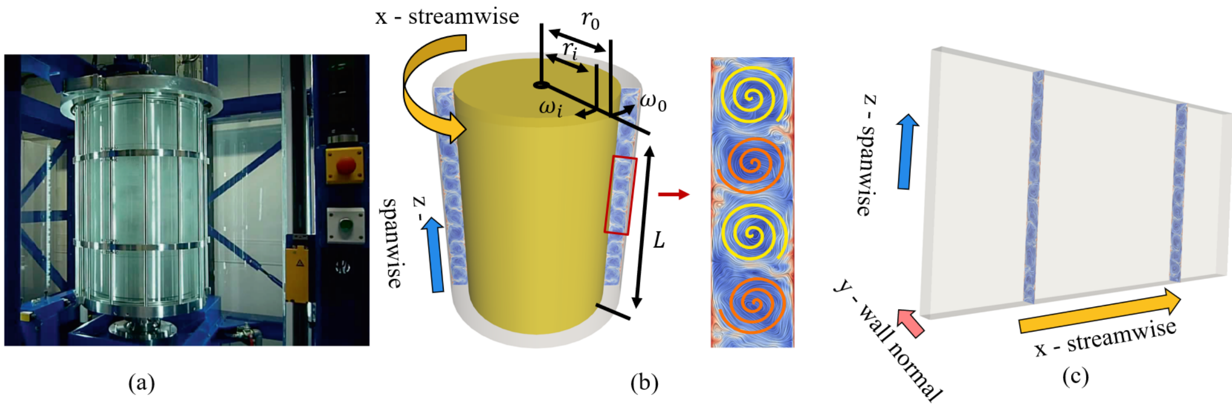


Figure 2: (a) Twente Turbulent Taylor-Couette (T3C) experimental facility [65]. (b) The Taylor-Couette system consists of two coaxial cylinders, which have an inner cylinder with radius of r_i and an outer cylinder with radius of r_o. Both cylinders are of height L. The inner cylinder rotates with an angular velocity ω_i and the outer cylinder rotates with an angular velocity of ω_o. (c) Domain transformation from the cylindrical to the Cartesian coordinates.

corresponds to vortices with small-scale structures due to its definition, i.e., it is based on vorticity and strain which are inherently local and small-scale quantities, while small Q values may exist both inside and outside the large structures. Furthermore, as shown later (Figure26), one single attribute with one single threshold value is not sufficient to clearly separate multi-scale coherent structures, especially when the cylinder rotation ratio changes. All these challenges force the fluid mechanic researchers to resort to the visualization of 2D cross section (Figure 2 (b)) through cylinders for TCF analysis, losing important 3D configuration information of the structures. To the best of our knowledge, no existing works provide an effective 3D visualization of TCF that separates the large- and small-scale structures to aid the study of the re-appearance of certain coherent structures (especially, Taylor vortices) in TCF under different rotation ratios.

To address this challenge, a first 3D visualization framework is proposed to support fluid mechanic experts in the analysis and exploration of the large- and small- scale structures in Taylor-Couette flows.

To overcome the limitation of the separation of structures in different scales with a single attribute, our framework employs multiple attributes for feature extraction and separation. By

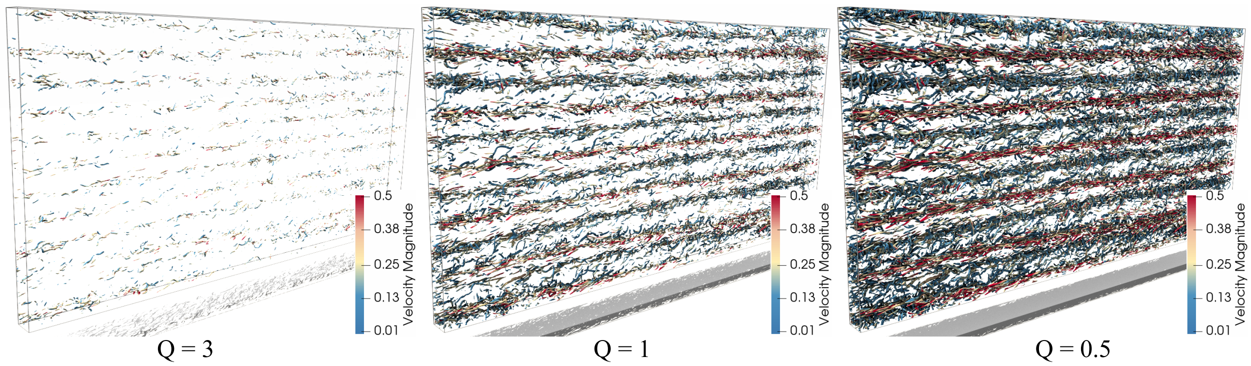


Figure 3: Traditional issues with Q threshold selection. Domain experts rely on the Q-criterion [38] to extract the vortices and suffer from the issues with the occlusion, because many different scale vortices overlap each others. Note that while visually these dense layout of vortices form groups that may indicate the formation of Taylor rolls/vortices, they do not align with the location of Taylor vortices as indicated by the velocity field as illustrated in Figure 2 (b).

extending the feature level-set method [40] to take into account values in a range, my novel framework can combine the characteristics of multiple attributes in one analysis to increase the difference between the small- and large- scale structures, making their separation easier and more accurate. To extract a 3D surface representation for the coherent structures, iso-surfaces from the kernel density estimate of the distance field obtained from the feature level-set are constructed. Considering the characteristic that TCF with certain rotation ratio may be depicted by its 2D cross sections, referred to as the **WS planes** formed by the wall-normal and spanwise directions (Figure 2 (a)(c)), a 2D summary configuration of vortices in TCF is computed by projecting the 3D information onto the **WS** plane. Compared to the visualizations generated with the conventional approaches in both 2D and 3D, the presented framework produces a much cleaner visualization when applied to TCFs simulated with different parameters.

One limitation of the above framework is that it heavily depends on the parameter selection (e.g., the KDE kernel size, the number of attributes, and the feature level-set threshold), reducing its application to other types of shear flows. Also, it does not support the analysis and visualization of coherent structures in time-dependent flows. This leads to the third contribution of this dissertation which attempts to overcome the parameter dependent issue and works for unsteady flows.

1.2.3 Dynamic Mode Decomposition for Large-Scale Coherent Structure Extraction in Shear Flows

To address the limitation of the above framework for TCF visualization and to separate large- and small-scale structures in general shear flows (**Challenge 2**), a DMD-based analysis and visualization framework is introduced.

Dynamic mode decomposition (DMD) is a data-driven and parameter-free method that provides a spatio-temporal decomposition of data into a set of relevant dynamical modes called DMD modes from a sequence of snapshots of an evolving system. It has been initially introduced by Schmidt [92] to provide the best *linear approximation* that sends the data from its current state to the next in a non-linear dynamics system. DMD computation relies on the parameter-free Singular Value Decomposition (SVD) method. Each DMD mode is considered as a spatial structure which is accompanied by time dynamics. The corresponding time dynamics of DMD modes can be characterized by their speeds (e.g., how fast or slow they move), making DMD a promising candidate to extract coherent structures (evolving with different speeds) in flow. Several works [129] have attempted to apply DMD to different kinds of flows. For example, Gilka et al. [28] performed a DMD analysis on the flow behind an actuated bluff body, and Schmid [92] tried DMD with the Gurney flap wake flow to capture the vortex shedding pattern. However, the capabilities of DMD in analyzing shear flows have not been fully investigated. In addition, DMD has not received much attention from the visualization community for the task of spatial and temporal feature extraction; hence there is no existing visualization system specialized for DMD, reducing its analysis ability.

To address this challenge, the relation between the large-scale structures in shear flows and the slowest DMD mode is first shown, which enables the extraction and visualization of the large-scale structures using the slowest DMD mode. To address the issue of the standard DMD, multi-resolution DMD (mrDMD) [48] is used. To enable mrDMD to be applied to large-scale 3D unsteady vector fields, a GPU implementation is introduced. The analysis and visualization based on the fast mrDMD is applied to a number of 2D and 3D Plane Couette (PC) and the Waleffe flows.

This is the first time mrDMD is applied to 3D Plane Couette and Waleffe flow snapshots for large-scale structure extraction. The results obtained with the DMD-based analysis are compared with those obtained using convolution kernel based smoothing, time average, and the proper orthogonal decomposition (POD) to demonstrate its advantages. To further evaluate the applicability of the proposed DMD-based analysis and visualization, it is applied to other types of flows than shear flows. The experiments show that only using the slowest DMD mode is not sufficient in revealing the essential behavior of the flow. Instead, modes selected from different levels of mrDMD can better capture the dynamics of those non-shear flows. These results can be used as an empirical guideline for the visualization community on the proper use of DMD, especially mrDMD, for turbulence flow analysis and visualization.

1.3 Structure of the Dissertation

The rest of the dissertation is organized as follows: Chapter 2 introduces background knowledge of vector field, shear flows including Taylor-Couette and Waleffe turbulence. Simulated physical attributes in the flow analysis are also presented in Chapter 2. In Chapter 3, we discuss the existing works for the geometry-based visualization as well as the state-of-the-art methods for coherent structure extraction. The physics-based geometric representation framework is described in Chapter 4. Chapter 5 proposes two methods to separate coherent structures with different scales for Taylor Couette flows. Chapter 6 presents the large-scale coherent structure identification framework for shear flows based on the recently popular dynamic mode decomposition (DMD). Chapter 7 concludes this dissertation, lists the limitations of the proposed methods and provides discussion on the possible future directions of this work.

2 Background

This chapter introduces the concepts of vector fields and other simulated physical attributes that are used in the flow analysis. The numerical simulation details of Taylor-Couette flow are also discussed. In addition, some related methods, like dynamic mode decomposition, for the analysis and visualization of flow data, are reviewed.

2.1 Vector Fields

Consider a spatio-temporal domain $\mathbb{D} = \mathbb{M} \times \mathbb{T}$ where $\mathbb{M} \subset \mathbb{R}^d$ is a d-manifold ($d = 2, 3$) and $\mathbb{T} \subset \mathbb{R}$, a general vector field can be expressed as an ordinary differential equation (ODE) $\dot{\mathbf{x}} = V(\mathbf{x}, t)$. For an *unsteady* (or *time-dependent*) vector field $V(\mathbf{x}, t)$, the trajectory of a particle starting at $\mathbf{x}_0$ and at time t_0 is called a *pathline*, denoted by $\mathbf{x}_{\mathbf{x}_0,t_0}(t) = \mathbf{x}_0 + \int_0^t V(\mathbf{x}_{\mathbf{x}_0,t_0}(\tau), t_0 + \tau)d\tau$.

2.1.1 Local Attributes in Vector Field

There are a number of local attributes that are of interest to domain experts. Given a steady vector field $\mathbf{v}$, its spatial gradient $\nabla_{\mathbf{x}}\mathbf{v}$ is referred to as its *Jacobian*, denoted by $\mathbf{J}$. $\mathbf{J}$ can be decomposed as $\mathbf{J} = \mathbf{S} + \mathbf{R}$, where $\mathbf{S} = \frac{1}{2}[\mathbf{J} + (\mathbf{J})^\top]$ and $\mathbf{R} = \frac{1}{2}[\mathbf{J} - (\mathbf{J})^\top]$ are the symmetric and antisymmetric components of $\mathbf{J}$, respectively. A number of flow attributes can be derived from $\mathbf{v}$, $\mathbf{J}$, $\mathbf{S}$ and $\mathbf{R}$ [73]. In the examples shown in the dissertation, the following local, $\mathbf{A}_l$, are used for experimentation.

- $\mathbf{A}_1$: vorticity magnitude, $||\nabla \times \mathbf{v}||$.
- $\mathbf{A}_2$: λ_2, computed as the second largest eigenvalue of the tensor $\mathbf{S}^2 + \mathbf{R}^2$ [41].
- $\mathbf{A}_3$: $Q = \frac{1}{2}(\|\mathbf{R}\|^2 - \|\mathbf{S}\|^2)$. [38]
- $\mathbf{A}_4$: local shear rate, defined as the Frobenius norm of $\mathbf{S}$.
- $\mathbf{A}_5$: norm of $\mathbf{J}$, $\sqrt{\sum_{ij} J_{ij}^2}$. [32]

The attributes selected in this work are widely used in fluid mechanics to characterize different physical properties of the flow. For example, Q, λ_2 and vorticity are the common attributes for vortex characterization. Local shear rate and the norm of Jacobian are used to study the local divergence and separation behavior in the flow. Note that although not included in the above list, other attributes provided in a given simulation (e.g., acceleration, kinetic energy, pressure and dye) can also be used with our framework. The correlation and dependency of these attributes has been studied in [3]. One important aspect about the derived attributes is that they change when the reference frames change. As a result, TACs are variant under the movement of reference frames. To make the TACs independent from the reference frame transformation, the hyper-objective measure from Günther et al. [30] is applied to find the optimal local reference frame in which the flow appears near-steady. Under this reference frame, the velocity and Jacobian become objective which means they are invariant under any smooth affine transformations (rotation, translation and uniform scale). The attributes computed in the near-steady field are also affine invariant. For the complete explanation about the changes of reference frames, please refer to [30].

2.1.2 Time Activity Curve (TAC)

In this section, how TACs are computed based on a given local attribute is described. The concept of split points is also introduced, which are used to segment a TAC into multiple intervals.

Definition of TAC. Given a local attribute $\mathbf{A}$, a Lagrangian TAC along a pathline $\mathcal{C}$ of a particle, seeded at $\mathbf{x}$ at time t, can be expressed as:

$$\Gamma_{\mathbf{A},\mathcal{C}}[i] = \mathbf{A}(\mathcal{C}(\mathbf{x}, t_i), t_i) | i = 1, 2 \ldots n \tag{1}$$

where $t_1, t_2 \ldots t_n$ are the sample times within the time window $T \subset \mathbb{T}$ and $\mathcal{C}(\mathbf{x}, t_i)$ is the location of $\mathbf{x}$ on the pathline $\mathcal{C}$ at time t_i. Similarly, Eulerian TACs of attributes can be measured at a fixed location over time. For the rest of the discussion, the focus is on Lagrangian TACs unless specified otherwise.

To simplify the notation, a Time Activity Curve is denoted as $\Gamma = \{\Gamma[i] | i = 1, 2 \ldots n\}$ where $\Gamma[i]$ is the local attribute value at time t_i. Figure 4(a) illustrates a TAC where the x axis indicates time indexes and the y axis shows the local attribute values at the corresponding times. The length

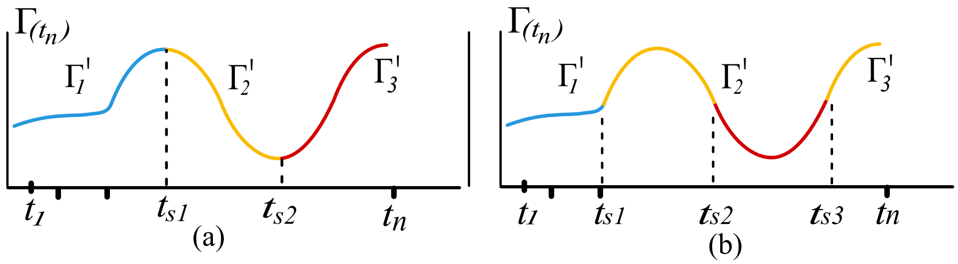


Figure 4: (a) An example TAC with three events. t_{s1} and t_{s2} are two split points defined at the extrema. (b) A similar TAC as in (a), but the split points are defined at the maxima and minima of the derivative curve

of a TAC is the number of values in the TAC, indicating the lifespan of the corresponding particle.

Split points. Since TACs are time series data, it is needed to analyze their behaviors in different time intervals. To facilitate the temporal characterization, a TAC is split into multiple sub-TACs by using either the extrema or inflection points as shown in Figure 4(a)(b). These split points are later used in the temporal clustering to segment all Γ in identical time intervals (Section 4.1.4).

2.2 Shear Flows

Shear flows are fluid flows which are driven by a velocity difference. The two types of wall-bounded shear flows that this dissertation work is focused on, i.e., Taylor Couette flows (TCF) and Waleffe flows, show large-scale structures which behave in an inviscid manner, and are hard to extract. It is believed that these structure behave in a quasi-linear manner [57, 103], thus making them a target for DMD. The following provides a brief introduction of these flows and the large-scale structures that domain experts are interested in.

Taylor Couette flows (TCFs). In a simple setting, Taylor Couette can be described as the flow between two parallel plates separated by a distance d, which move with equal, but opposite velocities $\pm U/2$. A solid body rotation Ω can be added in the spanwise direction that is either anti-cyclonic, i.e., opposite to the shear, or cyclonic, i.e. in the direction of the shear. The two non-dimensional parameters that define the flow are the (shear) Reynolds number $Re = Ud/\nu$, where ν is the kinematic viscosity of the fluid, and the rotation number $R_\Omega = 2d\Omega/U$. A schematic of the flow is shown in Figure 5.

Let us fix $Re_s = 10^4$, and vary from $R_\Omega = 0$ to $R_\Omega = 0.1$, which generate a sufficiently diverse range of length-scales, while also featuring fixed Taylor vortices [80].

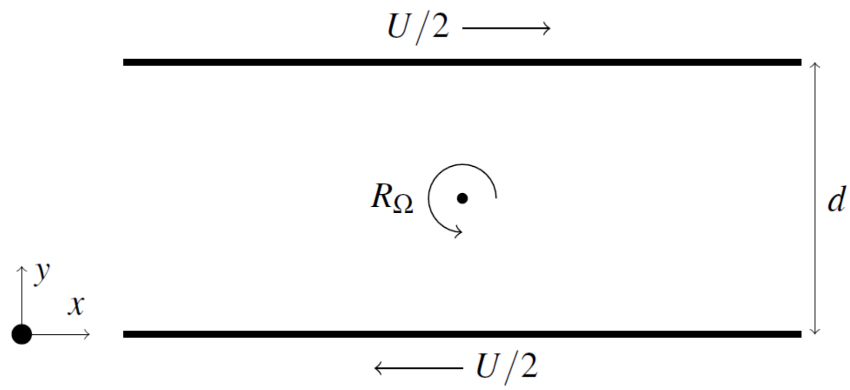


Figure 5: Schematic of the simulation domain. The third (spanwise) dimension z is omitted for clarity.

Depending on the solid-body spanwise rotation, large-scale structures can appear in the TC flow. For some values of R_Ω, statistical approaches [80] indicate that they can be pinned in space and regularized in the streamwise direction. As the flow becomes turbulent, structures of increasingly smaller of length- and time-scales appear, even while the large-scale vortices are relatively stationary in time. The large-scale structures modulate the appearance of small-scale structures, and can act as transport barriers. Detecting the large-scale structures in various parameter settings still remains an open problem that this dissertation work tries to address with the presented DMD framework.

Waleffe flows. Waleffe flow [20] can be thought of as a variation on plane Couette flow which substitutes the no-slip condition at both walls by a free-slip condition. The system then consists of two parallel plates d apart. The flow is forced by a sinusoidal body shear $F = \cos(\pi y/d)$, which gives a characteristic velocity $U = \sqrt{F/d}$. Waleffe flow has been used to study the effect of the near-wall cycle on the formation and pinning of large-scale structures in shear flow in the absence of a no-slip boundary condition. Pinned and streamwise invariant large-scale vortices were found in the anti-cyclonic regime at high Reynolds numbers [20]. These structures had a different vorticity distribution from those found for the analogous parameters in plane Couette flow, but showed similar temporal behavior when analyzing the separate energy components.

To perform the direct numerical simulations for both Taylor-Couette and Waleffe, the Navier-Stokes equation is solved using a second-order energy-conserving finite difference code [111]. The two wall-parallel directions are taken as periodic, with periodicity lengths $L_x = 25.12d$ and $L_z = 12.56d$ respectively. This means that the simulation covers a subset of the spatio-temporal domain, $[0, 25.12] \times [0, 1] \times [0, 12.56]$. The resolution of the simulation is $1024 \times 384 \times 1024$ (i.e., the number of grid points in the x,y and z dimension.).

2.3 Dynamic Mode Decomposition

Dynamic mode decomposition (DMD) is a method that provides a spatio-temporal decomposition of data into a set of relevant dynamical modes from a sequence of snapshots of an evolving system [92]. It is capable of *extracting flow structures which evolve linearly.* This characteristic makes it a promising candidate to extract the large-scale structures in wall-bounded turbulent flow, as it is known that these structures behave in a quasi-inviscid, quasi-linear manner [57, 103]. In this section, the general idea of the standard (or exact) DMD method is introduced first, followed by the description of its improved variant – multiresolution DMD (mrDMD).

Given an unsteady vector field with M time steps $x_1 ... x_M$ where $x_i \in C^N$ is a vector including the velocity field sampled at equal time intervals Δt. In other words, x_i can be generated by transforming the 2D or 3D flow domain into 1D vector structure in which N is the multiplication between the number of velocity components (i.e., 2 for 2D, 3 for 3D flows) and the domain resolution. Typically, the size of flow domain is substantially greater than the number of time steps, thus $N >> M$. The data can be arranged into two column-wise matrices X_1 and X_2 as follows:

$$X_1 = \begin{bmatrix} | & | & & | \\ x_1 & x_2 & ... & x_{M-1} \\ | & | & & | \end{bmatrix}, X_2 = \begin{bmatrix} | & | & & | \\ x_2 & x_3 & ... & x_M \\ | & | & & | \end{bmatrix} \tag{2}$$

DMD tries to find the dominant eigenvalues and eigenvectors of a best linear approximation that sends the data from its current state to the next state. The best-fit approximation can be simply expressed as:

$$x_{i+1} = Ax_i \tag{3}$$

or can be written in the matrix form and decomposed by using Singular Value Decomposition (SVD) as follows:

$$X_2 = AX_1 \text{ or } A = X_2X_1^{\dagger} = X_2V\Sigma^{-1}U^* \tag{4}$$

where A is a $N \times N$ matrix and $\dagger$ denotes the Moore-Penrose pseudo inverse. In practice, N is typically too large; hence, it is nearly impossible to store and compute matrix A directly. Fortunately, DMD circumvents this challenge by analyzing a smaller matrix $\tilde{A}$ obtained via projection onto the left singular vectors in U:

$$\tilde{A} = U^*AU = U^*X_2V\Sigma^{-1} \tag{5}$$

The matrix $\tilde{A}$ is of the size $N \times (M-1)$ which is substantially smaller than the original matrix A, but they have the same eigenvalues as proven by Tu et al. [108]. After applying the eigen-decomposition to $\tilde{A}$, the eigenvalues λ_i and eigenvectors w_i can be arranged in the matrices Λ and W such that:

$$\tilde{A}W = W\Lambda \tag{6}$$

The eigen-decomposition of A can be defined as:

$$A\Phi = \Phi\Lambda, \text{ where } \Phi = X_2V\Sigma^{-1}W \tag{7}$$

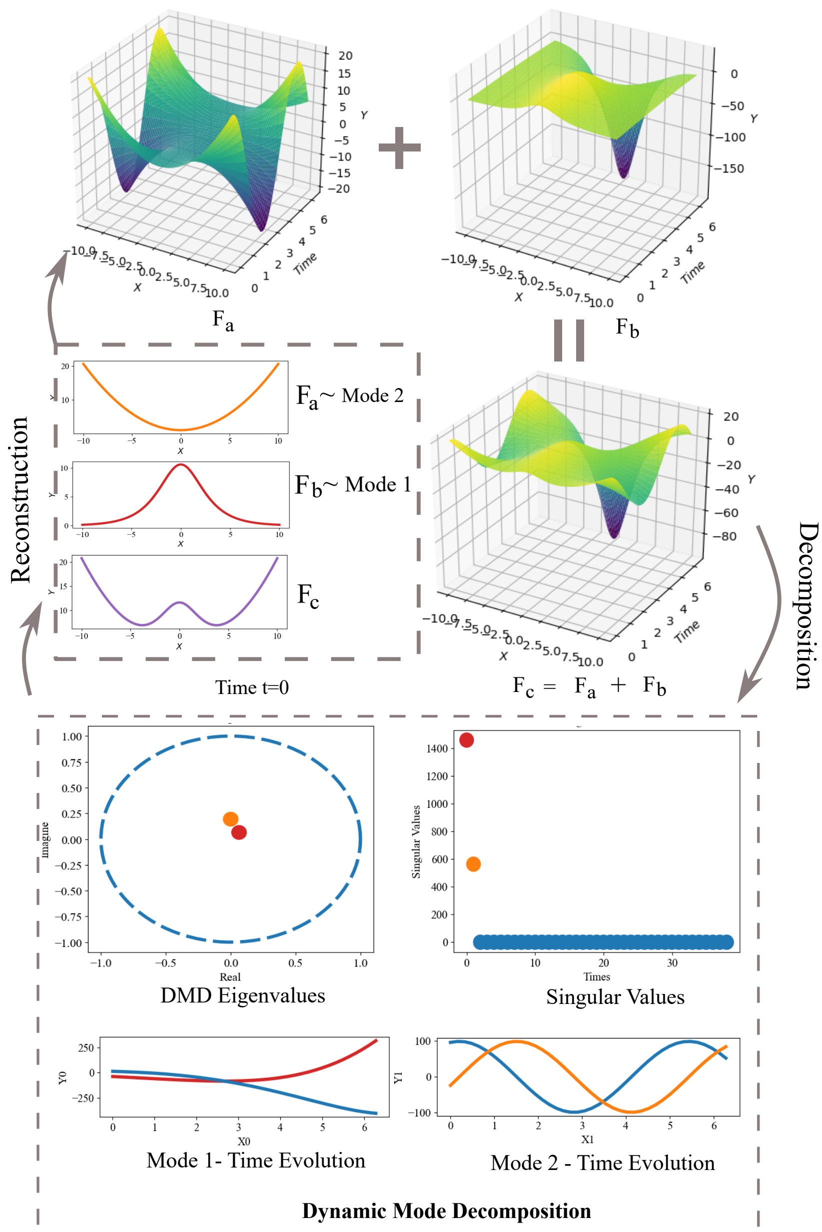


Figure 6: A sample DMD decomposition process. Given the 2D input function F_c which is the summation of two artificial time-dependent functions F_a and F_b, DMD extracts the singular values, eigenvalues, and a set of modes. The singular values reveal two high-energy modes. Each mode is a complex number and has a time dynamic behavior which is characterized by the position of eigenvalues with respect to a unit circle. In this example, DMD can decompose F_c back to F_a and F_b with the original time dynamics. The red and orange plots in the last row present the real values, while the blue plots demonstrate the imagine parts of the two modes. The X-dimension in the plots is the time, while the Y-dimension shows the frequency values of the mode over time.

The columns of Φ are called DMD modes. They have the same size and spatial configuration as x_i, except that each mode has a specific temporal behavior characterized by the corresponding eigenvalue λ in Λ. The DMD representation of a data snapshot based on the columns of matrices A and Φ can be derived as:

$$x_{DMD}(t) = \sum_{k=1}^{K} a_k \phi_k e^{(\delta_k + i w_k)t} \tag{8}$$

where δ_k is the growth rates, w_k is the frequency of the DMD modes ϕ_k which behave similarly as Fourier. Note that k is the number of modes. The DMD eigenvalues can be re-written as (which is a complex number):

$$\delta_k + i w_k = \frac{\log(\lambda_k)}{\Delta t} \tag{9}$$

The amplitudes a_k can be derived from the least squares fitting of the snapshots during the expansion. The entire process of DMD computation reveals several important components and characteristics of the method. Given a 2D random time-dependent sample data as illustrated in Figure 6,the singular values of X_1 can be visualized which show the low and high-energy modes. The number of high-energy modes can be used to reduce the dimensionality of the system and be utilized for the data compression task. Instead of storing the full spatial-temporal flow field, only the prominent modes need to be retained along with their corresponding eigenvalues for data reconstruction. Readers who are interested in how to choose the suitable number of modes are referred to the work by Jovanović et al. [42]. The visualization of DMD modes and their eigenvalues are shown in Figure 6. The dynamic behavior of a DMD mode depends on the values of its corresponding eigenvalue with respect to a unit circle. The mode either grows, decays, or neither if the eigenvalue is outside, inside or seats exactly on the unit circle. There is an oscillation if the eigenvalue has a *non-zero imaginary part.* The time dynamic evolution of the sample DMD modes are computed and visualized in the last row of Figure 6. *DMD modes also can be characterized by the speed of the time evolution.* For instance, the second mode has a higher frequency (as indicated by their quick changes) which makes it the fast modes, while the first mode is a slow mode as it moves/changes relatively slower.

Relation between slow DMD modes and large-scale coherent structures. As shown in previous works [70, 124], DMD modes can be used to extract spatial configuration and temporal information of certain coherent structures. Some works [129] argue that the DMD modes and their corresponding temporal evolution plots can separate coherent structures in spatial and spectral sense (according to their temporal behavior – grows, decays, or oscillates). Although this is not rigorously justified and not all flows have the exact same number of structures as the number of DMD modes, there is an important observation that motivates us to use certain DMD mode(s) to identify large-scale structures in the shear flow. That is, both the slow mode and the large-scale structures change/move slowly over time. For large-scale structures, this is because they have a larger inertia due to their higher kinetic energy. This observation leads us to develop an effective framework to separate large-scale structures from small ones in shear flows using DMD (Section 6.3). Note that, this relation between slow DMD modes and large-scale structures need not be true for other flows than shear flows (e.g., translational flows), which will be discussed in Section 6.5.

It has been shown that the standard/exact DMD cannot sufficiently handle the transient time phenomena (e.g., features exist for a short period of time) that often occur in the beginning of many turbulence flows (including shear flows) before a stable state is reached [47]. To address this issue and to develop a robust and generalized DMD-based analysis and visualization framework for different flows, a variant of DMD, called multi-resolution DMD (mrDMD) is needed.

2.3.1 Multi-Resolution DMD

Motivated by foreground/background subtraction in video processing, mrDMD tries to separate the slow and fast modes. As mentioned in the previous section, the slow mode has a relatively low frequency or slow growth/decay rate. They can be defined by small values of both the growth rates δ_k and frequency w_k. To obtain the slow modes, they can be plotted in a complex unit circle $(Re(|\log(\lambda_k)|), Im\,(|\log(\lambda_k))|))$. The nearer to the origin, the slower the corresponding DMD modes are.

mrDMD is a recursive process in which the slow modes are removed iteratively, and the remaining data is filtered for analysis of its higher frequency content. The entire process can be described below:

1. Compute DMD for the existing data.
2. Determine fast and slow modes relatively based on the values of w_k and δ_k.
3. Reconstruct data with only slow modes
4. Subtract the reconstructed data by the slow modes from the available data
5. Split the subtracted data in half in time
6. Repeat the procedure for the first and second half of the subtracted data separately (including this step)

Note that standard DMD can be considered as a special case of mrDMD when only one level mrDMD is performed with the additional benefit of the identification of the slow modes, which is needed for our problem.

3 Related Work

This chapter first discusses the existing works for the analysis and visualization of fluid flow data. Then, it presents closely-related works for coherent structure extraction and multi-scale separation of coherent structures for turbulent flow in both the fluid dynamics and the visualization communities.

3.1 The Analysis And Visualization Of Flow Data

Existing approaches for the analysis and visualization of flow data can be broadly divided into three main categories, namely dense/texture-based, topology-based and geometry-based methods. For time-dependent flows, time-varying series analysis also has been commonly used in scientific visualization in recent years.

Texture-based techniques encode flow data in textures to reveal patterns. Interested readers are referred to recent surveys for dense and texture-based visualization techniques [49]. **Topology-based methods** [50, 74] extract the topological skeleton that is comprised of first-order fixed points and their connectivity. Vector field topology provides a streamline classification strategy based on the origin and destination of the individual streamlines. Since its introduction to the visualization community [36], vector field topology has received extensive attention. A large body of work has been introduced to identify different topological features, including fixed points [76, 107] and periodic orbits [9, 102, 120]. Recently, Chen et al. [10] studied the instability of trajectory-based vector field topology and, for the first time, proposed a Morse decomposition for vector field topology computation, which leads to a more reliable interpretation of the resulting topological representation of vector fields. Szymczak et al. [100] introduced a new approach to converting the input vector field to a piecewise constant (PC) vector field and computing the Morse decomposition on a triangulated manifold surface. For the topological analysis of unsteady flow, *Lagrangian Coherent structures (LCSs)*, i.e., curves (2D) or surfaces (3D) in the domain across which the flux is negligible, were introduced to identify separation structures in unsteady flow. The computation of LCS was first introduced by Haller [33] by computing the *Finite-Time Lyapunov Exponent (FTLE)*, whose ridges indicate the LCS. FTLE has been compared with the separatrices in the steady case [84], and its computational performance has been improved substantially [27]. Recently, Fuchs et al. [24] presented an extended critical point concept to adapt the notion of vector field topology to unsteady flows. Sadlo and Weiskopf introduced a streakline-based topology based on generalized streaklines [85]. It successfully characterizes the saddle type of hyperbolic features and has been extended to study 3D unsteady flow topology [109].

Geometric-based methods [14, 59] utilize the geometric representation (e.g., integral curves including streamlines, pathlines, and integral surfaces) to depict the flow patterns. Salzbrunn and Scheuermann introduced *streamline predicates* that classify streamlines by interrogating them as they pass through user-specified features, e.g., vortices [88]. Later, this approach was extended to classifying pathlines [87]. At the same time, Shi et al. [93] presented a data exploration system to study the characteristics of pathlines based on various attributes, including winding angle. Recently, a statistics-based method was proposed to help select the proper set of pathline attributes to improve interactive flow analysis [73]. More recently, McLoughlin et al. [58] introduced the idea of a streamline signature based on a set of curve-based attributes including curvature and torsion. This streamline signature is used as a measure of the similarity between streamlines, pathlines, and helps domain experts place and filter streamlines for the creation of an informative and uncluttered depiction of 3D flow. Zhang et al. [127] extended Lagrangian accumulation to define an attribute field based on the accumulated values along integral curves. This attribute field employs an Eulerian representation of Lagrangian information in a similar fashion to texture-based techniques. It conveys a continuous representation of the variation associated with integral curve behavior to some extent.

Time-varying series analysis and visualization. TACs have been studied in scientific visualization in recent years [114, 122, 121, 31, 123, 51, 16]. Lee et al. proposed a visualization framework to analyze time-varying data sets with a TAC-based distance field [52]. This field provides a visualization to highlight the position of the features; however, it still does not provide certain details about an individual TAC, especially the temporal occurrence and period of an interesting feature. Wei et al. introduced a dual-space method to analyze turbulent combustion particle data, starting by clustering the time series curves in the phase space of the data, and then visualizing the corresponding trajectories of each cluster in the physical space [116]. The 2D time series curves are constructed using the correlation between temperature and mixture fraction. These curves are then clustered using the statistical model-based method. For spatio-temporal visualization of vortex features, Ferrari et al. [21] combined the vortex core lines extracted from the maxima score correlation of the two attributes λ_2 and vorticity along the time dimension to create an evolution surface of vortices. Ferstl et al. proposed a time-hierarchical clustering approach for analyzing the temporal growth of the uncertainty in ensembles of weather forecasts [22]. For a thorough overview of approaches for the time-varying data, please refer to the surveys [15, 55]. Our first TAC-based framework belongs to the time-varying series analysis and visualization category.

3.2 Coherent Structure Extraction

The visualization community has proposed numerous novel visualization techniques to help experts understand coherent features with a focus on the vortex detection. Vortices are flow regions where the flow motion is mostly rotating around an imaginary axis, called the *vortex core.* Vortices are one of the most important coherent structures in the fluid flows and a major component of turbulent flow. There is a large body of work for the detection and tracking of vortices over time [25, 43, 44, 71].

The foundations of many formal vortex definitions were laid out between the late 70s and early 90s. Most definitions [32] include two concepts that are still actively researched to this day, i.e., (1) vortex coreline – a line or curve that any mass of fluid moving around [56] and (2) an appropriate reference frame so that when instantaneous streamlines mapped onto a plane normal to the vortex core exhibit a roughly circular or spiral pattern [79]. Based on these two essential concepts, there are several groups of methods that identify the vortex corelines and vortex regions, respectively.

Line-based Methods. These methods mostly focus on the coreline of vortices. In steady flow, it can be identified using reduced velocity criterion, which considers the eigenvalues and eigenvectors of the Jacobian **J** [117]. However, in unsteady flow, there exists several methods, such as in magnetic fields or when the vector field is transformed into a reference frame, in which the flow appears steady [32]. The most popular coreline detection approach is via the Parallel Vector operator [71]. However, most line-based methods are numerically unstable and easily result in fragmented corelines, which may be challenging to clean up.

Geometric Methods. These methods mainly focus on constructing the skeleton of vortex tube. For 2D flows, Sadarjoen et al. [82, 83] presented two geometric approaches that are based on the shape of streamlines. Their *curvature center method* computes the density of curvature centers for a given set of streamlines. Their second method called *winding angle* is based on the angle between two subsequent polyline segments. Streamlines that have winding angles larger than a threshold are considered part of vortices.

Integration-based Measures. All the above methods are local. However, it was shown that there are classes of vortices that cannot be extracted by local methods, for instance attracting vortices that move on non-linear paths. As a solution, integration-based measures such as particle density estimation [119] and analyzing of Jacobian [117] were developed. These papers proposed to inject a number of particles and observe their attraction behavior over time.

Objective Methods. Objectivity refers to the invariance of a measure under a change of the reference frame. There are several measurements that could be used to identify vortex objectivity, for instance, relative vorticity tensor-based measures, strain tensor-based measures and vorticity-based measures [32].

Extraction of Vortex Boundaries. In addition to finding vortex corelines, the size of vortices is also of interest. With region-based methods, the size is determined by a threshold, which is typically difficult to set. Line-based methods may serve as a starting point for region growing approaches. There exist few works that aim to estimate the size based on streamlines or corelines [32]. To the best of our knowledge, the only work that has been focused on vortex boundary extraction was proposed by Haller et al. [35]. They proposed elliptic LCS, which preserves arc length and area in incompressible 2D flows and considered the outermost elliptic LCS, of a family of nested elliptic LCSs, as the boundary of a coherent vortex.

Region-based Methods. Region-based or Eulerian-based methods identify a volume of coherent structures' behavior based on some physical attributes such as pressure [39], vorticity, helicity, Q-criterion, Okubo-Weiss criterion and λ_2 criterion [32] and other derived attributes from the velocity field. Structure identification is dependent on thresholding. To detect vortices, for example, fluid mechanic researchers can play with the Q criterion [11] or λ_2 [34] to identify the volume of vortex-like behavior. Much fluid mechanics focuses on Eulerian-based approaches because they provide more physical information. It is important to note that the most commonly used physical attributes including Q, vorticity or λ_2 naturally focus on small-scale structures due to its definition, i.e., it is based on vorticity and strain which are inherently small-scale quantities. One way to characterize large (or coarse) scale structures or is to use a convolution kernel, usually an averaging box filter as proposed by Treib et al. [104]. Besides Eulerian methods, moving ensemble averages [67], two-point correlations [94] or Fourier analysis [66] have also been widely used in fluid dynamics. However, they suffer from the same issue. Choosing a proper threshold requires a substantial amount of time on trial and error experiments which might lead to undesired structures.

The last set of techniques that can be used to extract the coherent structures is vector field decomposition. In these techniques, the vector field is decomposed into different components with hope that the components would help to reveal some desired features. The most well-known method - Helmholtz-Hodge-decomposition (HHD) [4] - decomposes the flow into curl-free and divergence-free vector fields. However, HHD is not applicable for multi-scale coherent structure analysis. Another decomposition method which is considered the most related to DMD is proper orthogonal decomposition (POD) [45] or principal component analysis in statistics. In POD, the data is decomposed into an orthogonal basis of spatial correlated modes, called POD-modes. POD has been commonly applied in fluid mechanics to determine coherent structures from multiple flow snapshots as well as to extract high energy flow features [75, 115, 78] and to improve the vortex detection result [7]. In Section 6.4, we demonstrate that POD is not reliable for coherent structure extraction in shear flows. This challenge leads us to the third work which utilizes the novel Dynamic Mode Decomposition to extract the large-scale coherent structures in turbulent flows.

3.3 Multi-scale Processing of Flow Data

Given the increasing complexity of flow data and the multi-scale nature of the turbulence flow, multi-scale processing is required to delineate physical events arising across different scales. There are two different types of multi-scale processing of flow data, concentrating on either its geometric aspect or its physical aspect, respectively. The representative method for the former is the construction of the topological hierarchy of the velocity (or vector) fields. Both bottom-up and top-down strategies have been introduced. The bottom-up strategy is usually realized as some simplification process, that gradually merges (or cancels) pairs of topological features, including fixed points [53, 12, 106, 107, 101, 118, 126, 97, 96] and periodic orbits [9], based on certain proximity (e.g., Euclidean or flow distance) between features. In contrast, the top-down approaches usually start with a coarse structure and gradually refine (or split) it. The top-down strategies are only seen in techniques based on Morse decomposition [10, 100, 8], where the structure with different fineness can be obtained with different particle advection times (i.e., a multi-valued flow map). Nonetheless, for turbulent flows, a topology based approach may generate too many detailed structures that complicate the interpretation of the flow behavior.

Different from the above topology-based multi-scale processing, the non-topology-based methods define the scales in the frequency space. Mundane methods of filtering out scales for coherent structure extraction, such as Fourier transforms [2, 23, 125] have been applied successfully. They have seen wide applications in homogeneous isotropic turbulence, and provide glimpses on how the real-world turbulent cascade operates. However, they are inherently non-local and are in practice limited to periodic and homogeneously discretized domains. Wavelet methods [60, 17, 18, 19] were developed to overcome the limitations of Fourier transforms, but have not seen widespread adoption in fluid mechanics. This is because for turbulence analysis, the wavelets considered rapidly become complicated and problem-specific instead of universal [86].

4 Physics-based Pathline Clustering and Exploration

To incorporate more detailed physics into the analysis and visualization of unsteady flow, we propose a novel visual analysis framework based on the temporal behavior of local, physical attributes of interest measured along individual pathlines. Similar to Lee et al. [52], we refer to the temporal profile of the attribute along a pathline as time activity curve (TAC), specifically a Lagrangian TAC. Unlike Shi et al. [93] and other similar methods that also compute the pathline attributes, our framework takes into account the arbitrary movement of the observer. That is, the physical properties are computed from the new instantaneous vector fields after applying the optimal reference frames, which achieves a better alignment between physical features and the geometric representation of the flow, as demonstrated in [30]. The benefits of analyzing the flow behavior based on TACs are two-fold. First, they are 1D plots that are independent of the flow dimension (i.e., applicable to both 2D and 3D data). Second, the geometric characteristics of TACs (e.g., ascending, descending, peaks, valleys, etc.) reveal the interaction of flow particles with physical features over time. This enables us to explain the geometric characteristics of the corresponding pathlines and vice versa. By utilizing the advantages of analyzing TACs, we introduce a new spatio-temporal, hierarchical clustering of pathlines based on their respective TACs and a TAC-based pathline selection and exploration. Central to these techniques is a comprehensive distance metric for the comparison of two TACs, which we refer to as a TAC similarity metric (TSM) that incorporates the global correlation of pair-wise TACs and the spatio-temporal distances between them (Section 4.1.1). We develop a visual exploration system that integrates the aforementioned TAC-based analysis and exploration techniques with a modified edge-bundling visualization of TAC clusters. We have applied our TAC-based exploration system to a number of 2D and 3D unsteady flows. Our framework effectively reveals the two-layer configuration of a vortex and its decay over time in vortex shedding (Figure 14), which is difficult to reveal via conventional methods. We also facilitate interpretation of the temporal behavior of a 2D cross section of vortex ring, including its interaction with a wall and its breakdown.

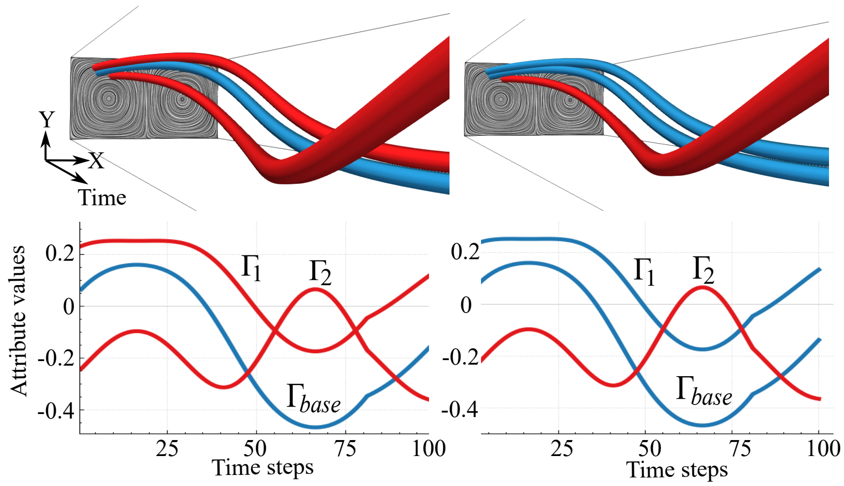
4.1 TAC-based Flow Exploration Framework

Overview. Our pipeline consists of two main phases: computation and exploration (Section 4.2). We concentrate on the computation phase in this section. First, we densely and uniformly sample the particles in the flow domain and compute pathlines. Depending on the attributes of interest, the corresponding TACs are derived and segmented into multiple time intervals (Section 4.1.4). Next, we perform a hierarchical clustering based on the characteristics of the entire TACs (Section 4.1.2). Based on the global clustering result, we perform a hierarchical temporal clustering of TACs to capture the level-of-detail characterization of their temporal behavior (Section 4.1.5). From the spatio-temporal clustering result, flow exploration (Section 4.2) is conducted from the following three perspectives: flow space, attribute space and temporal space. In the following we detail our TAC-based clustering.

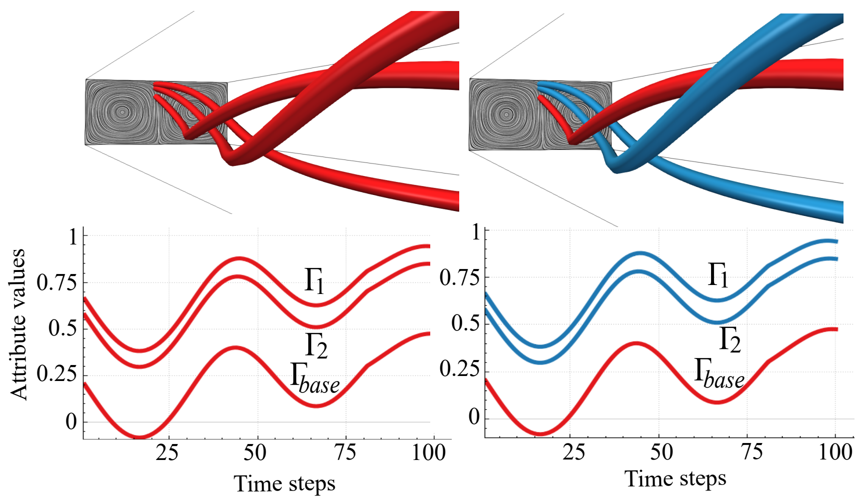
4.1.1 TAC-based Similarity Measure

To assist the spatio-temporal clustering of TACs, we first describe our similarity measure for TACs. To compare the difference in the characteristics of two TACs, the similarity measure takes both the temporal trends and magnitude of TACs into account. The traditional distance metrics, such as the Euclidean distance and the Pearson correlation coefficient, concentrate on either the trend or the magnitude of the TACs and cannot satisfy our needs. Figure 7 (left column of (a) and (b)) illustrates the limitations of Euclidean distance and Pearson correlation coefficient in characterizing the difference in a number of representative TACs. Another metric for measuring the similarity of two time series is Dynamic Time Warping (DTW) [52]. DTW considers both shift and deformation of the time series. However, the time stamp for each sample in the TACs has specific meaning which requires us to align the TACs based on the time stamps, making DTW less suitable in our cases.

To address the limitations of the existing similarity measures, we introduce a new TAC Similarity Measure (TSM) to calculate the similarity of TACs based on their spatial and correlation differences. The proposed measure requires linear time to compute; thus, it is practical to apply the measure



(a) Euclidean (left) vs Our measure (right)



(b) Pearson correlation (left) vs Our measure (right)

Figure 7: Comparison of TSM and (a) Euclidean and (b) Pearson correlation, respectively. The TACs are computed based on the λ_2 attribute on the Double Gyre simulation. Colors represent clusters. In both cases, the difference of TACs cannot be accurately measured by Euclidean distance or Pearson correlation (left column). (a) $D_e(\Gamma_{base}, \Gamma_1) = 42.32 > D_e(\Gamma_{base}, \Gamma_2) = 38.87$, using the Euclidean distance. (b) $D_p(\Gamma_1, \Gamma_2) = D_p(\Gamma_2, \Gamma_{base}) = 1$ using Pearson correlation, resulting in all of them belonging to the same group. TSM can differentiate the behavior of TACs more accurately in both cases (right column): (a) $D_{tsm}(\Gamma_{base}, \Gamma_1) = 62.32 < D_{tsm}(\Gamma_{base}, \Gamma_2) = 77.74$ (b) $D_{tsm}(\Gamma_1, \Gamma_2) = 25.32 < D_{tsm}(\Gamma_2, \Gamma_{base}) = 52.64$.

to large data sets. TSM is defined as follows:

$$D_{tsm}(\Gamma_1, \Gamma_2) = (1 + \mathrm{P}_c \; D_{corr}(\Gamma_1, \Gamma_2)) \; D_e(\Gamma_1, \Gamma_2) \tag{10}$$

$$D_{corr}(\Gamma_1, \Gamma_2) = 0.5 - \frac{cov(\Gamma_1, \Gamma_2)}{2\sigma_{\Gamma_1}\sigma_{\Gamma_2}} \tag{11}$$

where cov is the covariance and σ_Γ is the standard deviation of Γ.

$$D_e(\Gamma_1, \Gamma_2) = \sqrt{\sum_{i=1}^{n} (\Gamma_1[i] - \Gamma_2[i])^2} \tag{12}$$

In the above definition, $D_e(\Gamma_1, \Gamma_2)$ represents the Euclidean spatial distance between the two TACs Γ_1 and Γ_2. $D_e(\Gamma_1, \Gamma_2)$ addresses the challenges illustrated in the left column of Figure 7(a) where TACs exhibit similar trends but a different spatio-temporal distance measure. $D_{corr}(\Gamma_1, \Gamma_2)$ measures the global correlation between TACs Γ_1 and Γ_2. If $D_{corr}(\Gamma_1, \Gamma_2) = 1$, then the first term in Equation 10 $(1 + P_c \times D_{corr}({\Gamma_1}'_i, {\Gamma_2}'_i))$ equals 2 which means Γ_1 and Γ_2 have opposite trends. In contrast, if $D_{corr}(\Gamma_1, \Gamma_2) = 0$, Γ_1 and Γ_2 have the same trend and the value of the first term is 1. $D_{corr}(\Gamma_1, \Gamma_2)$ aims to resolve the ambiguity illustrated in Figure 7(a) where Γ_1 and Γ_2 have similar Euclidean distance, but different correlation distance relative to Γ_{base}. By using $D_{corr}(\Gamma_1, \Gamma_2)$, the TSM measure can differentiate Γ_1 and Γ_2; hence it groups Γ_1 to Γ_{base} rather than Γ_2 illustrated in the right column of Figure 7(a). $D_{corr}(\Gamma_1, \Gamma_2)$ introduces a penalty factor P_c, which represents a user-assigned importance for the spatial difference and the global correlation, respectively. The higher value of P_c, the more weight given to the global correlation. By default, we set $P_c = 1$.

We use multiplication instead of addition to combine the first and second terms in Eq. 10 due to the relation between the two terms, and their value range difference. The first term is equal to 1 when the two TACs contain similar trends. In this case, the distance between two TACs is completely based on the second term (i.e., Euclidean distance). If the two TACs have opposite trends, then the distance between the TACs is expected to be large. By multiplying, we magnify the second term by a maximum of two when the TACs have inverse trends.

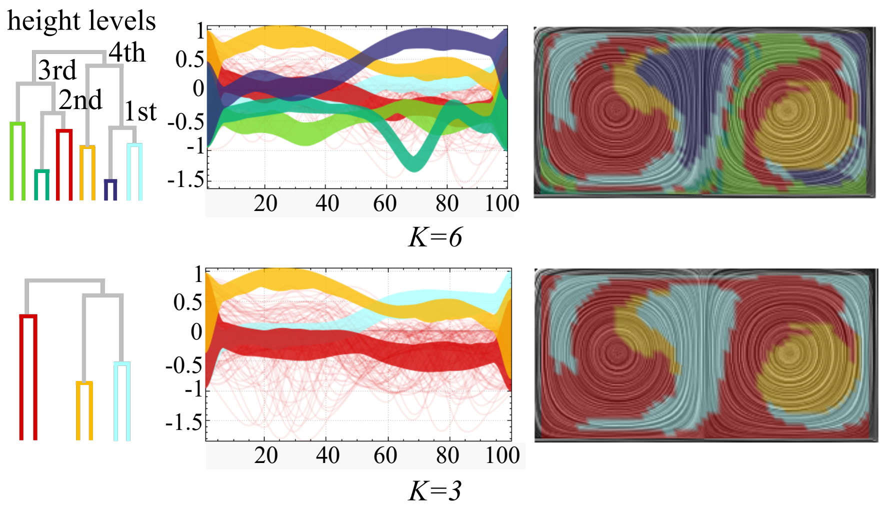


Figure 8: Clustering results of the Double Gyre flow using Lagrangian TACs of attribute curl with different numbers of clusters (K=6, 3 from top to bottom, respectively), specified by the user in the hierarchical tree view (left column). The height levels of the hierarchical tree indicate the merging order. The lower level, the sooner the leaf nodes will be merged. The middle two columns show the TAC clusters, stacked plots and their representative TAC curves. The right column shows the clusters in the flow domain. The time window is $T = 100 \times 0.01$ with 100×50 sampling points.

Figure 7 illustrates the advantages of TSM (right column) over the Euclidean distance and Pearson correlation. In both cases, TSM can differentiate the behavior of TACs more accurately.

With the above similarity measure, we can develop a TAC-based pathline selection and exploration strategy which allows us to highlight pathlines whose TACs exhibit certain specific characteristics (i.e., the distance of their TACs to a reference TAC is smaller than a threshold). We will defer the discussion of this functionality until the results.

4.1.2 TAC-based Clustering

In order to provide different levels of detail for flow behavior w.r.t. the local attributes, we perform the clustering of TACs using the new similarity measure over all temporal samples, coupled with the popular agglomerative hierarchical clustering (AHC). The linkage type used in this work is the complete linkage since it is better for finding compact clusters of approximately equal diameter [13].

AHC gives rise to a hierarchical tree with each node representing a cluster and each bifurcation representing a merging (see an example shown in view 2 of Figure 13). The different heights of the hierarchical tree which correspond to the distance values of the two clusters indicate the merging order. The lower the level, the sooner the leaf nodes will be merged. A sample merging order can be found in Figure 8 where the hierarchical tree with six leaf nodes is reduced to three nodes. As the number of clusters represents the level of abstraction, increasing the number of clusters results in more details to be revealed as shown in Figure 8 (top row). Choosing a suitable number of clusters of pathlines to reveal the most interesting flow behavior is not trivial, and it is a trade off between the details of flow behavior and the clearness of the cluster structures. In practice, it is an exploration process. Since the hierarchical tree is a natural product of AHC, a user can interactively select the number of clusters to show after the clustering computation (i.e., no re-clustering is needed).

The traditional AHC runs in $O(N^3)$ time where N is the number of TACs. This running time makes the AHC impractical for large datasets. We implement the parallel, locally-ordered AHC proposed by Walter et al. [113], which runs in sub-quadratic time. This greedy algorithm is based on the following observation: if two clusters, A and B, are nearest neighbors, they will eventually be merged together. Thus, it is better to find all possible A and B pairs, and cluster them immediately. To quickly identify the closest clusters for merging, we store clusters in a K-d tree which requires $O(\log n)$ running time.

4.1.3 TAC-based Temporal AHC

Two TACs that belong to two clusters may possess local segments having similar behavior (Figure 9(c-d)), which cannot be captured in the above global clustering along the entire time range. To address this, we propose a hierarchical clustering algorithm in the temporal dimension, i.e., a temporal AHC.

We aim to address following technical challenges to achieve temporal AHC: (1) identify the appropriate temporal partitioning; (2) perform the AHC within each time interval obtained from step one; (3) handle the transition of AHC results between consecutive time intervals. Solving these problems is not trivial. First, all TACs have different temporal behavior. Even the TACs that belong to the same cluster may exhibit slightly different behavior, which makes the selection

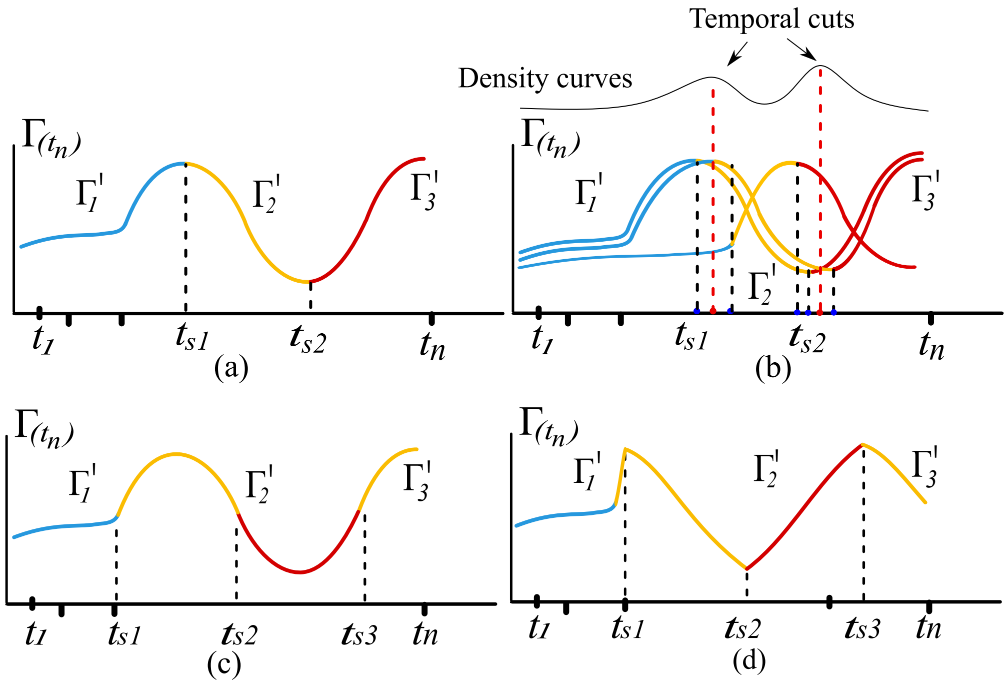


Figure 9: (a) An example TAC with three events. t_{s1} and t_{s2} are two split points defined at the extrema. (b) An example of identifying temporal cuts for temporal clustering. Cuts are selected at the maxima of the density curve. (c) A similar TAC as in (a), but the split points are defined at the maxima and minima of the derivative curve (d).

of cutting points (or cuts) for temporal partitioning difficult. Second, the AHC performed on individual time intervals and along the time axis should be consistent in terms of the error threshold and similarity characterization of the clusters. Third, the AHC results obtained in consecutive time intervals may not be identical. It is important to keep track of their transition relation (i.e., bifurcate or merge) across the cuts. In the following, we detail our solutions to these challenges.

4.1.4 Time Interval Segmentation of TAC

To study TACs in a level-of-detail fashion, we apply time interval segmentation to a group of TACs. The time intervals that segment TACs must preserve TAC characteristics. In other words, one primitive trend of a TAC is not expected to be segmented into two time intervals, which causes fragmentation. For an individual TAC, we can simply apply 1D Morse decomposition to generate the temporal sequences of TAC segments, as shown in Figure 9(b). However, for a group of TACs, it is not guaranteed that the segment split points are identical. To address this, we utilize a 1D Gaussian kernel density estimation (KDE) and choose the point with the highest estimated density as the split point. Specifically, we first identify the inflection points for each TAC. Let $x_1, x_2, ..., x_n$ be a set of 1-dimensional inflection points on $\mathbb{R}$ and let H be a positive definite bandwidth value. The univariate fixed bandwidth kernel estimator is defined as [95]: $f(x) = \frac{1}{NH}\sum_{i=1}^{N} K(\frac{x-x_i}{H})$, where K is the Gaussian kernel $K(x) = \frac{1}{\sqrt{2\pi}}e^{-x^2/2}$. Selection of the bandwidth value H is important in KDE as it can make the density estimate smoother or noisier. However, in our case, the value of H does not affect the position of the point with the highest estimated density. Thus, by default we set H to 1.

The cutting points obtained from the above KDE segment the entire time period T into a number of intervals of varying length, referred to as $T = \langle T_1, T_2, ..., T_m \rangle$. In this way, all TACs are segmented by these splitting points which attempt to preserve the most common characteristics of all TACs. An example of time interval segmentation is illustrated in Figure 9(c).

4.1.5 Temporal Hierarchical Clustering

After performing the temporal partitioning and obtaining the local time intervals, we now perform AHC within each interval. We apply the proposed similarity measure (Eq. 10) for temporal clustering.

Assigning the cluster number for each time interval is difficult because, on one hand, the number of time intervals in time-hierarchical clustering varies, while on the other hand, the cluster distances in different time intervals may be different. To show the consistent changes across time intervals, the same treatment needs to be applied uniformly. Therefore, we use the distance threshold ε for the global clustering to guide the clustering within individual time intervals. Specifically, the distance threshold ε_i for time interval T_i is determined by the time range of the interval, i.e., $\varepsilon_i = \frac{|T_i|}{|T|}\varepsilon$. In this way, it is foreseeable that there are more clusters generated in the time intervals where the TACs behave more diversely, i.e., when TACs have larger dissimilarities.

The goal of temporal clustering is to build up a hierarchical tree of the input m time intervals obtained in the previous temporal partitioning, i.e., m leaf nodes of the tree, so that the level-of-detail of a TAC's behavior can be observed in the temporal dimension. In contrast to the spatial hierarchical clustering, in which any two clusters can be selected for a merging operation, in temporal hierarchical clustering only two clusters that are contiguous in time can be merged together, which makes the merging operation simpler. In our implementation, starting from the initial m leaf nodes (i.e., m initial time intervals), a distance array $D \in \mathbb{R}^{(m-1)}$ is created. Each entry indicates the dissimilarity after a pair of consecutive time intervals are merged into one. $D[i]$ can be computed as follows.

$$D[i] = \eta(T_i) + \eta(T_{i+1}) - \eta(T_i \cup T_{i+1}) \tag{13}$$

where $\eta(T_k)$ is the average pairwise dissimilarity within a time interval T_k, $T_i \cup T_{i+1}$ is the new time interval obtained by merging T_i and T_{i+1}.

$$\eta(T_k) = \sqrt{\frac{\sum_{i=1}^{|T_k|} \sum_{j=i}^{|T_k|} (D_{tsm}(\Gamma_i, \Gamma_j))^2}{|T_k|(|T_k| - 1)}} \tag{14}$$

$\eta(T_k)$ reflects the compactness of the TACs in the time interval T_k. The larger the value of $\eta(T_k)$, the further the TACs in T_k are located from the centroid.

In spatial AHC, the two clusters with the smallest distance are selected for merging. Similarly, in temporal hierarchical clustering, the two time intervals with the smallest dissimilarity changes are merged together first. In other words, time interval T_k and its neighboring time interval T_{k+1} that satisfies $D[k] \leq D[i], \forall 1 \leq i \leq M$, are first merged together to generate a new time interval $T_k + T_{k+1}$ and then removed from the node list. Consequently, a new $m-2$ dimension distance array $D_{(m-2)}$ is generated with the remaining $m-1$ nodes. The above merge process is iterated until only one time interval, i.e., the entire time period, remains as the root of the temporal hierarchical tree. The height of the temporal hierarchical tree built on m time intervals is $m-1$. On the i^{th} level of the tree, i.e., the height is i, there are $m-i$ time intervals. This temporal clustering strategy can be applied to all TACs or a subset of TACs grouped based on the global clustering results from

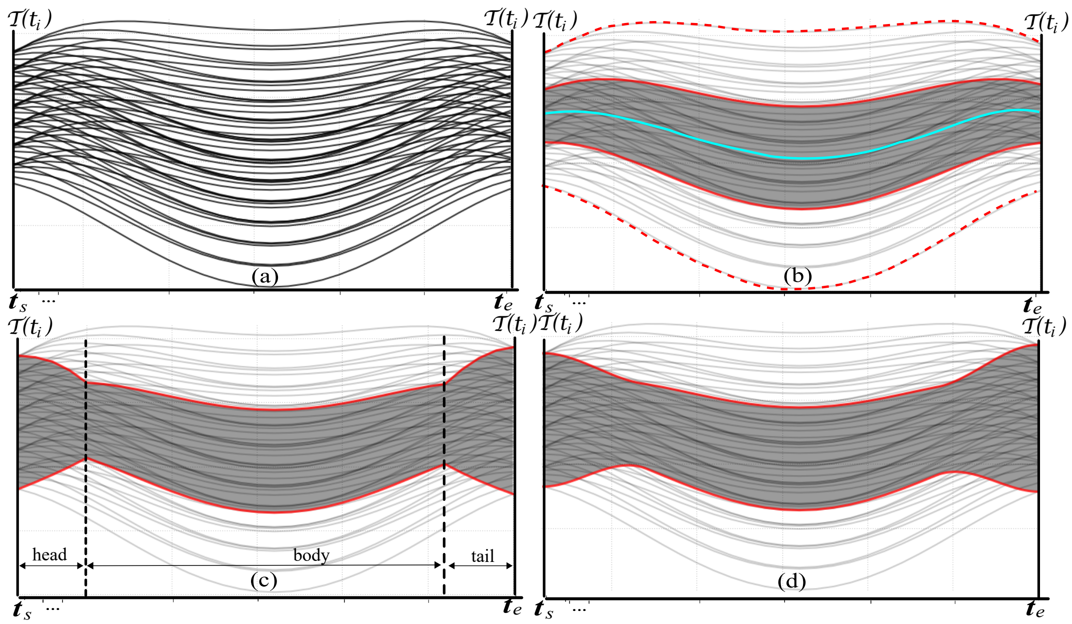


Figure 10: Illustration of edge-bundling method for TACs cluster visualization. (a) individual TACs, (b) edge-bundling rendering of a group of TACs by thinning toward the representative TAC, (c) preservation of the range of a group of TACs by adding head and tail segments, (d) the final result.

Section 4.1.2. For the latter, different numbers of temporal cuts (and temporal segments) may be resulted for different global clusters, depending on the overall TAC behavior of each cluster (see Figure 16(c) for an example.)

4.1.6 Visualization of TAC Clusters

An improved edge bundling technique. Traditionally, visualizing clusters is achieved by assigning each cluster a specific color. However, showing all TACs with colors assigned based on their cluster IDs will result in clutter, making it difficult to recognize the behavior encoded in TACs as demonstrated in Figure 10(a). To address this issue, we adapt the edge bundling technique for parallel coordinate plot (PCP) visualization developed by Palmas, et al. [68].

Given a cluster of TACs, the average (cyan curve) and boundary TACs (dotted red curves) are first derived (Figure 10(b)). Then we offset the two boundaries towards the centroid. The offset operation does not change the overall behavior of the TACs in the cluster, while the range of the

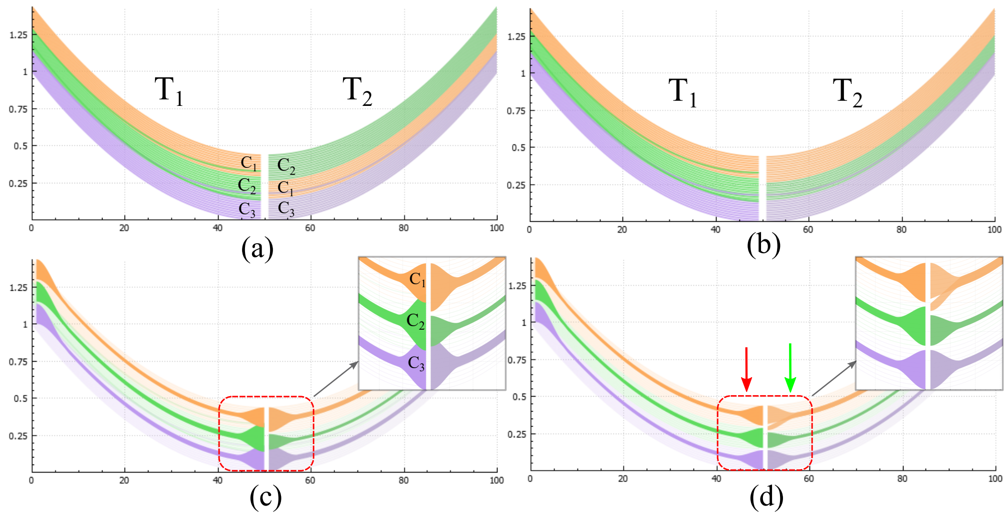


Figure 11: Visualization of transition between time intervals. Results before (a) and after (b) cluster ID adjustment, respectively. (c) edge-bundling visualization of the result. (d) modified edge-bundling visualization. Magnified views show the transition between two time intervals.

cluster, i.e., the coverage of the attribute values at the two ends of TACs t_s and t_e is changed. To preserve the coverage information of the cluster, we create a head and a tail for the edge-bundling by keeping the maximum and minimum of the attribute values at the two ends (Figure 10(c)). After we apply the edge-bundling method, the clusters in Figure 10(a) are shown as Figure 10(d), which greatly reduces the clutter.

To ensure the color consistency for the temporal clustering visualization of TACs, we assign a color to a cluster C_p based on its *main source cluster*, i.e., the cluster from which most TACs in C_p originate in the previous time interval. For example, in Figure 11(a), cluster C_1 in T_1 is the main source of cluster C_2 in T_2. Thus, the color of C_2 in T_2 will be set in a manner consistent with C_1 in T_1 (Figure 11(b)).

Visual overlapping persists at two ends of edge bundle as shown in Figure 11(c). To address this limitation, we offset proportionally to clusters' size, whose heads or tails are overlapping. As illustrated by the red arrow in Figure 11(d), the minimum value of C_2 at the tail end is increased and the maximum value of C_3 at the tail end is decreased, eliminating the overlapping between C_2 and C_3 while preserving the relative range size simultaneously. Removing overlapping at the tail

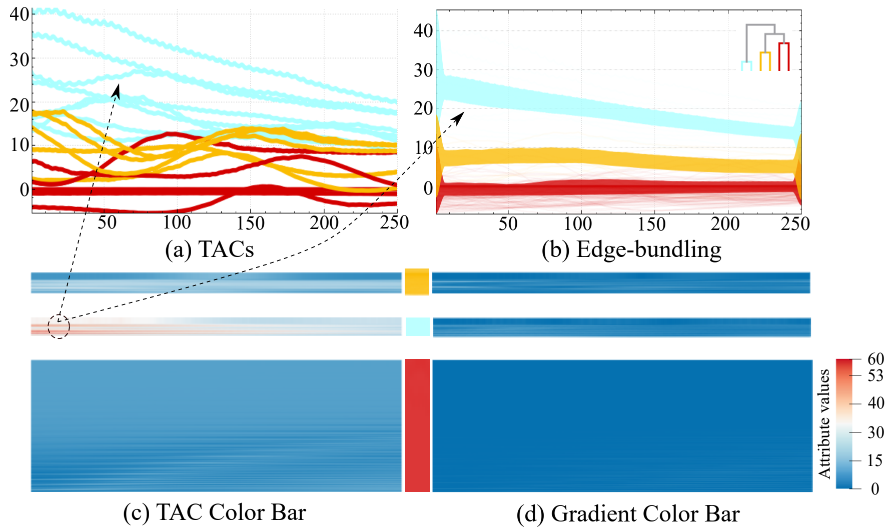


Figure 12: The visual comparison of edge-bundling with 1D TAC color plot. (a) Actual TACs exhibit occlusion. (b) Edge-bundling visualization provides an overview of each group of TACs. (c) The smooth 1D stacked color plot means that the TACs are grouped effectively. (d) The uniform blue color in the gradient color plot indicates the similarity among neighboring TACs in each cluster.

of T_{k-1} makes the boundaries of source clusters clear. To fully resolve the connections among time intervals, we visualize both main and minor sources at the head of a cluster. From Figure 11(d), we can easily ascertain the transition of clusters between two time intervals.

A 2D stack plot. Although edge-bundling visualization is an effective way to provide an overview about a group of TACs , the detailed behavior of individual TACs and the difference between TACs in the group is not conveyed effectively (see the cyan TAC cluster shown in Figure 12(b)). To address this, we visualize each TAC using a 1D bar, whose colors are determined by attribute values of the TAC over time. We then stack these 1D color plots to form a 2D color plot (Figure 12(c)). Note that TACs belonging to the same cluster are rendered next to each other. With this condense representation, one can easily assess the clustering quality. That is, if the color in this 2D plot is smooth, it means that the neighboring TACs have similar characteristics, indicating a good clustering result. The distance between two neighboring TACs is also converted to the 1D color bar to create the gradient plot. If neighboring TACs have similar patterns, the gradient between them is small, then the plot exhibits mostly uniform color.

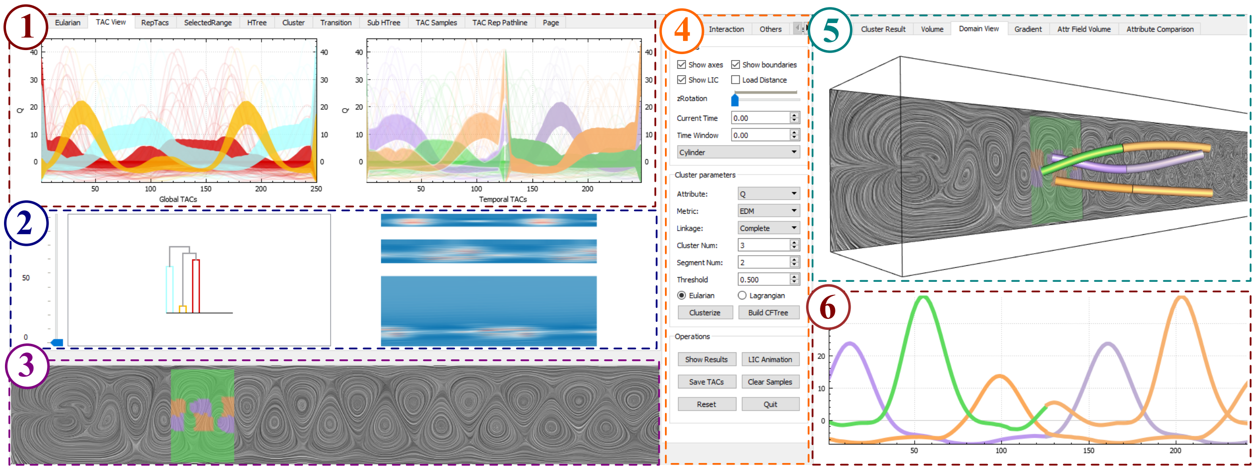


Figure 13: User interface of our system. (1) Multiple tab views show TACs with temporal and global clustering results. (2) Hierarchical Tree generated from the AHC and the TAC color bars. (3) the Line Integral Convolution (LIC) image generated for the first time step, allowing users to select the region of interest (ROI). (4) The control panel providing functions for user interaction. (5) Pathlines in the flow domain and (6) the detailed TACs in the selected region of interest.

4.2 TAC-based Flow Visualization System

A multiple coordinated view system is developed to enable visualization and exploration of unsteady flows using TACs. Users can flexibly switch between different views to perform analysis and comparison. Our system provides visualizations in the temporal space, attribute space and the original flow domain. Figure 13 illustrates the user interface. ① The top left view has multiple tabs showing TACs with temporal and global clustering results. ④ The center view is the control panel that provides functions for user interaction. ③ The bottom view shows a Line Integral Convolution (LIC) [6] image generated for the first time step, and enables users to select the region of interest (ROI). ⑤ The top right view visualizes pathlines in the space-time domain and the detailed TACs (⑥) in the selected region of interest. View ② shows the hierarchical clustering tree. Users can interact with the system and explore the clustering results in four ways: (1) The user can select a region of interest in view ③ and inspect the behaviors of the particles seeded within this region. A volume rendering coupled with iso-surfacing is used to visualize the behavior of the pathlines, which are colored based on the spatio-temporal clustering in view ⑤. (2) The user can choose a specific cluster in view ① to analyze its TACs' behaviors and highlight specific TAC and its corresponding pathline (in view ⑤). (3) The user can choose a temporal cluster to analyze TACs' behavior within a specific time interval. (4) The user can manipulate the error threshold or number of clusters in view ② to inspect the abstract flow behavior with different levels of details (see Figure 8 for an example).

4.2.1 TAC-based Exploration

In addition to the TAC-based clustering framework introduced above, which is suitable for providing the overview of the behaviors of the unsteady flows. Our visual analytic system also supports a number of exploration functionality based on TACs.

TAC-based pattern search. Similar to the pattern search from a set of integral curves based on some template curve and its geometric characteristics, the TSM distance metric for TACs can be applied to perform the pattern search based on TACs' behaviors. Unlike clustering which considers all pairwise TAC distances, the pattern search starts with a reference TAC that can be selected by domain experts. The system then returns the most similar TACs and their corresponding pathlines. This feature provides the freedom to users to customize the TAC characteristics interesting to them (Figure 18(b)).

Input filtering. Instead of using a set of densely placed pathlines in the first time step of the data sets, the experts may wish to focus on a specific subset of pathlines based on some prior knowledge about the flow. For instance, in the study of the relation between shear layer and the vortex formation, the users may select a subset of pathlines seeded within the shear layer of the flow (i.e. regions with negative Q values) and study the characteristics of their corresponding TACs in hope with finding the pathlines starting from the shear layer that may participate in the formation of vortices (i.e., entering vortex regions) at a later time (Figure 19(b)). On the other hand, the experts may be interested in certain features arising in a later time and wish to see the origin of the particles that enter these features. In this case, rather than computing the pathlines using forward tracing, backward tracing can be used to compute the pathlines starting from the feature areas. In all these ad-hoc exploration, our system can perform clustering and visualization on only a subset of pathlines.

4.3 Applications

We have utilized our TAC-based exploration framework to help our experts from aerospace engineering and mechanical engineering analyze vortex structures in different situations (Section 4.3.1) and help reveal subtle difference in the seemly symmetric flow behavior (Section 4.3.2).

4.3.1 Vortex Structure Analysis

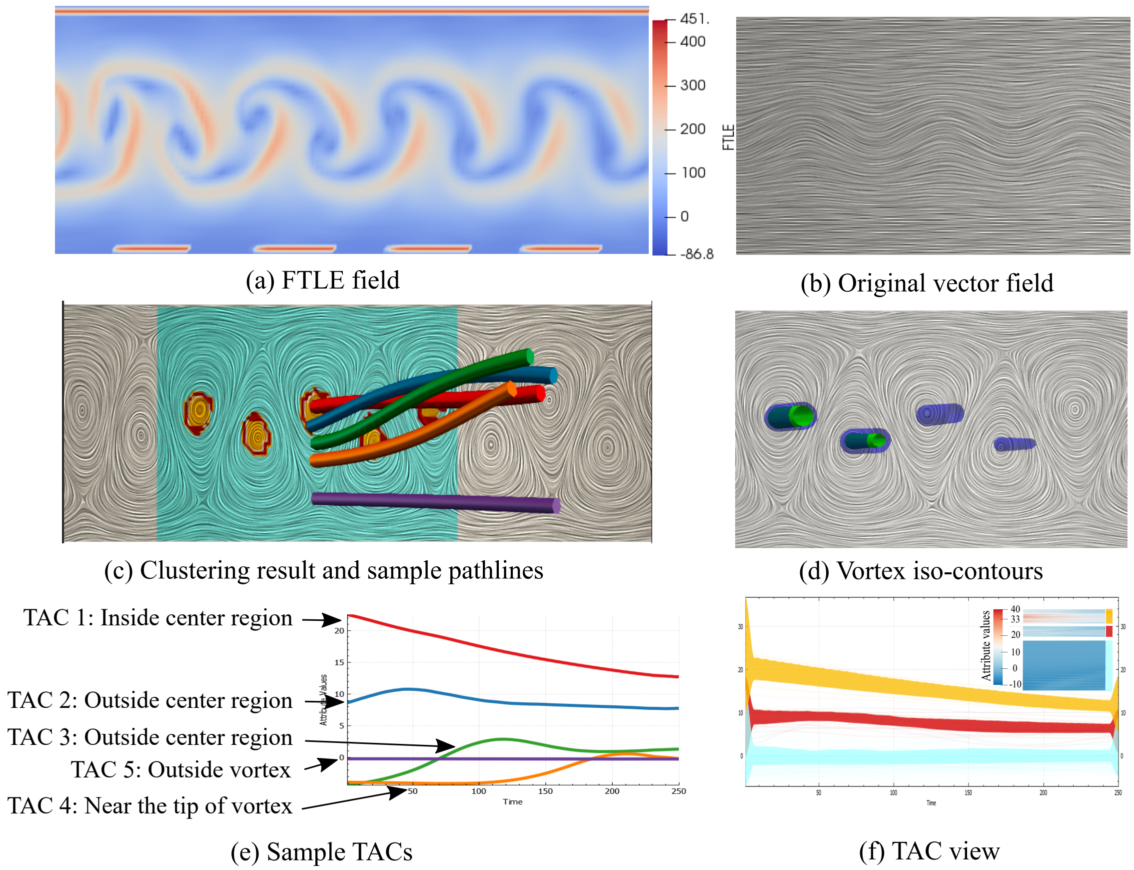
In the following, we demonstrate how we apply the proposed TAC-based clustering to help study vortex behavior in a number of 2D and 3D unsteady flows. Vortices are one of the most important dynamics in flow that often relate to energy/material transport and mixing [54]. The attributes applied in the following studies are mostly Q, λ_2, and vorticity (or curl in 2D). Although these attributes have different temporal trends, they often result in similar clustering results as shown in the supplemental document. In the following, we only provide the representative result for each data set using one of these attributes. In practice, the user should explore different attributes to identify the attribute that best reveals the flow behavior of interest.

Double Gyre. We first apply our method to a simple analytic flow, i.e. Double Gyre flow. We used the time window $T = 100 \times 0.01$ with 100×50 sampling points. Figure 8 shows the hierarchical clustering result for the flow based on Lagrangian TACs of the curl attribute with different cluster numbers. Since the number of clusters represents the level of abstraction of the representation, increasing the number of clusters causes more details to be revealed. The clustering of the flow domain elucidates the LCS structure. One noteworthy observation is that the vortex structure is stable when the cluster number is changed.

2D flow behind a cylinder. Next, we apply our technique to a 2D simulation of the flow behind a square cylinder with a Reynolds number of 160 [117]. The simulation covers a subset of the spatio-temporal domain, $[-0.5, 7.5] \times [-0.5, 0.5] \times [15, 23]$, where the vortex shedding is fully formed. According to the domain experts, the core region of a vortex in this flow has a motion close to that of a rigid body rotation, which helps to preserve the shape of the vortex. However, the concentrated vorticity in the vortex cores will diffuse due to viscosity (i.e., friction) and the absence of an external forces to maintain the rotation [54]. The diffused vorticity will reach the outer layer of vortices where it interacts with vorticity from other vortices, thus losing the coherent character. The overall structure of the vortices is stable due to interleaving and somewhat symmetric configuration of the counter-rotating vortices.

We choose the first 250 time steps of this simulation and use a spatial resolution of 1200×150 to compute pathlines and measure the attributes along them. As demonstrated in Figure 14(a), our framework identifies three regions using the TACs of the Q attribute without significant user intervention: the viscous vortex core where the vorticity is concentrated, the outer layer of the vortices where vorticity diffuses and grows and the region outside of the vortices where the flow is irrotational. In addition, the TACs' visualization in Figure 14(d) informatively characterizes the attribute behaviors within different flow regions. Specifically, the decay of the rotational momentum of the vortex core as expected by the experts is clearly depicted by the monotonic decrease of the orange TAC that illustrates the vortex core behavior. In contrast, the traditional iso-contouring (or iso-surfacing in the space-time domain) has a difficult time to depict this configuration. For instance, Figure 14 (c) shows two iso-surfaces computed with two different Q values. Due to the decrease of Q concentration from left to right in space, the selected thresholds may not lead to iso-surfaces to depict the behavior of vortices in the far right of the flow, whose Q concentration may be similar to other regions without a vortex.

In a detailed study of the above behavior of vortices, we sample 5 pathlines along a vertical line passing the center of a vortex (Figure 14 (a-b). Clearly, we see three different types of TAC behaviors: (1) the decaying of Q concentration over time along the core (the red TAC/pathline); (2) the increasing and shifting of the peak Q values of the TACs corresponding to the pathlines seeded at locations gradually moving away from the vortex core (blue, green, and orange TACs/pathlines);



(a) FTLE field (b) Original vector field

(c) Clustering result and sample pathlines (d) Vortex iso-contours

(e) Sample TACs (f) TAC view

Figure 14: Clustering result of the 2D cylinder flow using TACs of attribute Q reveals a three-layered configuration of the vortex system ($p_c = 100$). (a) FTLE field shows the Lagrangian coherent structures. (b) The LIC texture computed from the original velocity field does not reveal the vortex structures. (c) Pathlines sampled along a vertical line passing the center of a vortex and (e) their corresponding TACs. The TAC (red) of the pathline seeded in the center region decreases monotonically over time, indicating the diffusion of the concentrated vorticity, which gradually increases the vorticity in the outer regions of the shedded vortices (TACs 2-4). TAC 5 corresponds to the pathline seeded outside of vortex region, which exhibits stable characteristic. (d) The two iso-contours (blue and green) with Q values of 17 and 24, respectively, cannot fully capture the vortex configuration. The LIC texture is computed from the velocity in the optimal reference frame (f) TAC profiles of our clustering results.

and (3) a flat TAC/pathline (purple). While types (1) and (3) are easily understandable, the behavior of the TACs in group (2) is interesting. On the one hand, one can see the correlation of the peak locations of these TACs with the changing direction (or turning) of their respective pathlines. This is important, as it associates the geometric characteristics of pathlines with relevant physics. On the other hand, the shifting of the peaks in part indicates the propagation of the rotation momentum outwardly from the vortex core. The increase from negative Q values to positive Q values for the green and orange TACs also associate the shearing layer ($Q < 0$) with its corresponding vortex region. Such a detailed behavior, though known by experts, has not been studied for the flow behind cylinder in the visualization community and cannot be easily obtained with other methods alone.

2D cross section of vortex ring. The next 2D data set simulates a vortex ring hitting a wall with a Reynolds number of 2000. During the interaction, the vortex ring approaches the wall and causes a boundary layer to appear. As the vortex slides against the wall, the boundary layer becomes unstable and is lifted up as a secondary vortex, which in turn lifts up the primary vortex. This data set helps us analyze the role of coherent structures interacting with boundaries, and the generation of turbulence in wall-bounded flows. Our temporal segmentation result using the λ_2 attribute is shown in Figure 15, which demonstrates that our method can detect the moment when the vortex impacts the wall, and automatically generates a temporal cut at that time. The candidate cuts for each TAC are determined based on the extrema of the TAC.

The global clustering results with the Q attribute are shown in Figure 16(a1). The 2D stack color plot in Figure 16(a2) allows us to assess the quality of the clustering results, which is hidden in the edge-bundling visualization. The patterns of TACs in each cluster looks alike. In particular, the red group has stronger positive Q values indicating the vortex core area, while the yellow group corresponds to the particles having negative Q. Also, the peak location of the yellow group indicates the appearance of the shear layer when the secondary vortex is induced and lifted from the wall. To separate the two vortices in this simulation, we apply the temporal clustering to the

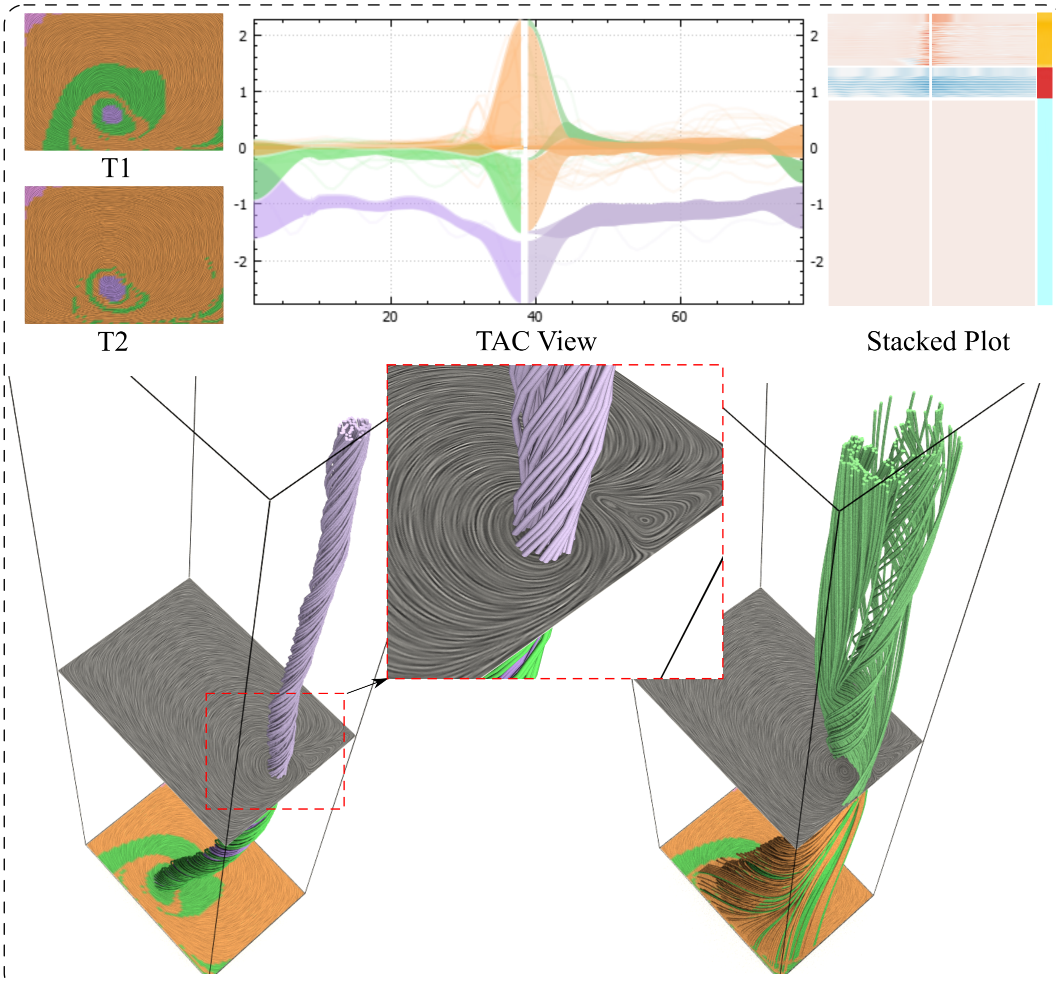


Figure 15: Temporal clustering results of the 2D cross section of vortex ring ($p_c = 1$). TACs are computed by using the λ_2 attribute. The split points for the KDE computation are selected at the extrema values. The temporal cut indicates the moment when the main vortex impacts the wall.

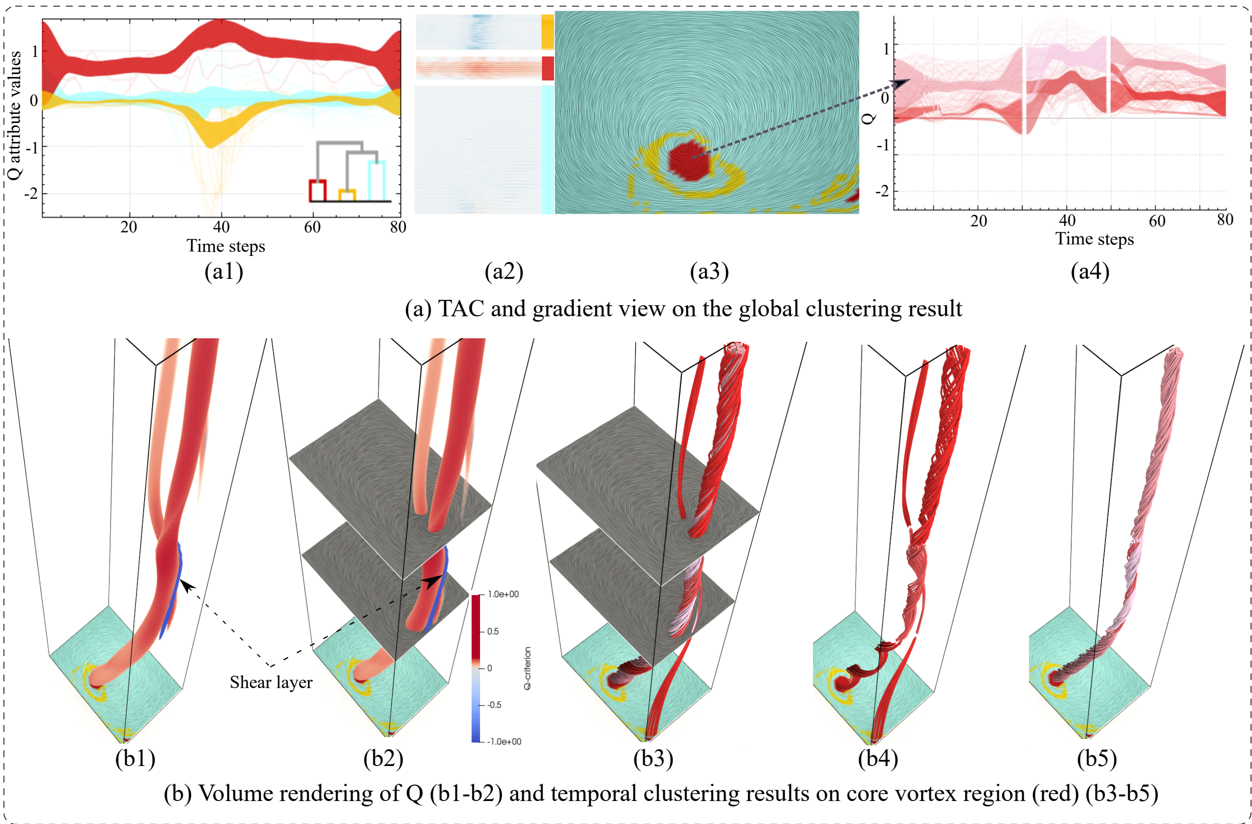
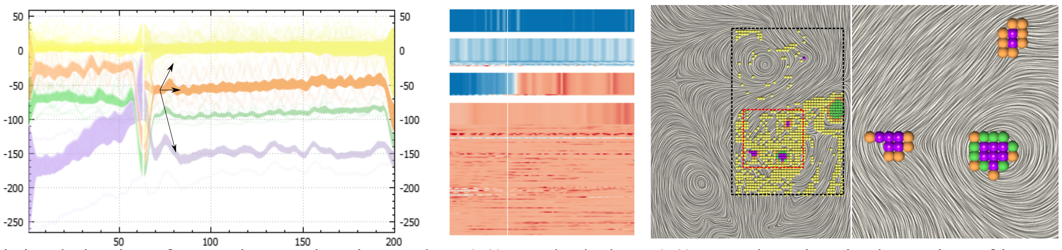


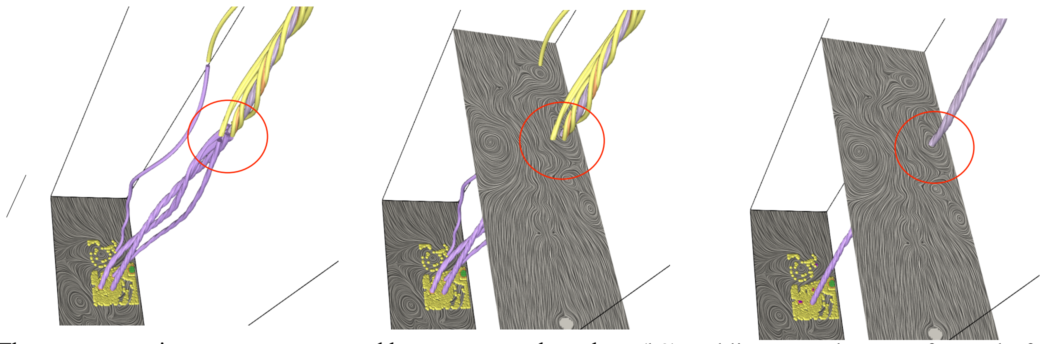
Figure 16: Clustering results of the vortex ring ($p_c = 1$). TACs are computed by using the Q attribute. (a) Global clustering result.(a1) - TAC view, (a2) - Stacked plot (a3) - LIC texture view (a4) - The temporal clustering results on the core vortex region (red). (b1) Original volume rendering of Q. (b2) Volume rendering of Q with two temporal cuts generated by our methods. The second temporal segment aligns with the shear layer when the main vortex collides and sticks with the boundary wall before the secondary vortex is created. (b3-b5) Pathline view of the temporal clustering results computed based on pathlines and their respective TACs in the core vortex region (a4) reveals the behaviors of the two main vortices in the simulation.

dark red group which corresponds to the vortex core region. The results shown in Figure 16(b) illustrate that we can extract the main vortex (pink) and the secondary vortex (red) thanks to the difference in their respective TAC profiles. In comparison with traditional techniques such as volume rendering Figure 16(b1-b2) or iso-surfacing, our method requires fewer selections of thresholds to produce the similar result. In addition, pathlines can provide more detail behaviors of the vortices (e.g., the rotation degree and direction (b4)). It also illustrates the origins of the particles that involve in the generation of the secondary vortex (i.e., the small cluster of pathlines in Figure 16(b3,b4)), which has not been shown previously for this flow. Knowing this is important to understand the dynamics of boundary shear layer of the flow. The three generated temporal segments (of TACs) in Figure 16(a4,b) reveal three main physical events that domain experts care about: before/during/after the main vortex hitting the wall. Here, the temporal cuts were generated based on the split points determined by the derivative curve of the TACs (Figure 9(c-d)). In particular, the event of impact starts with the occurrence of the shear layer (shown in Figure 16(b1-b2) as well as indicated by the yellow cluster in Figure 16(b1)) and ends with this shear layer becomes the secondary vortex.

Boussinesq. Figure 17 shows the temporal clustering results of TACs based on the λ_2 attribute for the Boussinesq flow [30]. This flow has numerous vortices with varying sizes that in part rotate around and collide with each other. The simulation has a dimension of 150×450, and we used 1000 time slices which correspond to one half of the time window. Figure 17(a1) shows two temporal segmentations of four TAC clusters. The multi-layered vortex structure is also revealed here. According to the λ_2-criterion, a region is considered as a vortex if its λ_2 value is smaller than zero [32]. Thus, the green and violet clusters correspond to the core regions of the vortices, while the orange cluster captures the outer vortex regions. The yellow TACs are outside of the vortex region. For the small-scale vortices, the core areas are encoded in the violet cluster because of their smaller negative λ_2 values. The positions of the vortices are highlighted in (a3). Three nearby vortices in the violet clusters coalesce into a new vortex in a later time. Our temporal



(a1) Splitting behavior of TACs in two time intervals (a2) Stacked Plot (a3) Sample points in the region of interest

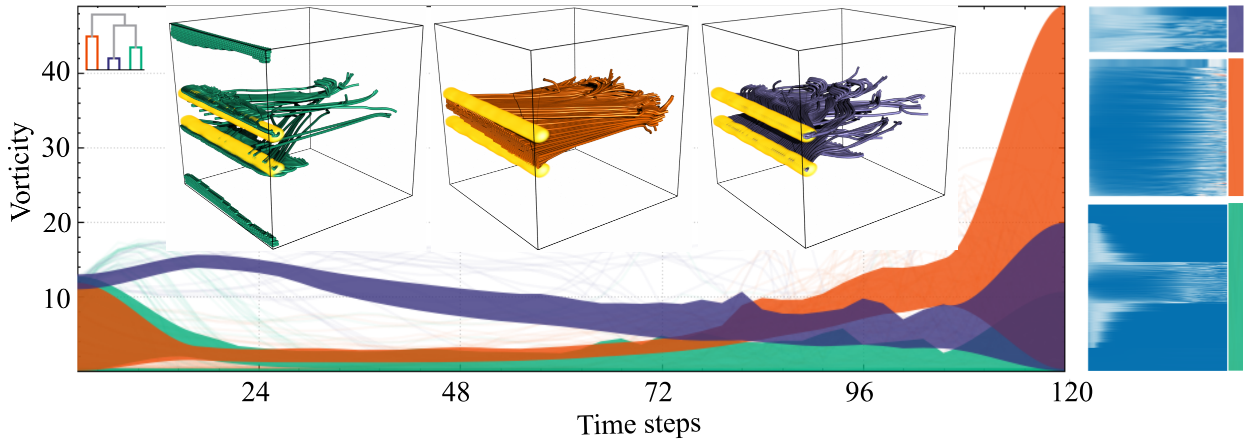


(b1) The vortex merging events are captured by our temporal result (b2) Pathlines near the core of a newly-formed vortex

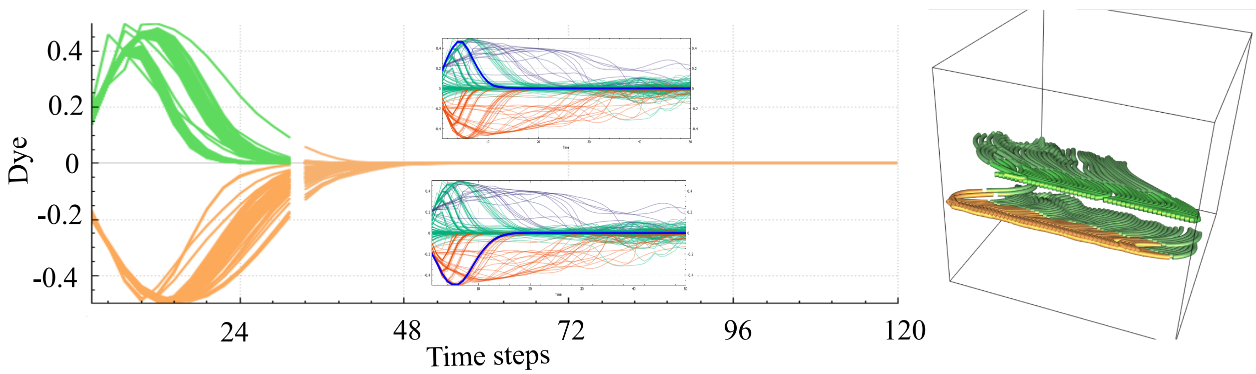
Figure 17: Temporal clustering result of the Boussinesq flow using Lagrangian TACs of the attribute λ_2 (p_c=1). (a1) The TAC view shows the splitting behavior of TACs in the two contiguous time intervals. (a2) The stacked plot. (a3) Sample points from four clusters in the region of interest. The close up view on the right focus on three clusters enclosing the vortex regions. (b1) Pathline visualizations show the vortex merging events. (b2) Pathlines near the core of the newly formed vortex. Colors present different clusters.

segmentation can not only capture the moment the event happens as shown in (b1), but also reveal the physical transportation of particles in the simulation. In particular, the splitting behavior of TACs in (a1) indicates that some particles from the core region (violet) remain in the core, others move to the outer vortex core region (orange) or exit the vortex area (yellow). Such a detailed temporal behavior of small-scale vortices is not easy to obtain with the previous techniques (e.g., geometric-based pathline clustering or thresholding).

3D vortex tube simulation. We also performed experiments using a 3D vortex tube simulation, which simulates two parallel, counter-rotating vortex tubes at a circulation-based flow with $Re = 3500$ and a distance of 2.5 radii apart. The two vortex tubes undergo an elliptical instability [91] that ends with a vortex disintegration. The two vortices interact with each other mainly through the strain produced by the differential velocities induced. The simulation has dimensions



(a) Global clustering result



(b) Pattern search results based on two TAC profiles

Figure 18: (a) Global clustering result of the 3D vortex tube simulation with TACs of the vorticity attribute that reveals a three-layered structure. (b) Pattern search results: Given two groups of TACs with opposite trends, we can find a group of similar TACs and their corresponding pathlines that have symmetric geometric configuration.

of $360 \times 360 \times 360 \times 120$ in a volume of $[0, 2] \times [0, 2] \times [0, 2]$. The total size of this simulation is $36GB$. Again, the pathlines are seeded at the left boundary plane, and vorticity is used here. The results shown in Figure 18 reveal a three-layered structure. The blue cluster includes particles with higher vorticity residing in the two vortex core areas. The green cluster involves particles residing in the outer layer of the two vortices. Their respective TACs are relatively stable. The orange cluster corresponds to particles seeded in the region between two vortices which becomes turbulent at a later time. Accordingly, the vorticity values for this group of particles increase substantially at later times when the flow becomes turbulent.

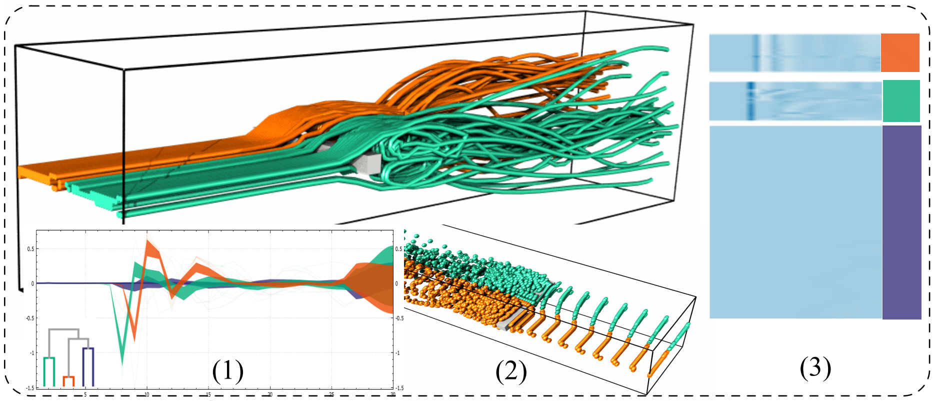
4.3.2 Exploration of Other Flow Features

In addition to the above analysis of the temporal behaviors of vortex structures within various 2D and 3D unsteady flows, we also apply our framework to study other flow behaviors.

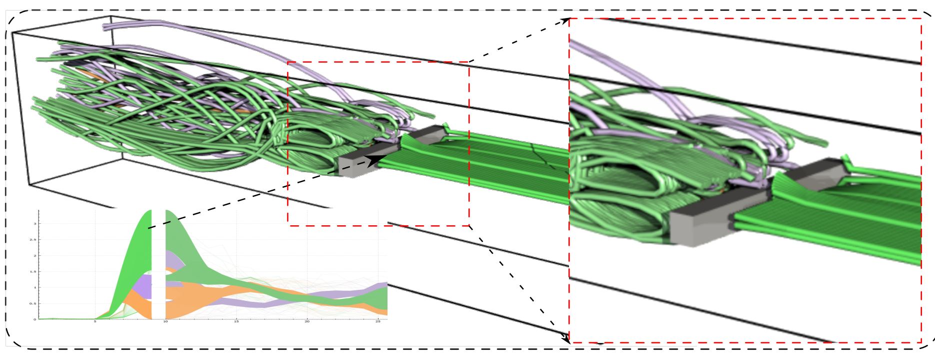
Study subtle differences in axis-symmetric flows. Figure 19 shows the clustering results using TACs of the attribute Q for the 3D simulated flow behind a cylinder [112] with dimensions of $192 \times 64 \times 48$ in the volume of $[-12, 20] \times [-4, 4] \times [0, 6]$ with 101 time steps. Considering the transitional nature of this flow, we select a seeding plane near the left boundary (i.e., $X = -11$) with 64×48 uniform samples. Pathlines that leave the domain earlier are discarded. From the result, we see that the symmetric pathline behavior is captured by the clustering (i.e., the orange and fern clusters in Figure 19(a1-a2)). In addition, their TACs reveal a clear shifting in their temporal trends, indicating that their similar pathline configurations occur at different times. This information cannot be easily obtained by inspecting only pathlines.

In the example shown in Figure 18(b), our domain experts observe two interesting TACs that have opposite trends. They select two sample TACs and perform the pattern search. The results show an intrigued pattern. The two set of TACs are symmetric and the directions of their corresponding pathlines are opposite. Figure 20 demonstrates another use case with the 3D flow behind a cylinder where our users can find a group of particles that move stably and form loops after the flow collides with the cylinder.

Studying separation behaviors in flow. Flow separation in unsteady flows is an important dynamic that experts are interested. Finite Time Lyapunov Exponent (FTLE) [33] is typically used for highlighting the locations where flow separation is the strongest. However, it cannot reveal the cause of the separation. To demonstrate how our framework can help explain this to some extent, we compute the FTLE field on the 2D cross section of vortex ring simulation, and use the obtained FTLE ridge (i.e., with top 20% FTLE values) to select a set of pathlines for study. Specifically, we perform both the global and temporal clustering of this set of pathlines based on their TACs

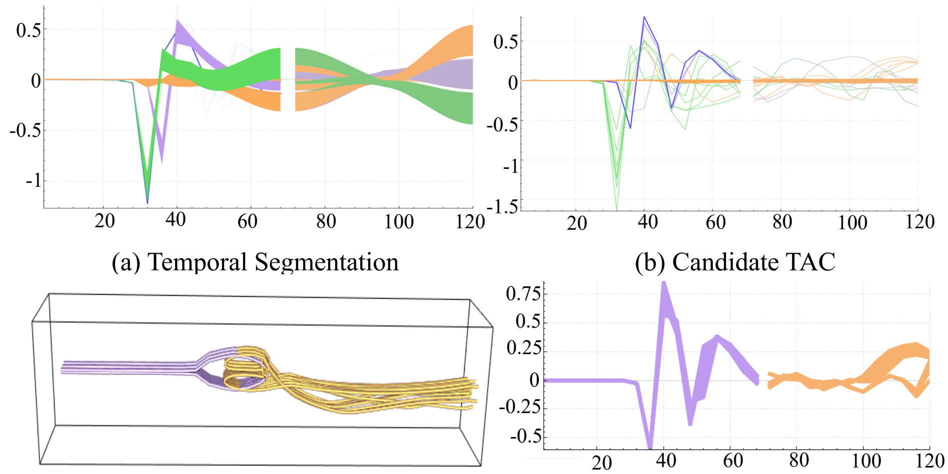


(a)



(b)

Figure 19: Clustering result of the 3D Cylinder flow ($p_c = 1$). (a) Global clustering result with TACs computed from the Q attribute. The TAC profiles corresponding to the two symmetry pathline groups at the center location show a shifting, indicating particles in these two regions exhibit vortex shedding at different times. (b) Temporal clustering result with TACs computed from the local shear rate. The pathline view focuses on the green cluster whose corresponding pathlines have different behaviors in the second time interval (i.e., splitting into three clusters). From the physics point, the vortex regions (i.e., with high positive Q values) usually have little or no shearing flow (i.e., with low local shear rate), and vice versa. This negative correlation can be also observed with our TACs in which the moments that TACs have negative Q are identical to the moments that they have high positive local shear rate. Colors present different clusters.



(a) Temporal Segmentation (b) Candidate TAC

(c) Pathlines (d) Similar TACs

Figure 20: Based on the segmentation result (a), users can choose an interesting TAC (b) and perform the pattern search which results in the similar TACs (d) and their corresponding pathlines(c).

of the local shear rate attribute. Figure 21 shows the results, which highlight three groups of particles having high degrees of separation: the red group indicates the separation between the main and secondary vortices, the light blue particles move to the vortex boundaries, while the yellow particles transport to the physical wall boundary locations. That said, there are at least three different pathline behaviors that cause the same FTLE ridge. Our technique can help reveal this more effectively.

Studying relation between shearing layer and vortex formation. To reveal the relation between shearing layer and vortex formation, we apply the $Q < 0$ criterion on the 3D cylinder flow to select a set of pathlines that have negative Q values in the earlier time. We then perform clustering on these pathlines based on their TACs of the local shear rate attribute. The results are shown in Figure 19(b). As can be seen, there is a strong correlation between the negative Q and local shear rate since the highest shearing values align with the lowest Q values right after the flow encounters the cylinder, where a shearing layer is formed. In a later time, pathlines with strong shear rate participate the vortex shedding formation behind cylinder (i.e., the green and violet groups).

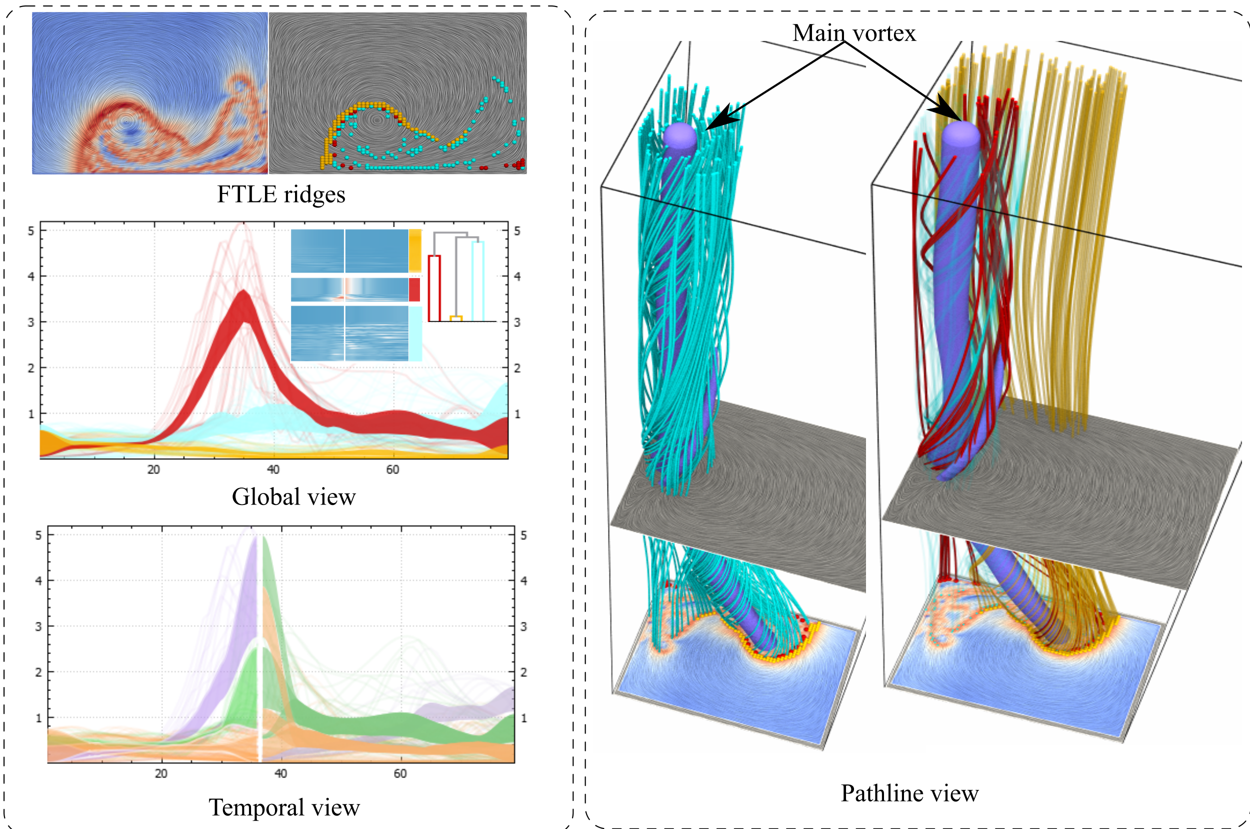


Figure 21: FTLE filtering on the 2D cross section of vortex ring simulation. The seeding points are placed in the FTLE ridges and TACs are generated by using the *shearing* attribute. It can be seen from both global and temporal views that the TAC analysis results based on our clustering method highlight three groups of particles in the high FTLE value region. The yellow particles have high separation degrees near the wall boundary. The light blue group indicates the separation around the boundary of vortices. The red group has large shear values indicating the separation between the main and secondary vortices. Here, the blue iso-surface of the main vortex is provided as a reference. Colors present different groups of TACs and their corresponding pathlines.

4.3.3 Extension to Vector-valued Attributes

Inspired by the vector-based correlation computation [3], we extend our TSM metric to compute the similarity between two TACs with **vector-based** attributes (e.g., velocity vector and acceleration). Specifically, we use the similarity metric between two vectors from the work [3] to compute the correlation between two vector-valued attribute. To compute the distance between two vectors, we use the following approach, i.e., we replace $D_e(\bullet, \bullet)$ in Eq.(10) of the dissertation with the following $d(\mathbf{g_i}, \mathbf{g_j})$.

$$d(\mathbf{g_i}, \mathbf{g_j}) = (2 - d_{\mathbf{d}}(\mathbf{g_i}, \mathbf{g_j}))\, d_{\mathbf{m}}(\mathbf{g_i}, \mathbf{g_j}) - 1 \tag{15}$$

$$d_{\mathbf{d}}(\mathbf{g_i}, \mathbf{g_j}) = \frac{\langle \mathbf{g_i}, \mathbf{g_j} \rangle}{||\mathbf{g_i}||\ ||\mathbf{g_j}||}, \quad d_{\mathbf{m}}(\mathbf{g_i}, \mathbf{g_j}) = 1 + \left| \frac{||\mathbf{g_i}|| - ||\mathbf{g_j}||}{||\mathbf{g}||_{max}} \right|$$

where $||\mathbf{g}||_{max}$ is the maximum magnitude of the vector-valued attribute of interest in the entire data set.

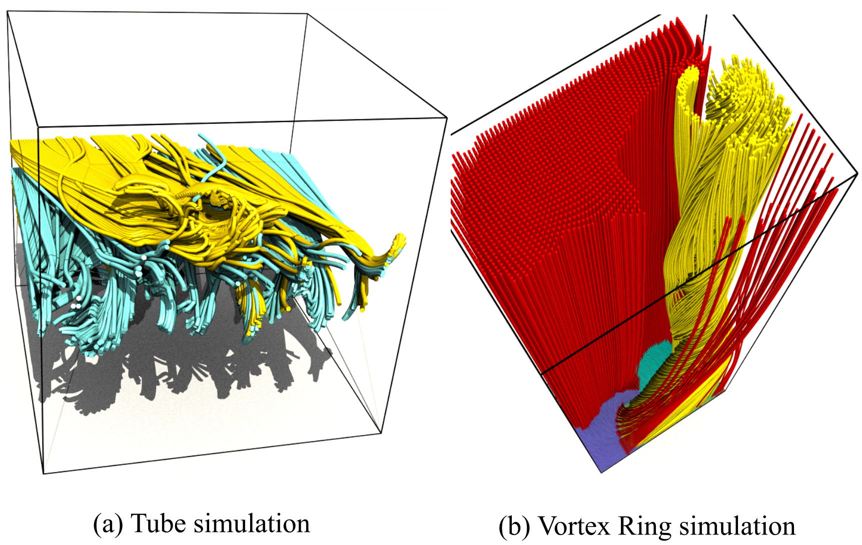


Figure 22: Pathline clustering results with velocity vector TACs for 3D vortex tube (a) and 2D cross section of vortex ring (b) simulations, respectively.

Figure 22 provides the clustering result of pathlines of the 3D vortex tube and 2D cross section of vortex ring simulations using the TACs of velocity vectors. The shapes of pathlines can be distinguished and grouped into different clusters. In particular, for the clustering result in (a),

Table 1: Performance of AHC clustering on four datasets

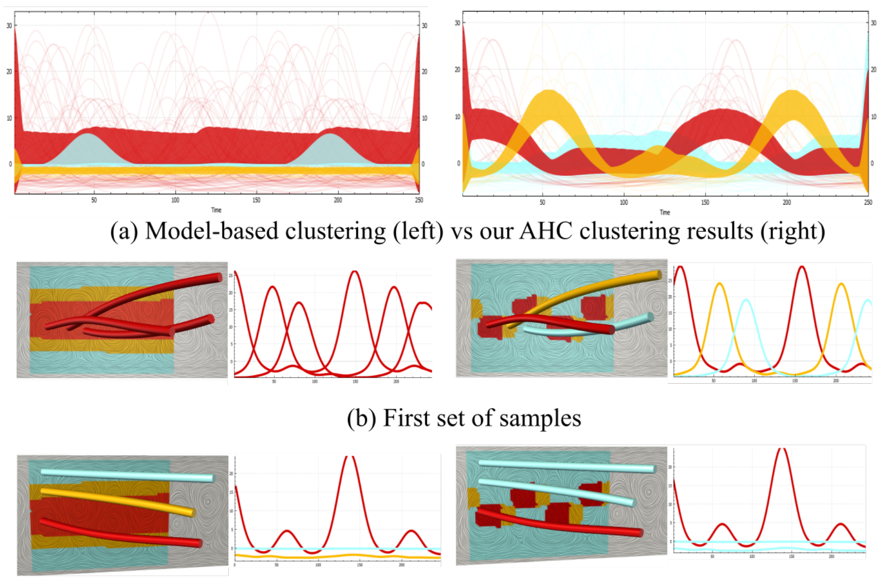
Simulations	**The Number of TAC**	**Time steps**	**Running Time**
Flow behind Cylinder 2D [117]	20000	500	48.6s
Vortex Ring [64]	16384	80	12.4s
Boussinesq 2D [30]	67000	1000	144.8s
Flow behind Cylinder 3D [112]	30720	102	30.5s
Tube 3D [91]	129600	52	182.7s

the pathlines that exhibit upward bending (yellow) during the vortex ring breakdown are roughly separated from those bending downward. For the result in (b), the pathlines in the two clusters are different in term of the representation. The red group corresponds to the ambient area where the vortex does not exist, while the yellow particles belong to the vortex region. Although visualizing the 3D vector-based attribute TACs is not trivial and we plan to address in the future, the results lead us to another utility of TACs which can be used to cluster pathlines.

4.3.4 Performance and Comparison with Other Methods

Performance. All numerical experiments are carried out on a PC with an Intel Core i7-3537U CPU and 128GB RAM with a NVIDIA Quadro 4000 graphic card. The most time consuming task in the system is the AHC clustering. The detailed average running time of AHC clustering on five unsteady flow simulations is reported in Table 1.

Comparison with the model-based clustering. For time-series data (e.g., TACs) clustering, both similarity-based methods and model-based methods are usually applied. AHC is a representative of the similarity-based method in which a customized distance metric needs to be designed to measure the similarity between two TACs. In contrast to similarity-based methods, model-based methods assume that the data are generated by a mixture of underlying probability distributions. The probabilistic representation enables the derivation of consistent expectation-maximization (EM) learning algorithms for the clustering problem, which eliminates the need for a similarity metric. [26]. Here, we compare the clustering results obtained using the proposed AHC



(a) Model-based clustering (left) vs our AHC clustering results (right)

(b) First set of samples

(c) Second set of samples

Figure 23: Comparison between AHC and model-based clustering. Model-based method classifies TACs only based on the magnitude while AHC considers both magnitude and shapes of TACs.

and new metric with those produced by the model-based methods for our TAC data. Figure 23 shows the comparison between the two clustering methods. From the comparison, we see that the model-based method tends to classify TACs only based on the magnitude while AHC consider both magnitude and shapes of TACs. As in our settings both the trend (or shape) and the magnitude of the TACs are important in characterizing their similarity, we opt for AHC over the model-based method.

Other reason for choosing the AHC in our case is due to the performance of the clustering during the user interaction. Although the classic AHC has large computational complexity as the similarity has to be measured between every pair, the performance of model-based clustering method will decrease (i.e it will run slower) if we increase the number of clusters. This is because AHC produces a hierarchy of the clusters, which enables the interactive selection of different numbers of clusters to show. In contrast, model-based method has to recompute the clusters if the number of the clusters is changed during user interaction. In addition, the model-based method requires computation of a probability to assign each TAC to a cluster for each iteration during the expectation-maximization (EM) computation. Therefore, the numbers of computations will increase with the increase of the number of clusters.

One thing we wish to point out is that, in contrast to the model-based method, AHC is sensitive to TACs of varying length, which can have consequences in our clustering. To address this, we only perform clustering on a group of TACs that have similar lengths. Our assumption is that if two TACs have rather different lengths, they do not have similar overall behaviors considering the importance of their life span in unsteady flow. Nonetheless, we can still focus on the portions where the two TACs have similar lengths using the proposed temperal clustering.

4.4 Discussion

In this work, we propose an interactive visualization framework for analysis and exploration of unsteady flow based on TACs. Given a vector field, we first compute the TACs over the entire flow domain and apply time interval segmentation to all TACs. To describe the behavior of a TAC, a sub-TAC extraction method is introduced to identify one or more interesting temporal trends. To measure the similarity of two TACs properly, we introduce a new similarity measure, called the TAC Similarity Measure (TSM) to calculate the dissimilarity of TACs based on their events.

We implement the Agglomerative Hierarchical Clustering algorithm with the new TSM measure for the clustering of Lagrangian TACs on different temporal intervals. The clustering results provide different levels of details for flow behavior in both space and time, which facilitates data exploration. We also improve an edge-bundling technique to better represent the general behavior of TACs in a cluster and the connection of clusters among different time intervals. We introduce a 2D stack plot to visualize the TAC clusters without occlusion. Our framework has been evaluated on multiple unsteady flow simulations, and helps domain experts analyze vortex structures and other flow features.

Limitations. There are a number of limitations of our current system. First, the clustering computation is the most time consuming task as we haven't fully optimized the AHC algorithm. Second, our current TAC-based framework concentrates on scalar attributes. However, it may be extended to other attribute types, such as vector-valued and tensor-valued attributes. The supplemental document provides a couple examples on the extension to vector-valued attributes (e.g., velocity vectors). Nonetheless, the visualization of the clustering results in the TAC space needs to be addressed. Third, the 2D stack plots can provide a summary view of the global clustering results, but they cannot properly visualize the transition between neighboring temporal clusters, which we plan to address. Fourth, our framework has been evaluated via the vortex structure analysis and the exploration of other relevant features such as shearing layers and symmetric behavior, and the flows shown in this work have relatively simple configurations (except for the Boussinesq flow). In the future, it is important to apply our framework to more complex turbulent flows for the study of energy transport to further evaluate it's scalability. Finally, it would be interesting to extend our framework for the clustering of path surfaces to provide a more informative visualization for the study of 3D unsteady flow behaviors. However, to achieve that, effective path surface seeding and placement to achieve sufficient spatial coverage while reducing overlap needs to be addressed, as well as the design of an effective similarity measure for the comparison of two surfaces, which we plan to explore in the future.

5 Multi-scale Coherent Structure Extraction For Taylor-Couette Turbulence

5.1 Problem Overview and Contributions

To overcome the difficulty of separating structures with different scales in turbulent flows, we propose a 3D visualization framework that enables the clear separation of large- and small- scale structures and apply it to the Taylor-Couette Turbulence (TCF). TCF is the turbulent fluid motion created between two concentric and independently rotating cylinders. It has been heavily researched in fluid mechanics thanks to the various nonlinear dynamical phenomena that are exhibited in the flow. As many dense coherent structures overlap each other in TCF, it is challenging to isolate and visualize them, especially when the cylinder rotation ratio is changing. To identify the different scale structures in TCF, domain experts traditionally rely on the rudimentary visualization techniques (e.g., volume rendering, iso-surfacing, and thresholding) based on the physical attributes, such as vorticity or Q-criterion [32]. As illustrated in Figure 3, different parameter settings (here, different threshold values for Q) may lead to structures with significantly different shapes and density, making the identification and tracking of the desired structures difficult. Making this worse is that attributes such as vorticity and Q-criterion in small structures have different value ranges from those in large structures as reported previously [110]. Often, high positive Q-criterion corresponds to vortices with small-scale structures due to its definition, i.e., it is based on vorticity and strain which are inherently local and small-scale quantities, while small Q values may exist both inside and outside the large structures. Furthermore, as shown later (Figure26), one single attribute with one single threshold value is not sufficient to clearly separate multi-scale coherent structures, especially when the cylinder rotation ratio changes. All these challenges force the fluid mechanic researchers to resort to the visualization of 2D cross section (Figure 2 (a)) through cylinders for TCF analysis, losing important 3D configuration information of the structures. To the best of our knowledge, no existing works provide an effective 3D visualization of TCF that separates the large- and small-scale structures to aid the study of the re-appearance of certain coherent structures (especially, Taylor vortices) in TCF under different rotation ratios. To overcome the limitation of the separation of structures in different scales with a single attribute, our framework employs multiple attributes for feature extraction and separation. This is also supported by a recent work on 2D TCF [81] that shows that streamwise velocity aligns reasonably well with the position of the Taylor roll in the flow domain. By extending the feature level-set method [40] to take into account values in a range, we can combine the characteristics of multiple attributes in one analysis to increase the difference between the small- and large- scale structures, making their separation easier and more accurate. To extract a 3D surface representation for the coherent structures, we construct iso-surfaces from the kernel density estimate of the distance field obtained from the feature level-set. Considering the characteristic that TCF with certain rotation ratio may be depicted by its 2D cross sections, referred to as the **WS planes** formed by the wall-normal and spanwise directions (Figure 2 (b)(c)), we compute the 2D summary configuration of vortices in TCF by projecting the 3D information onto the **WS** plane. Compared to the visualizations generated with the conventional approaches in both 2D and 3D, our method produces a much cleaner visualization. Our framework is implemented in CUDA, making it suitable for the efficient analysis of TCF.

Our framework has been applied to three different TCFs simulated with different control parameters. These three TCFs have different levels of turbulence, making the separation of large-scale structures from the small ones extremely challenging. Our framework successfully reveals the configurations of large-scale coherent structures in all three TCFs. This is the first time the large-scale structures in TCFs are visualized in 3D. From these new visualizations, domain experts not only can see a clear separation of large- and small- scale structures, but they also can see the 3D configuration of large-scale structures in TCFs with low cylinder rotation ratios to better understand the procedure of the formation of Taylor rolls (or Taylor vortices). The efficient computation of our framework will enable the study of the impact of the rotation ratio to the 3D configuration of Taylor rolls in a finer scale and set up the foundation for the study of the time-dependent behavior of TCFs in the future.

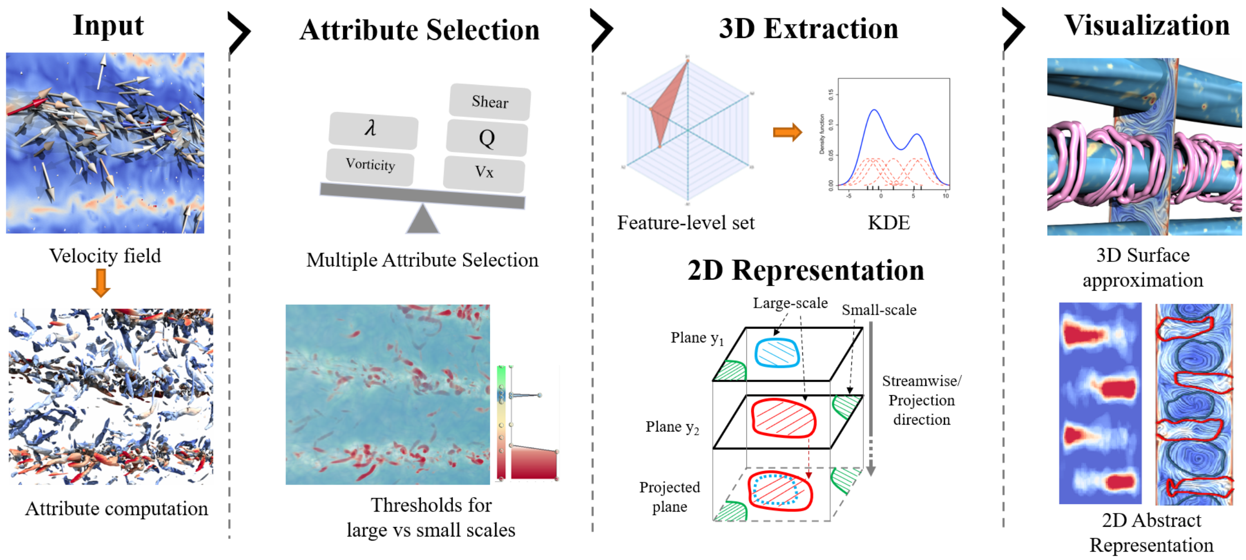


Figure 24: The pipeline of our framework.

5.2 Proposed Pipeline

As stated earlier, our goal is to separate large-scale vortex structure from the small-scale ones and to provide an intuitive and interactive visual representation so that fluid mechanics experts can quickly evaluate their hypotheses.

To achieve this goal, we propose a pipeline as illustrated in Figure 24. We first derive multiple attributes from the input velocity field. By analyzing the characteristics of the attributes in each cylinder rotation ratio and with knowledge from the expert, we select attributes and their corresponding value ranges that can partially indicate the existence and location of large- and small-scale structures (Section 5.2.1). **To extract and visualize the structures in 3D**, we adapt the recently introduced feature level-set [40] to compute a distance field to combine the selected attributes and their corresponding value ranges to better locate and separate large- and small-scale structures (Section 5.2.2). To achieve a smooth representation of the shape of the large- or small- scale structures, we further apply a kernel density estimation on the regions extracted from the obtained distance field (Section 5.2.3), from which a surface that approximates the geometric configuration of the large- or small- scale structure is extracted for visualization. Attribute information can be visualized on the extracted surface to provide additional information of the structure.

To provide a summary view of the TCF in 2D, we project the areas in the individual 2D cross sections (parallel to the WS plane) with the attribute values falling within the respective selected value ranges onto the first cross section. This process is illustrated in Figure 24 under **2D Representation**. This aggregated information is color coded on top of the LIC texture of the 2D flow in the first cross section. An example of this summary 2D visualization is shown in Figure 28 (e) and (f).

5.2.1 Attribute Selection

Physical attributes for the region-based vortex extraction. Given a steady vector field $\mathbf{v}$, its spatial gradient $\nabla_{\mathbf{x}}\mathbf{v}$ is referred to as its *Jacobian*, denoted by $\mathbf{J}$. $\mathbf{J}$ can be decomposed as $\mathbf{J} = \mathbf{S} + \mathbf{R}$, where $\mathbf{S} = \frac{1}{2}[\mathbf{J} + (\mathbf{J})^\top]$ and $\mathbf{R} = \frac{1}{2}[\mathbf{J} - (\mathbf{J})^\top]$ are the symmetric and anti-symmetric components of $\mathbf{J}$, respectively. In addition to the velocity magnitude and the individual (u, v, and w) components of the velocity field, a number of flow attributes related to coherent structures, especially vortices, can be derived from $\mathbf{v}$, $\mathbf{J}$, $\mathbf{S}$ and $\mathbf{R}$ [73], such as:

- vorticity magnitude, $||\nabla \times \mathbf{v}||$.
- λ_2, computed as the second largest eigenvalue of the tensor $\mathbf{S}^2 + \mathbf{R}^2$ [41].
- $Q = \frac{1}{2}(\|\mathbf{R}\|^2 - \|\mathbf{S}\|^2)$.
- local shear rate, defined as the Frobenius norm of $\mathbf{S}$.

These scalar attributes can help to identify regions of vortical behavior. For example, if the Euclidean norm of the vorticity tensor $\mathbf{R}$ is greater than the magnitude of strain rate tensor $\mathbf{S}$, then the region contains a vortex. This criterion is equivalent with the condition $Q > 0$. The main drawback of the region-based methods is that the local attributes with a single threshold value tend to focus on small-scale features. Thus, the conventional thresholding approach is not sufficient for the extraction of large-scale structures. Nonetheless, for TCF, structures with different scales may be associated with certain value ranges of some attributes. By carefully selecting these ranges of certain attributes, small- and large-scale structures may be better separated for the subsequent processing.

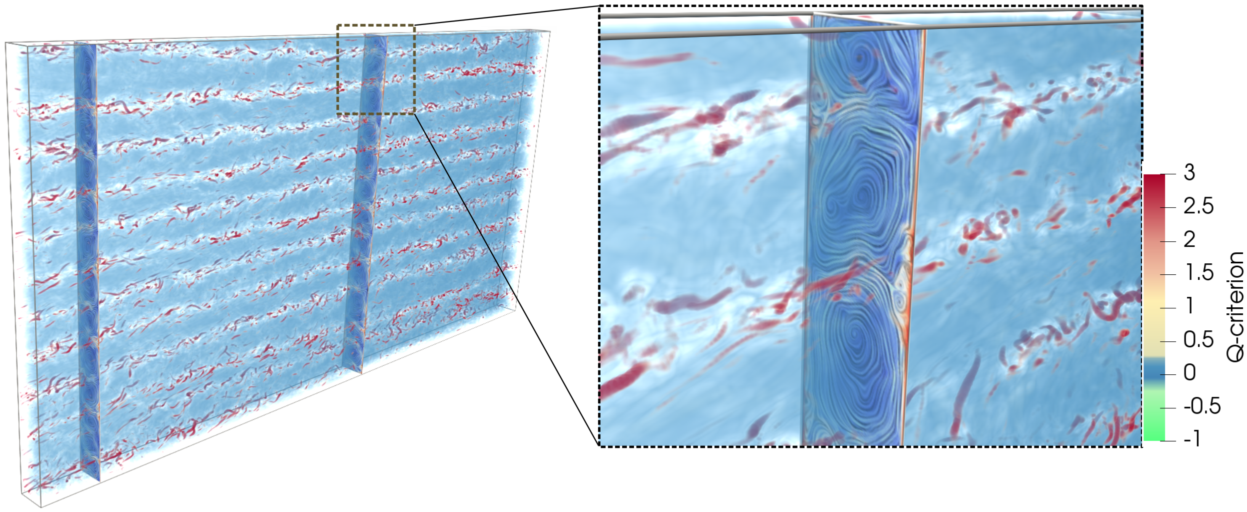


Figure 25: Volume rendering of Q-criterion corresponding to the large-scale structures (blue with $Q \in [0, 0.1]$) and small-scale structures (red with $Q > 2$), respectively.

In our pipeline, domain experts can select a single attribute, and then analyze it through statistical-based methods [46] or with the aid of volume rendering to obtain the desired value ranges which potentially reveal the location and shape of the coherent structures. As illustrated in Figure 3, small vortices can be extracted with $Q > 2$ for the simulation setting $R_\omega = 0.1$. In contrast, the large-scale structures can be observed via the volume rendering of Q with values $\in [0, 0.1]$ (Figure 25). Interestingly, for TCF, we found a correlation between the rotation ratio R_ω and the value range of Q. In particular, we can use a value range of $[0, R_\omega]$ for Q to roughly reveal the large-scale structure. This observation has also been previously confirmed [110].

In addition to the value selection for a single attribute, the experts can also select multiple attributes to better extract large-scale structures. The motivation for combining different attributes comes from the issue of using a single attribute like Q, which may only perform well in the TCF with a high cylinder rotation ratio R_ω. With lower rotation ratio, Q produces noisy data, making the extraction more difficult. As illustrated in Figure 26, Q is noisier than the streamwise velocity and $Shear$ in the wall-normal and streamwise direction for $R_\omega = 0.05$. Either the streamwise velocity or $Shear$ along is NOT sufficient in both directions; but by combining the features revealed by both the streamwise velocity and $Shear$, we can fully capture the prominent structures in the 3D

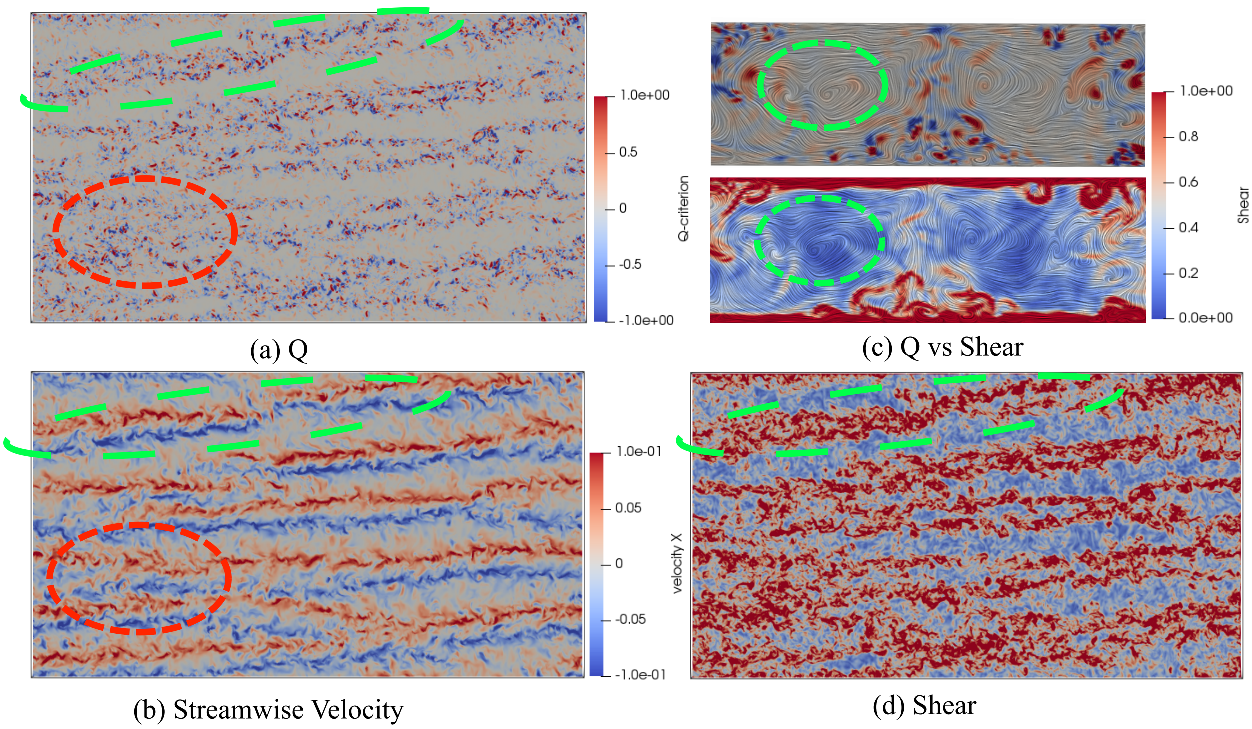


Figure 26: Visualization of (a) Q, (b) Streamwise velocity, (d) *Shear* in 2D cuts along the wall-normal direction, and (c) in the streamwise direction with $R_\omega = 0.05$. Q is noisier than the streamwise velocity and *Shear* in the wall-normal and streamwise directions and not able to separate structures with different scales here. By combining the streamwise velocity and *Shear*, we can better capture the prominent structures in both wall-normal and streamwise direction.

configuration (Figure 31).

Even though we can see the assembling of the structures in the volume rendering after selecting the proper value ranges of certain attributes, as illustrated in Figure 25, effectively isolating and extracting them is not trivial. Next, we will describe how we address this challenge with the help of feature level-set extraction.

5.2.2 Feature Level-Set Extension

To extract 3D coherent structures, we adapt the feature level-set method to combine the characteristics of the selected attributes. Feature level-set is the generalization of the concept of level-set (i.e., iso-surface) from uni-variate to multi-variate data. It is superior to the method proposed by Schneider et al. [95] in which the authors try to find features in the spatial overlap of iso-surfaces extracted from two attributes. That method does not work if there is no intersection between the iso-surfaces. Feature level-set does not suffer from the issue. In the original version, feature level-set only works with a single (threshold) value for each attribute. We modify the distance metric so it can work with values within a range. Assume that we have N attributes $\mathcal{A} = \{a_1, a_2, \cdots, a_N\}$ for each point $\mathbf{p}$ in the spatial domain. We denote $\{\{v_{i,1}, v_{i,2}\} \mid v_{i,1} \leq v_{i,2}, i \in [1, N]\}$ as the selected value ranges for either small-scale or large-scale features for these attributes. The scalar distance or level to each point p is defined as follows:

$$d_{\mathbf{p}} = \begin{cases} 0, & \text{if } \exists i,\ a_i \in [v_{i,1}, v_{i,2}] \\ \\ \min\limits_{i \in [1,N]} \min \left\{ \|a_i - v_{i,1}\|, \|a_i - v_{i,2}\| \right\}, & \text{otherwise.} \end{cases} \tag{16}$$

Given the attribute thresholds for small- or large- scale features, the feature level-set outputs a distance field in which the regions with smaller distance value belong to the small- or large- scale structures. To increase the smoothness of the distance function, we attempted to replace the Frobenius norm with the sigmoid function. However, the result did not improve significantly. Frobenius norm is ultimately selected because of its high accuracy and low computational complexity. It is important to note that we normalize the attribute values before computing the distance field. Thus, all attributes are in the same range, making the distance comparison feasible.

5.2.3 Kernel Density Estimation

Once we compute the distance field, the straightforward method to extract the surface representation for coherent structures is to apply iso-surfacing with a small distance threshold. In practice, however, the distance field is not always smooth as shown in Figure 27(b), leading to the disconnected and noisy iso-surfaces. To overcome this, we propose to group the (likely disconnected) regions that correspond to the large- (or small-) scale structures based on their spatial proximity, then provide the abstract visualization for each structure. To characterize their spatial distribution, we apply the kernel density estimation (KDE) to map the distance values into a density field in

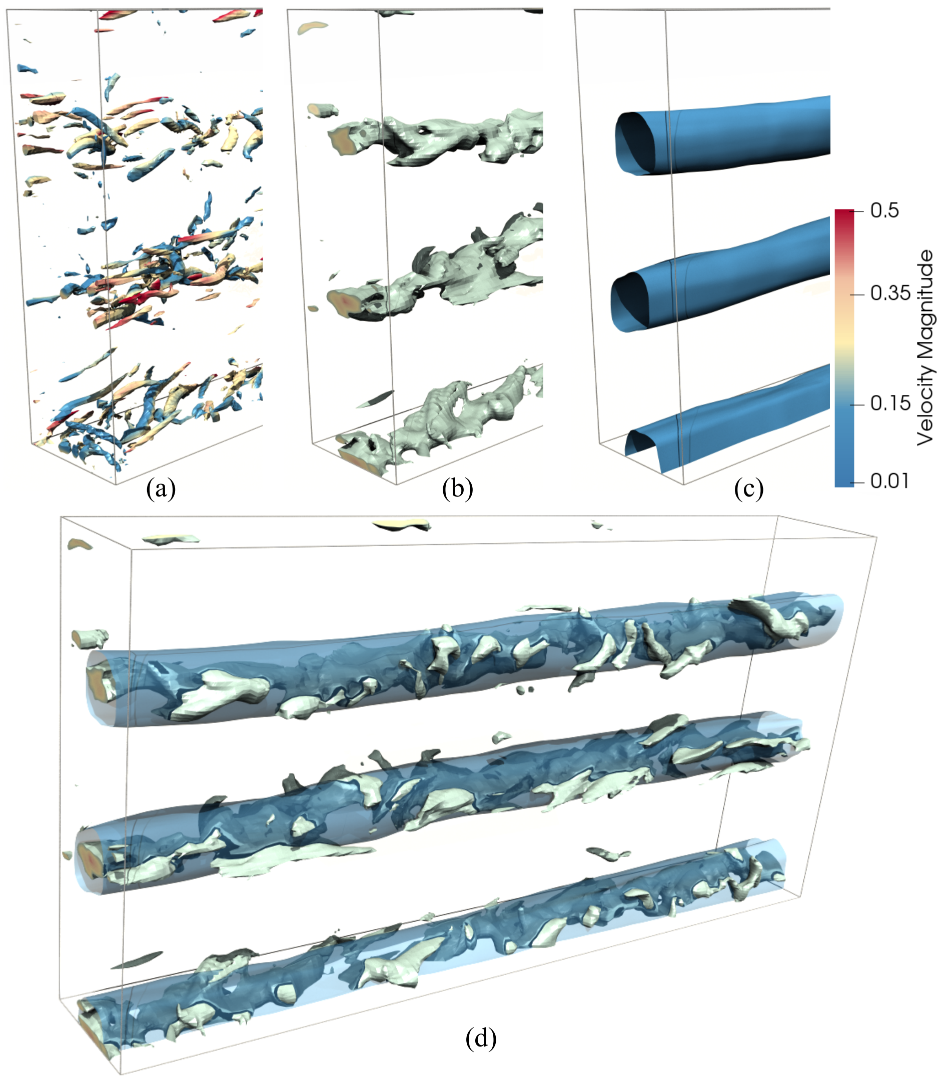


Figure 27: Motivation for the kernel density estimation. (a) Original iso-surfaces using Q-criterion lead to visual clutter. (b) The extracted iso-surfaces from the distance field computed with the feature level-set. (c) Iso-surfaces extracted from the density field provide cleaner visualization. $\sigma = 0.56$, $h = 0.6$, $m = 8$. (d) The combined visualization of (b) and (c).

which higher density values correspond to the area containing more structures with similar scale. KDE was first proposed by [69], and it is a well-established method to achieve a non-parametric estimation for spatial density. We approximate the density function using a Gaussian kernel density estimate similar to [63] in a cube-like neighborhood area P centered at x with size $n = m^3$:

$$f(x) = \frac{1}{n(h\sqrt{2}\pi)^d} \sum_{i=1}^{n} exp(-\frac{||x - p_i||^2}{2h^2}) \times w \tag{17}$$

with $||x - p_i||^2$ being the Euclidean distance between grid points, h being the bandwidth of KDE, d being the dimension of the grid. We utilize the Silverman method [95] to compute the optimal bandwidth which gives us $h = 1.06\sigma N^{-\frac{1}{5}}$ with σ the standard deviation and N the number of grid points. Note that we assume the true distribution of the data is Gaussian. Compared to the conventional KDE, we add a weighing term, w, which sets the density values to zero for the grid points not enclosed by a large- (or small-) scale structure. A point belongs to a coherent structure if the distance value is zero. Thus, the value of the weight term w is equal to 1 if the distance value is zero, otherwise $w = 0$. We do not set w to the actual values of distance values because the purpose of KDE in this step is to reveal how dense the small distance values are in a certain neighborhood area. Based on the computed density field, iso-surfaces are extracted to provide a visual representation of large- (or small-) scale structure groups. As can be seen in Figure 27, we can derive three main groups for a sub-volume of the flow, which are represented by three large iso-surfaces (blue) in (c) from eighty nine small regions in (a). The density iso-surfaces provide a cleaner visualization.

5.2.4 Composite Visualization Generation

Selecting the threshold for KDE iso-surface is a trial and error process. Based on our experiments, however, we observe that setting the threshold to half of the density value range can produce a reasonable result. If the density value at a grid point is greater than the half of the value range, it means that the majority of its neighbors belong to the desired coherent structures. Half of the density value range is set by default for all of our results. The color of KDE iso-surfaces is mapped to one of the selected physical attributes in the feature level-set extraction step. A smooth color surface indicates a better physical alignment of the obtained surfaces.

Streamlines can be used to depict the flow motion, and verify the correctness of the extracted coherent structures as they should wrap around the structures in the simulation with high R_ω values. To generate a small set of such streamlines, we use a straight seeding rake that is close and parallel to the surface of the large (or small) scale structures. Some sample streamlines are shown in Figure 28(b)(d) and Figure 31. Combined with the extracted boundary surfaces, they provide a more informative visualization.

As described in Section 2.2, the TC simulation is carried out in Cartesian coordinates where the cylindrical streamwise coordinate is unwrapped onto a straight line. To provide a more intuitive visualization, we wrap back the streamwise dimension from the straight to the cylindrical setting. Coordinate transformation is a widely solved math problem. Assume R_1 is the radius of the inner cylinder. We denote NX as the number of points in the streamwise direction, s_x, s_y, s_z are the spacings between two neighboring points along the streamwise, wall-normal, and spanwise directions, respectively. For each grid point (i,j,k) in the regular grid in the Cartesian space, we can derive the corresponding point (x,y,z) in the cylindrical coordinate as follows:

$$\begin{cases} x &= r\cos(j * 2\pi/NX) \\ y &= r\sin(j * 2\pi/NX) \\ z &= ks_z \end{cases} \tag{18}$$

where $r = R_1 + i * s_y$. R_1 is a user-specified value and can be obtained from the simulation spatial information. In our results, $R_1 = 4$.

5.3 Results

We have applied our analysis and visualization framework to Taylor-Couette flows simulated with three different cylinder rotation ratios, $R_\Omega = 0.1, 0.05, 0$ (Section 2.2), respectively. The velocity fields of these three simulations are stored in the VTK binary format with the size of 12GB each. In this section, we first demonstrate how the proposed framework can produce a first effective 3D visualization of TCF to depict the 3D behaviors of TCF. We then discuss how our 3D visualization helps domain experts analyze the 3D behaviors of TCFs with low cylinder rotation ratios (i.e., $R = 0.05$ and $R = 0$) that was not possible with the traditional 2D visualizations. We compare our method with the most widely used approach in multi-scale structure extraction – the convolution kernel filter, and point out a potential combination of the convolution kernel filter with the proposed method. Last but not least, we report the performance of our implementation.

TCF with $R_\Omega = 0.1$. Figure 28 provides the visualization of the TCF with $R = 0.1$. It is known by the expert that the large-scale Taylor rolls are fully formed in this simulation. These large-scale Taylor rolls can be depicted by their corresponding Taylor vortices with opposite rotation in the 2D cross section.

Figure 28(a) shows the combined visualization of the extracted small-scale vortices (the red iso-surfaces) and large-scale structures (the blue tubes). The closeup view is presented in Figure 28(c). From this overview visualization, the expert can easily see the configuration of the small- and large-scale structures that form interleaving spatial layout along spanwise direction. This is verified by the conventional 2D visualization (e.g., the LIC texture shown in Figure 28(f)). The feature level-set is computed by using the two attributes Q and the velocity magnitude in which $Q \in [0, 0.1]$, and the velocity magnitude $\in [0.05, 0.1]$.

Another prominent characteristic revealed in this visualization is that the large-scale structures (blue surfaces) are mostly following the streamwise direction (i.e., their orientation is aligned with the horizontal boundary). The red arrow indicates the streamwise direction in Figure 28(b). This characteristic is known as the *streamwise invariant* property of Taylor rolls, which means that the

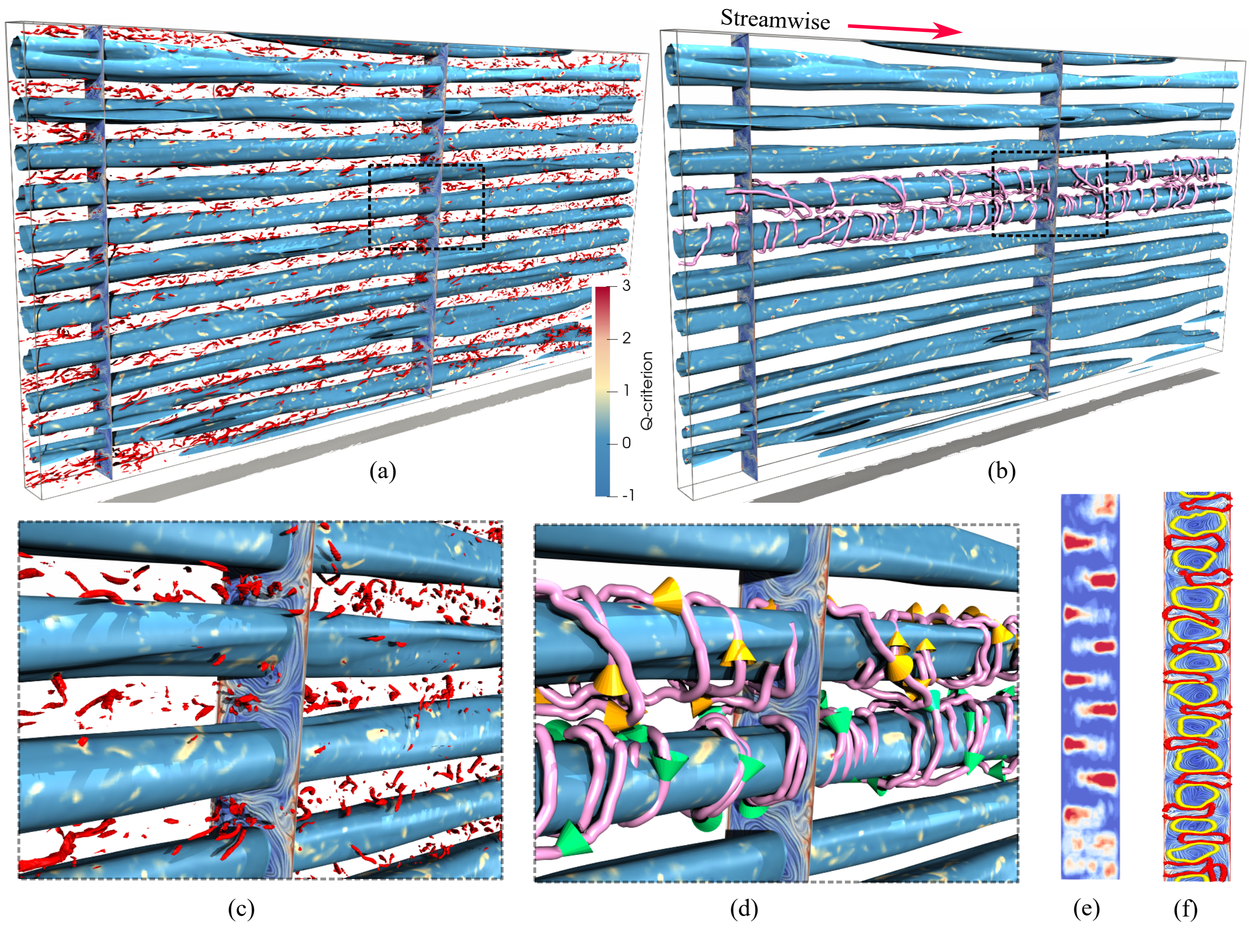


Figure 28: Visualization of the large-scale structures for TCF with $R = 0.1$. (a) shows the combined view of the small-scale vortices extracted with the threshold $Q > 2$ (red) and the large-scale structures (blue) using the proposed method. (b) visualizes the large-scale structures that are alike Taylor rolls with streamlines swirling around them. The surface color is mapped to Q. The red arrow indicates the streamwise direction. (c)(d) provide the close-up views of (a) and (b), respectively. (e) A 2D abstract representation focuses on the high concentration of small-scale structures (red). (f) shows a **WS** plane with the LIC texture and the corresponding boundaries of the small- (red) and large- (yellow) scale structures.

behavior of the flow on a cross section perpendicular to the streamwise direction is almost identical and independent of the location of the cross sectional cut along the streamwise direction. This is the reason why domain experts often use 2D visualization to study TCF with large cylinder rotation ratio. Nonetheless, 2D visualization emphasizes the large Taylor roll vortices as they are visually dominant in the LIC texture and may not capture the small-scale vortices that do not intersect with the 2D cross section.

To verify that the large-scale features extracted with our approach approximate the Taylor rolls in the TCF, we compute a number of streamlines along these features (Figure 28(b)). We see that these streamlines are swirling around the respective surfaces, indicating the rotation motion of the flow. The rotation bundles formed by these streamlines mostly match the geometry of the extracted surfaces. Figure 28(d) provides a closeup look. In addition to streamlines, we also color the extracted surfaces using the Q attribute measured on them. Our assumption is that if the color is close to constant on the surface, the surface is aligned well with the corresponding attribute. As can be seen in the provided visualization, most parts of the surfaces have similar colors except for a few spots, indicating a good alignment with the Q attribute. In summary, our visualization demonstrates that the extracted surfaces do approximate the Taylor rolls.

When compared to the 3D visualization using simple thresholding shown in Figure 3, our visualization provides a much cleaner separation of the regions dominated by the small-scale and large-scale features, respectively. It also enables the depiction of the large-scale flow motion, which is not easy to capture with the local physical attributes (i.e., vorticity, Q, and/or λ_2). This is because this large-scale motion corresponds to low values of those physical attributes. Our method can successfully identify them.

Figure 28(e) shows a density field which is computed based on the projection of small-scale structures onto a 2D cross section of the streamwise axis (i.e., parallel to the WS plane). N_x cross sections uniformly distributed in the streamwise direction are used ($N_x = 512$ in our experiments). The places with high concentration of the density field (i.e., corresponding to regions where more small-scale features reside) coincide with the places with high velocity magnitude (f). Note that this density field visualization indicates that the majority of the small-scale structures stays close to the boundary and form a clear periodic spatial distribution layout, indicating the formation of the Taylor rolls/vortices. Such a summary view is not easily obtained with the conventional approach.

With the separation of the regions with small-scale and large-scale structures, the expert can now study their different behavior separately. While Figure 28(b) focuses on large-scale structures, Figure 29 reveals the detailed behavior of the small-scale structures. The red iso-surfaces represent the small vortices extracted by using the threshold $Q > 2$. The gap between the 3D transparent surfaces that enclose all small vortices are the places where 3D Taylor rolls reside.

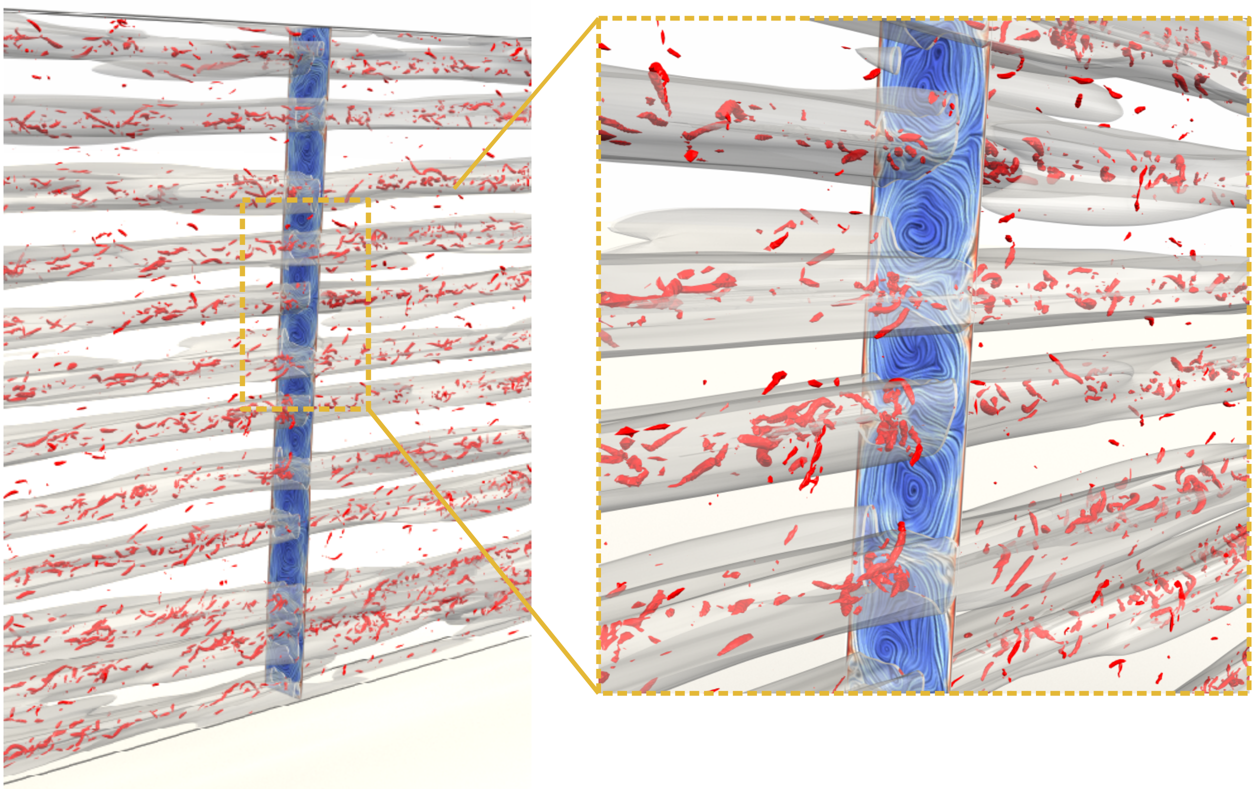
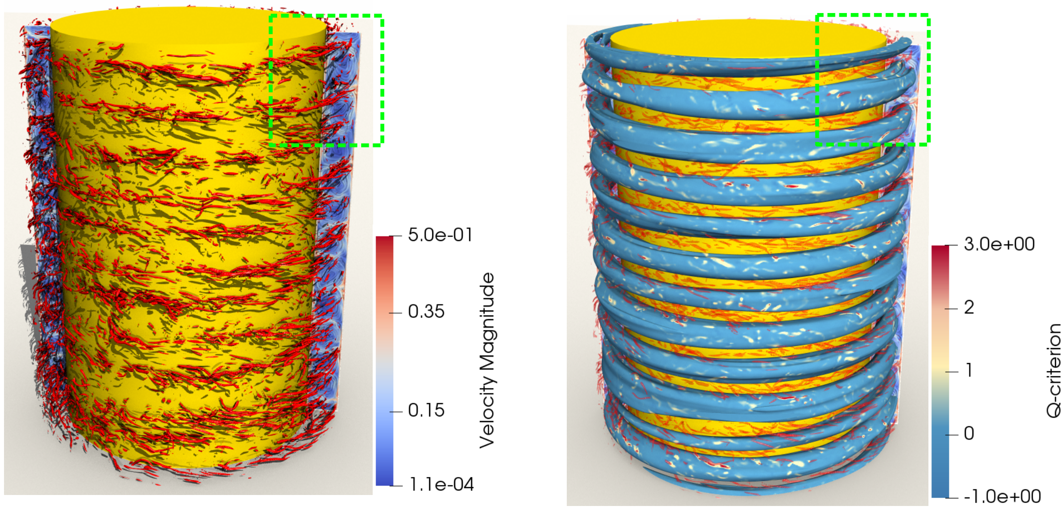
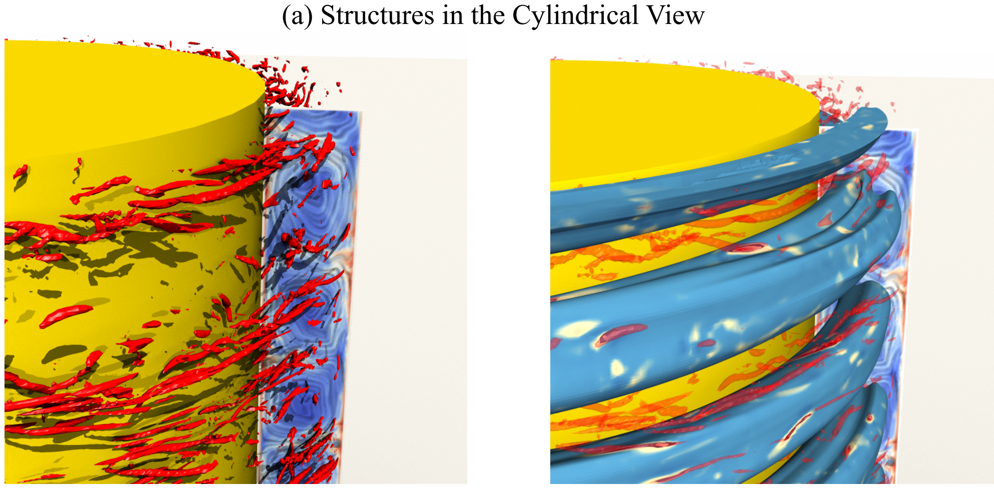


Figure 29: Visualization of the regions (enclosed by transparent surfaces) with small-scale structures not belonging to the Taylor rolls for $R_\omega = 0.1$.

To enable an intuitive understanding of TCF in the real-world setting (Figure 2), we transform the extracted features with our method to a cylindrical coordinate system. Figure 30 provides such a visualization. Specifically, Figure 30 (a) provides the entire flow in the cylindrical view for both



(a) Structures in the Cylindrical View



(b) Close-up view

Figure 30: (a) Visualization of the small-scale (left) and large-scale (right) features in TCF with $R = 0.1$ in the cylindrical coordinate reveals the well-known Taylor roll structure. (b) provides the close-up view for the small- (left) and large-scale (right) structures, respectively.

small-scale (left) and large-scale (right) features. Figure 30 (b) provides a close-up view.

TCF with $R_\Omega = 0.05$. Next, we apply our framework to a TCF with $R_\Omega = 0.05$ (i.e., half of the rotation ratio used to generate the above TCF). This TCF is less ordered than the one seen above with $R_\Omega = 0.1$. As shown in Figure 26 (a), Q attribute is not sufficient in capturing the large-scale structures in this case. By examining other attributes, we found that both local shear rate and the streamwise velocity can reveal a cleaner configuration of the large-scale structure (Figure 26 (b)(d)). Therefore, we use these two attributes in the feature level-set computation. The value ranges for these two attributes are [0.05, 0.2] for the streamwise velocity and [0.8, 2] for shear, respectively.

Figure 31 (b) shows the overview of the flow with the red small vortices identified with $Q = 2$. Compared to the TCF with $R_\Omega = 0.1$ (Figure 3), these small-scale vortices are less structured (in other words, more chaotic). Figure 31 (a)(c) show the visualization of this TCF using our framework. As can be seen, the identified approximate boundary surfaces of the large structures (curious blue) are not fully connected. This indicates the insufficient separation of the large-scale structures from the small-scale vortices, suggesting that the Taylor roll is not fully formed. This is better conveyed in the projected density field of the small-scale features shown in Figure 31 (g)(h), in which the high concentration of small-scale features (in red) are less structured and rather noisy. Nonetheless, these large structures still resemble Taylor rolls seen in the TCF with $R_\Omega = 0.1$. This is indicated by the seeded streamlines (Figure 31(c)). These streamlines still warp around the extracted surfaces, though not well-organized. Figure 31(f) offers a closeup look at these streamlines. Again, we color the extracted surface using one of the selected attributes (shear in this case). Based on the color distribution shown on the surface, we can see that the geometry of the surfaces aligns well with the local shear rate of the flow.

TCF with $R_\Omega = 0.0$. Finally, we apply our method to the TCF simulated with $R_\Omega = 0.0$. This is the most chaotic scenario in which the rotation is minimum. As shown in Figure 32(a), the small-scale vortices extracted using Q criterion cover almost the entire domain, making the separation of

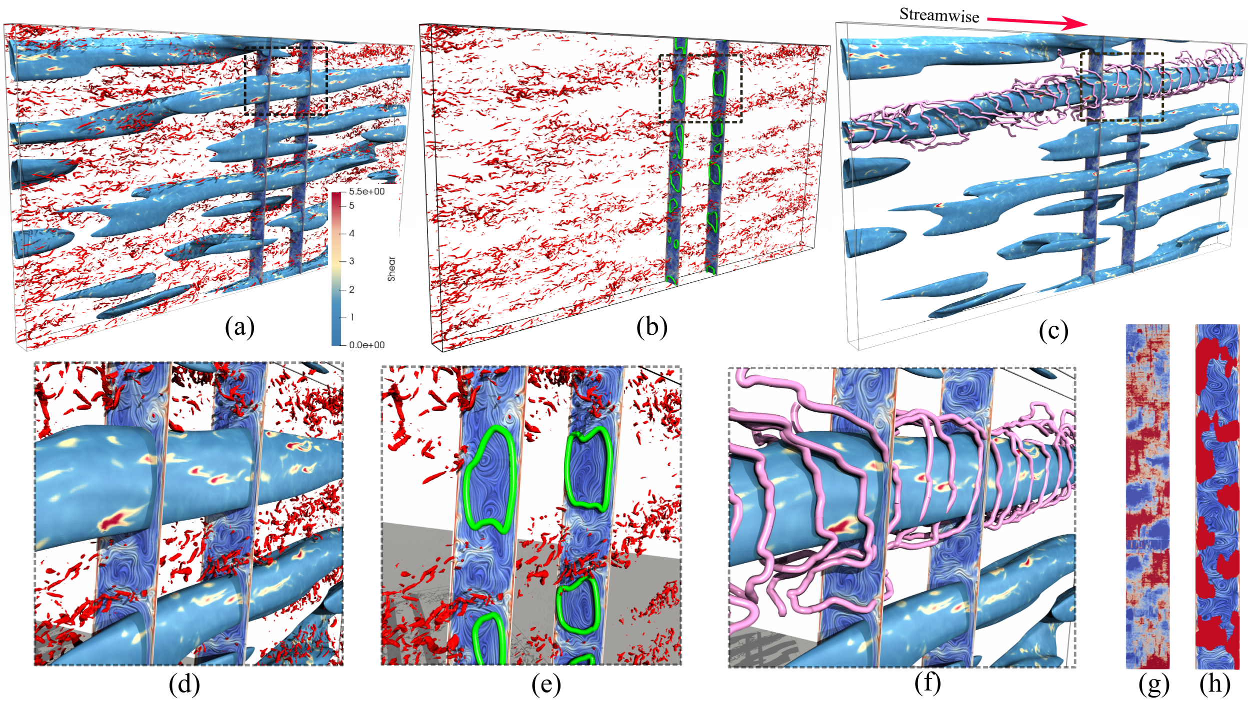


Figure 31: Result on a TFC with $R_\Omega = 0.05$. (a) the overview of the large-scale structures. The surface color is mapped to the shear attribute. (b) the small-scale vortices (in red) extracted with $Q = 2$, and slice cuts show the intersection of large structures and the WS planes. (c) streamlines wrap around a large structure. (d)(e)(f) provide the closeup views at the corresponding black rectangles in (a)(b)(c). (g) 2D abstraction representation indicates the regions with high concentration of the small structures, and (h) the projection of the small structures (red) on a LIC texture.

the large- and small- scale structures using attribute Q impossible. It is also known to the experts that there are no Taylor rolls in this TCF. Nonetheless, there may still exist other large-scale coherent structures. To reveal them, we again employ local shear rate and the streamwise velocity to compute the feature level-set for the extraction of the large-scale structure. The value ranges for these two attributes used here are [0.075, 0.15] for the streamwise velocity and [2.0, 5.0] for shearing, respectively.

Figure 32(b) visualizes the extracted surfaces approximating the large-scale structures. The streamline seeded at one of the surfaces corresponding to the large-scale structure has nearly flat geometry (i.e., not circulating around the corresponding surface as seen in the previous TCF results). This indicates that the flow motion at or near these large-scale structures is not rotational,

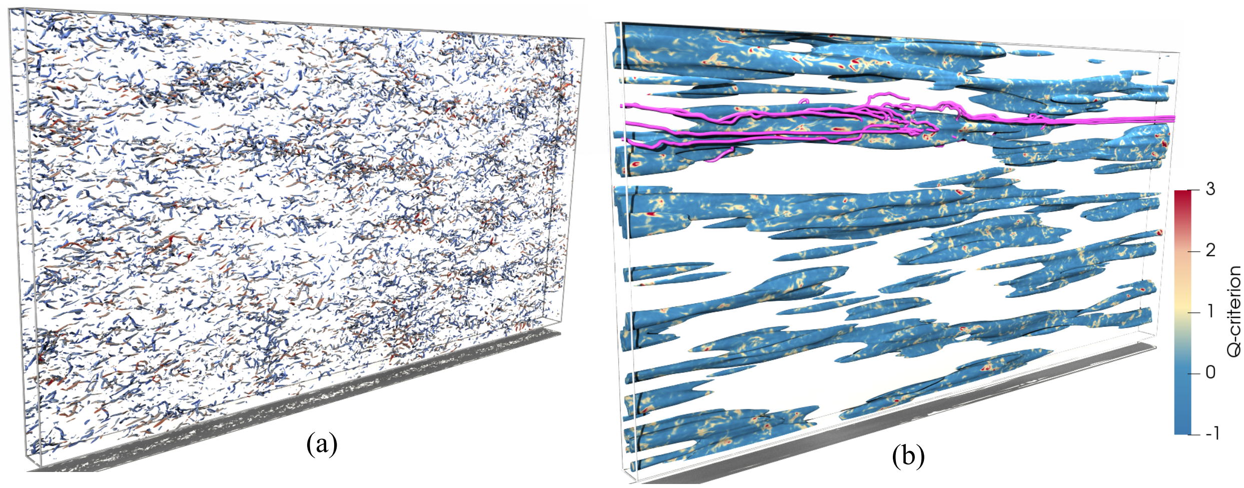


Figure 32: Visualization of (a) the vortices extracted by using the threshold $Q > 2$ for a TCF simulation with R_ω=0. It is impossible to visually distinguish the difference between the large- and small- scale structures as the vortices are distributed everywhere in the spatial domain, and the Taylor rolls are not clearly formed. (b) The large-scale structures are extracted using local shear rate and the streamwise velocity. The seeded streamlines (magenta) in the bottom have an overall flat configuration, indicating the absence of Taylor rolls.

which in turn shows that these structures are NOT Taylor rolls.

Performance. All numerical experiments reported are carried out on a PC with an Intel Core i7-9750H CPU, 128GB DDR3 RAM and a NVIDIA GeForce GTX1660Ti 6G graphic card. The framework is implemented with C++, and we use CUDA to parallelize the attribute, feature level-set and KDE computations. Note that the computation at each grid point is independent from the other points; hence, the computational complexity is reduced from N^3 for the sequential implementation (N is the spatial resolution in each dimension) to nearly a constant for our parallelized version.

5.4 Expert Evaluation

We have shown our results to an expert from the fluid mechanics community. In his impression, the visualization for the TCF with $R_\Omega = 0.1$ adequately captures the regions of vortex generation, and the "quiet" regions. Furthermore, the method is able to summarize the regions where vortices are formed from small-scale data, and this shows promise for application in a wide variety of problems like thermal convection, shear flows and wall turbulence. More impressively, the method is also able to extract the coarse-grain regions where the large-scale rolls are not streamwise invariant, but show a more complicated geometry (e.g., for the TCF with $R_\Omega = 0.05$), providing a rapid intuition of where transport barriers can arise.

In addition, the expert pointed out that our 3D visualization enables a more thorough inspection of the 3D behaviors of TCFs that were not possible with his existing 2D visualizations. Specifically, when $R_\Omega = 0.1$, each Taylor roll rotates in the opposite direction to its neighboring rolls, as observed in the respective 2D cross sections (Figure 2(a)). The circulating motion of the flow is nicely depicted by the sampled streamlines that wrap around the extracted tubular structure, which give an idea of how many wraparounds happen as the fluid particles transverse the flow. This large-scale configuration is relatively streamwise invariant (i.e., they are not shifting up and down) and remain stable in streamwise direction. In contrast, when $R_\Omega = 0.05$, the large-scale structure extracted is not streamwise invariant (moving up and/or down). In addition, this structure is not always connected along streamwise direction. This indicates the well-shaped Taylor rolls as seen in the $R_\Omega = 0.1$ simulation are not formed. In this case, inspecting only 2D cross sections of the flow is not sufficient, as some cross sections may have well-shaped vortices while the others do not.

For TCF with $R_\Omega = 0$, there is no existing method that can effectively isolate the large-scale structures from the small ones. Although the proposed method does not completely overcome this challenge, the extracted surfaces partially reveal the shapes and positions of the structures that the expert hypothesizes. This suggests that the proposed framework can be applied to a more systematic study on how R_Ω impacts the configuration of Taylor rolls and to identify the threshold value of R_Ω that leads to the fully developed Taylor rolls.

Although the presented method is able to isolate the Taylor roll successfully, the robustness of the isolation of the vortex remains to be properly assessed, and in the expert's opinion we have to go beyond visualization to quantify several aspects of the roll dynamics, such as its energy and position through time. The current approach cannot accurately capture the Taylor rolls for their quantitative analysis. Nonetheless, the expert believes that the method developed here could be used as springboard for low-order modelling of these structures, a very relevant research aim of the turbulence community in the present.

Comparison with the convolution kernel approach. To demonstrate the advantage of our framework over the traditional methods that usually perform smoothing to remove small-scale features, we compare our result with the most commonly used convolution kernel approach [105]. Figure 33 (b) shows the results using a convolution kernel - usually an averaging box filter with a kernel size of $12 \times 12 \times 12$ to the velocity field. Although the convolution kernel can generate a smoother result and better-connected surfaces, our approach aligns better with physics. This is because the convolution kernel may falsely enclose areas that are occupied by small-scale features. As highlighted by the yellow dashed circles in Figure 33, a region that is dominated by the small-scale vortices becomes part of the large-scale structure after applying convolution kernel (Figure 33 (b)). In contrast, our method correctly detects this area and excludes it from the construction of the large-scale structure.

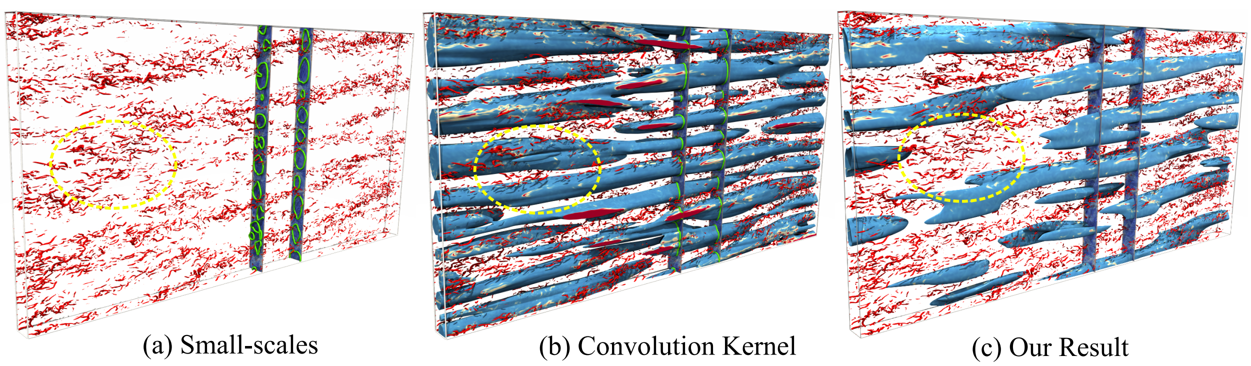


Figure 33: Comparison with the convolution kernel on a TFC with $R_\Omega = 0.05$. The large-scale structures extracted with the convolution kernel falsely includes an area occupied by small-scale features. In contrast, our method correctly excludes this region from the extracted large-scale structure.

In addition, we can see that the attribute values on the surfaces extracted using the convolution kernel have large variations (e.g., a few large areas colored in red on the those surfaces, indicating that the geometry there does not align well with the physical attribute (i.e., those parts should not belong to the extracted structure).

5.5 Parameter Discussion

Our approach depends on two types of parameters, i.e., the value ranges of the selected attributes, the neighborhood size m for the KDE computation and the iso-value for surface extraction. The discussion on how to properly select value ranges for the feature level-set computation is provided in Section 5.2.1. In this section we provide a more detailed discussion and an example regarding the effect of kernel size m to the smoothness of the distance field computed using the feature level set method. We also include an example showing the iso-surfaces extracted from the obtained KDE field for the representation of the large-scale structures in a TCF.

Effect of neighborhood size m. As described in the work, a local neighborhood centered at each vertex x is constructed for the density estimation in our KDE computation. The size of this neighborhood is controlled by parameter m. For example, if $m = 5$, a $5 \times 5 \times 5$ neighborhood around a central vertex x will be considered for the density estimation. The larger kernel size m is, the smoother the obtained surfaces will be. However, if m is too large, the density field can be overly smoothed, resulting in connected large-scale features. Figure 34 shows a few iso-surfaces extracted from the KDE density field computed with different values of m (i.e., neighborhood with different sizes). The iso-values used to extract these surfaces are the middle values of the data ranges of

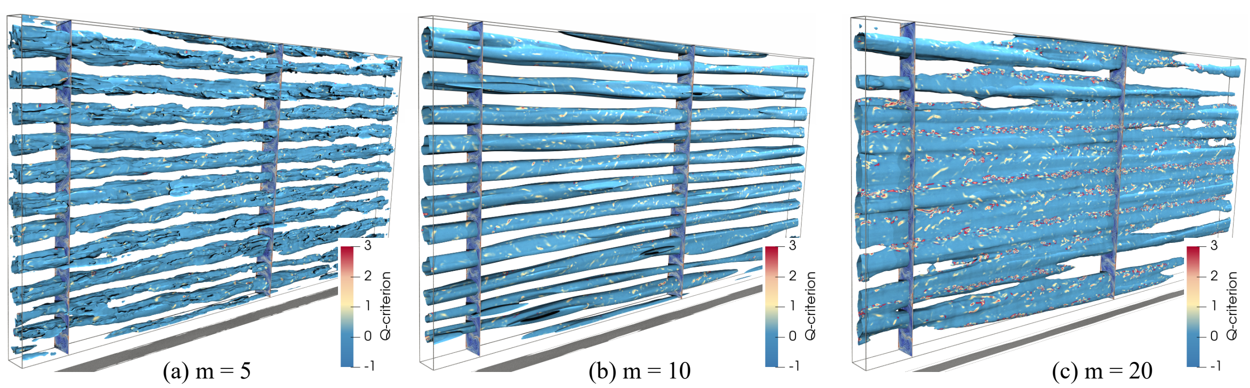


Figure 34: The effect of the size m of the neighborhood $\mathbf{P}$ in the KDE computation to the obtained surface representation of large-scale structures for the TCF with $R_\Omega = 0.1$. From the left to the right, $m = 5, 10, 20$, respectively. The corresponding iso-values are (a) 0.04, (b) 0.06, (c) 0.035.

the obtained KDE density fields. With the increase of m, the extracted surfaces become smoother.

However, when $m = 20$ or larger, the individual surfaces extracted become connected, violating our goal of well separation of the structures. In addition, the overly smoothed field leads to larger surfaces that include more small-scale structures, which is undesired. In practice, we found that $m = 10$ works the best for all three TCFs we experimented with.

Effects of different iso-values. Another parameter (or threshold) that the user needs to specify in our framework is the iso-value that is used to extract the iso-surfaces from the obtained KDE density field for the representation of the large-scale structures. Figure 35 shows the iso-surfaces extracted with different iso-values (i.e., 0.01, 0.06, and 0.08m, respectively) from the same KDE density field. As can be seen, smaller value tends to result in surfaces enclosing smaller volume area, while larger value leads to surfaces enclosing more volume region. Surfaces enclosing smaller volume region may miss important large-scale structures, while surfaces enclosing too many regions may include more small-scale structures. This is shown by the accuracy measurements (TP and FP) of the three surfaces. In particular, the TP and FP for the result with iso-value 0.01 are 96.2% and 18.7%, respectively, while the TP and FP for the result with iso-value 0.08 are 82.3%, and 3.2%, respectively. For comparison, the TP and FP for the result with iso-value 0.06 are 94.4% and 7.3%. Specifically, the iso-value 0.06 is the middle value of the data range of the obtained density field.

To extract the iso-surfaces with an ideal covering percentage of coherent structures, the iso-value is set to half of the density value range. Table 2 reports the optimal KDE parameter values and the optimal covering percentages (explained next) of large-scale structures for the three TCFs described above.

Accuracy measurement. To measure the accuracy of our results, we compute two covering percentages: (1) the percentage of the large-scale structures enclosed by the obtained surfaces and (2) the percentage of the small-scale structures enclosed by the surfaces. The former is computed as the ratio between the number of voxels enclosed by the surfaces that intersect with the large-scale structures (determined by the value ranges of the selected attributes) and the number of voxels

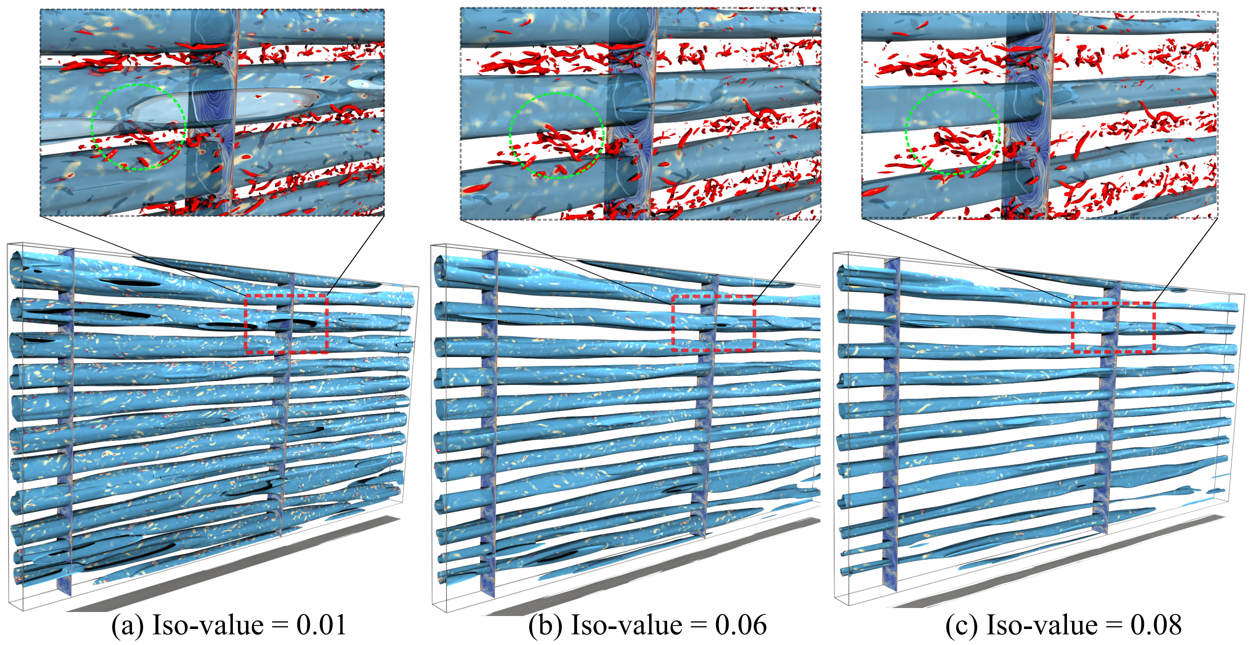


Figure 35: The iso-surface extracted with different threshold values from the KDE field for the representation of large-scale coherent structures for the TCF with R_ω=0.1, (a) 0.01 (b) 0.06, (c) 0.08, respectively. The red fibers are the small-scale features extracted with the threshold $Q > 2.0$. With a small value (a), the obtained iso-surfaces are big and include more small-scale structures (the green circle). On the other hand, if the threshold value is too large (c), the surfaces may miss a large amount of large-scale features (the yellow circle). The result in (b) provides the best coverage among all three values. In particular, the TP and FP values for the result with iso-value 0.01 are 96.2% and 18.7%, respectively, while the TP and FP values for the result with iso-value 0.08 are 82.3%, and 3.2%, respectively. For comparison, the TP and FP values for the result with iso-value 0.06 (middle) are 94.4% and 7.3%.

Table 2: Parameters used for all TCFs with $m = 10$ and their accuracy measurement.

R_Ω	**Bandwidth**	**Iso-values**	**Large-scale CS covering (TP)**	**Small-scale CS covering (FP)**
0.1	0.72	0.061	94.4%	7.3%
0.05	0.60	0.050	89.8%	9.5%
0.0	0.42	0.043	85.7%	17.6%

in the entire domain that intersect with large-scale structures. This essentially estimates the true positive (**TP**). The latter is computed as the ratio between the number of voxels enclosed by the surfaces that intersect with the small-scale structures and the number of all voxels enclosed by the surfaces. This is measuring the false positive (**FP**). An ideal surface should have TP close to 100% and FP close to 0. Table 2 reports the accuracy measurement of our results. As a comparison, the TP and FP values for the convolutional kernel result shown in Figure 33 (b) are 92.32% and 18.4%, respectively. Although it achieves slightly better coverage of large-scale structures than our result (Figure 33 (c)), it covers significantly more small-scale structures than it should, which makes it less accurate than our result.

5.6 Discussion

In this work, we present a framework for the visualization and analysis of Taylor-Couette flow (TCF), a well-known turbulence flow that is frequently studied in various situations. However, existing methods cannot effectively separate the large-scale structures (i.e., Taylor rolls/vortices) from the dense, space-filling small-scale structures for the study of transport barriers. To address this issue, we propose a novel visualization framework. First, we derive physical attributes, and find a combination among these attributes via the feature level-set technique to better capture the difference between the large- and small- scale features. Second, we extract the iso-surface from the kernel density estimation of the distance field obtained from the feature level-set computation. We also provide 2D abstract representation through the plane projection along the streamwise direction to highlight the concentrated positions of structures with different scales. Our method is simple yet effective, enabling the separation of the large-scale structures from the smaller ones. It leads to cleaner visualization of this turbulence flow, facilitating its analysis for the experts. We have applied our framework to three TCFs simulated with different parameters to assess its effectiveness. We show that our framework can be used to distinguish TFC with different configurations.

Though we successfully separate regions with small-scale features from those with large-scale ones for TCF, there are still a number of limitations that need to be addressed in the future. First, the quality of the extracted surface representation for large-scale structures still depends on the proper selection of the value ranges of certain relevant attributes. This is not trivial and the thresholds can be arbitrary, as criticized by the expert. Second, we do not really extract the large-scale structures precisely, instead, we only provide an approximation on the regions where they may reside. In the meantime, the expert wishes to see the more accurate transport barriers of this flow, i.e., the boundaries of Taylor rolls, to quantify their dynamics. Finally, we focus on one time step for each TCF despite TCFs are unsteady flows. Nonetheless, the efficient computation of our framework enables us to explore an in-situ visualization and analysis of TCFs during their simulations. We plan to address these limitations in the future work.

6 Dynamic Mode Decomposition for Large-Scale Coherent Structure Extraction in Shear Flows

6.1 Problem Overview and Contributions

A very relevant type of turbulent flow is *wall-bounded shear flows*, which are generated in a fluid between two surfaces with different velocities. These have been extensively used to explore new concepts in fluid mechanics such as instabilities [99], non-linear hydrodynamics [98], and pattern formation [5]. Two popular wall-bounded shear flows are plane Couette flow [61] and Waleffe flow [20]. The former is the flow between two infinite plates which move with different velocities, while the latter is the flow bounded by two infinite stress-free plates and forced using a body shear forcing force.

Problem Overview: Large-scale structures have been observed in shear flows [20], but extracting these structures is not trivial. Many approaches [14, 50, 74, 89] have been developed by the fluid mechanics and visualization community to assist experts understand coherent structures and flow dynamics. In fluid dynamics, moving ensemble averages [67], two-point correlations [94], and Fourier analysis [66] have been commonly used to detect and analyze coherent structures. The fluid dynamic researchers also employ visualization techniques such as volume rendering and iso-surfacing. The main disadvantage of these methods is that they often include the relatively arbitrary choice of a cut-off to isolate the structures, be it in the shape of a cut-off frequency in Fourier space [66], or in the shape of a threshold value for a certain attribute [1]. Furthermore, these methods are usually tuned for one particular setup of the system; but as the control parameters of the flow change, the shape of the coherent structure also changes. As a result, conventional methods can completely fail to track the structures across certain dimensions of parameter space. Finding a method which requires minimum thresholding still remains a challenge.

Dynamic mode decomposition (DMD) is a data-driven and parameter-free method that provides a spatio-temporal decomposition of data into a set of relevant dynamical modes called DMD modes from a sequence of snapshots of an evolving system. It has been initially introduced by

Schmidt [92] to provide the best *linear approximation* that sends the data from its current state to the next in a non-linear dynamics system. DMD computation relies on the parameter-free Singular Value Decomposition (SVD) method. Each DMD mode is considered as a spatial structure which is accompanied by time dynamics. The corresponding time dynamics of DMD modes can be characterized by their speeds (e.g., how fast or slow they move), making DMD a promising candidate to extract coherent structures (evolving with different speeds) in flow. Several works [129] have attempted to apply DMD to different kinds of flows. For example, Gilka et al. [28] performed a DMD analysis on the flow behind an actuated bluff body, and Schmid [92] tried DMD with the Gurney flap wake flow to capture the vortex shedding pattern. However, the capabilities of DMD in analyzing shear flows have not been fully investigated. In addition, DMD has not received much attention from the visualization community for the task of spatial and temporal feature extraction; hence there is no existing visualization system specialized for DMD, reducing its analysis ability.

Contributions: To address this challenge, we propose to use multi-resolution Dynamic Mode Decomposition (mrDMD) [48] which is a variant of the standard DMD for large-scale structure extraction in shear flows. In particular, we show that the slow motion DMD modes extracted by mrDMD are able to capture large-scale structures in shear flows. To speed up the computation of mrDMD, we provide a fast GPU-based implementation. We also introduce a visualization framework built on the fast mrDMD implementation which can help users effectively analyze the obtained DMD modes and their time dynamics for the study of the large-scale structures in shear flows. We have applied our framework to 2D and 3D Plane Couette (PC) and Waleffe flows using both 2D cuts and full 3D flow fields to demonstrate its effectiveness. This is the first time mrDMD is applied to 3D Plane Couette and Waleffe flow snapshots for large-scale structure extraction. We find that for the shear flows experimented in this work, the slowest mode of the first level mrDMD can already sufficiently capture the large-scale structures in the flows. This allows us to develop a parameter-free large-scale structure extraction based on DMD for shear flows. We compare our DMD based method with other existing methods, such as convolution kernel based smoothing, time average, and the proper orthogonal decomposition (POD) to demonstrate its advantages. In addition, we performed experiments of DMD on other types of flows than shear flows and found that the slowest DMD mode may not always extract useful structures. Instead, modes selected from different levels of mrDMD can better capture the dynamics of those non-shear flows, which indicates the needs of mrDMD. We report these experiments and attempt to provide some empirical guideline for the visualization community on the proper use of DMD, especially mrDMD, for turbulence flow analysis and visualization.

6.2 Parallel Implementation for mrDMD

There have been numerous literature studies investigating the feasibility of DMD with fluid flows [108, 47]. Due to the high memory consumption, however, they were only able to perform the experiments on 2D or small 3D simulations. As described in Section 2.3, the size of the input matrices X_1, X_2 for DMD computation is $N \times M$ where the number of rows N is substantially greater than the number of columns, resulting in a so-called tall-and-skinny (TS) matrix. For a large dataset, it is impossible to perform SVD on a single processor. Parallel algorithms are required to make DMD more broadly applicable to the large-scale simulations. Recently, many efforts [72, 90] have been made to improve the computation of DMD. Pendergrass et al. [72] proposed a parallelized algorithm to compute the dynamic mode decomposition (DMD) on a graphics processing unit using the streaming method of snapshots singular value decomposition. Sayadi et al. [90] employed the parallel Tall-Skinny QR(TSQR) algorithm to the DMD, allowing the decomposition of very large datasets. To our best knowledge, there is no existing parallelized implementation for mrDMD. Although the main components of mrDMD still comes from the standard DMD, it is worth providing a complete solution for mrDMD.

The fundamental computation of DMD is based on SVD. We can compute SVD of a matrix A through the QR decomposition as follows:

$$A = U\Sigma V^* = QR = QU_R\Sigma_R V_R^*, \text{where } [U_R, \Sigma_R, V_R^*] = svd(R) \tag{19}$$

The left singular vectors U_i of A are computed by using $U_i = Q_i U_R$. The singular values and the right singular vectors are already stored in Σ_R and V_R^*, respectively. The parallel pipeline of QR is illustrated in Figure 36. The input matrix A is decomposed into sub-sets which have the same number of columns, but smaller number of rows. Traditional QR decomposition is performed on each subset in a single processor core.

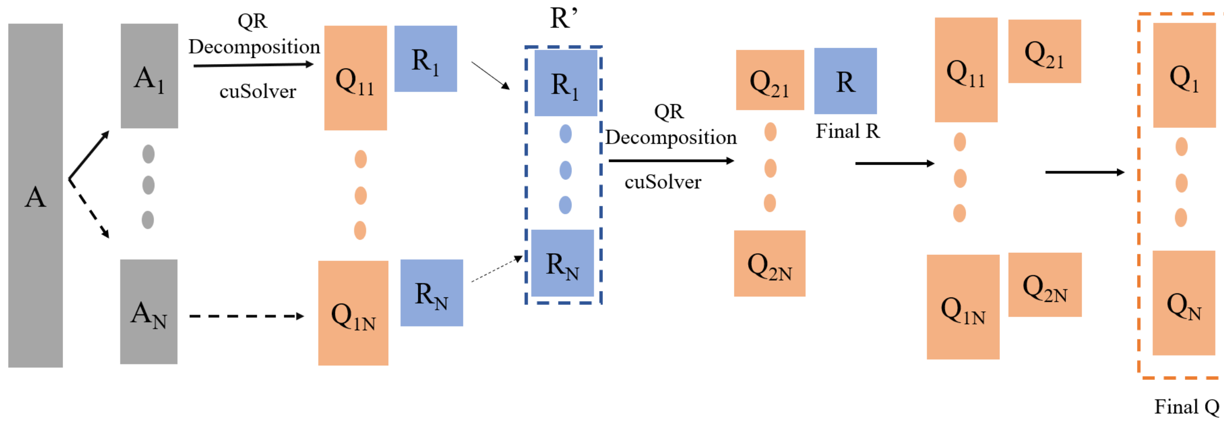


Figure 36: The pipeline of the direct TSQR algorithm in which the tall-and-skinny input matrix A is decomposed into smaller pieces so that they can fit into the GPU memory. QR decomposition is performed on each subset of the matrix A, then the outputs are combined in order.

Once we construct the DMD computation based on the parallelized SVD, mrDMD is implemented based on the algorithm described in Section 2.3.1. We utilize the open-source cuSOLVER [62] to perform the Eigenvalue and QR decomposition, as well as cuBLAS for the large matrix multiplication.

Performance of the proposed parallel mrDMD. To evaluate the accuracy and performance of our GPU-based implementation, we converted the existing Python-based mrDMD available at [77] to C++. The matrix manipulation tasks including SVD, multiplication, and Eigendecomposition rely on Eigen [29] – a CPU-based open-source library. We then compare the differences between the obtained DMD eigenvalues and modes from the CPU- and GPU- based versions on ten artificial 2D datasets which have the spatial resolution ranging from 250×250 to 1200×1200. Two time windows $NT = 50$ and $NT = 100$ are used to determine the number of snapshots for mrDMD computation. The data size varies from 17MB to 2.3GB. Figure 37(c) shows the average relative errors for the DMD eigenvalues and modes in each test cases. For both eigenvalues and modes, the relative errors are smaller than 1e-5. Figure 37(a,b) shows the running times of both the CPU and the GPU versions. For small datasets, the CPU-based implementation runs faster. However, when data size increases, the GPU-based version starts outperforming the rival. The performance gain of our GPU-based implementation versus the previous CPU-version increases significantly when the data size increases. We observe that the CPU-GPU data transfer is the most time consuming step. For each large matrix operation (e.g., SVD and Eigendecomposition) we need to copy the entire matrix from CPU to GPU, then copy back the result from GPU to CPU. This limitation can be addressed with a better optimized algorithm in the future.

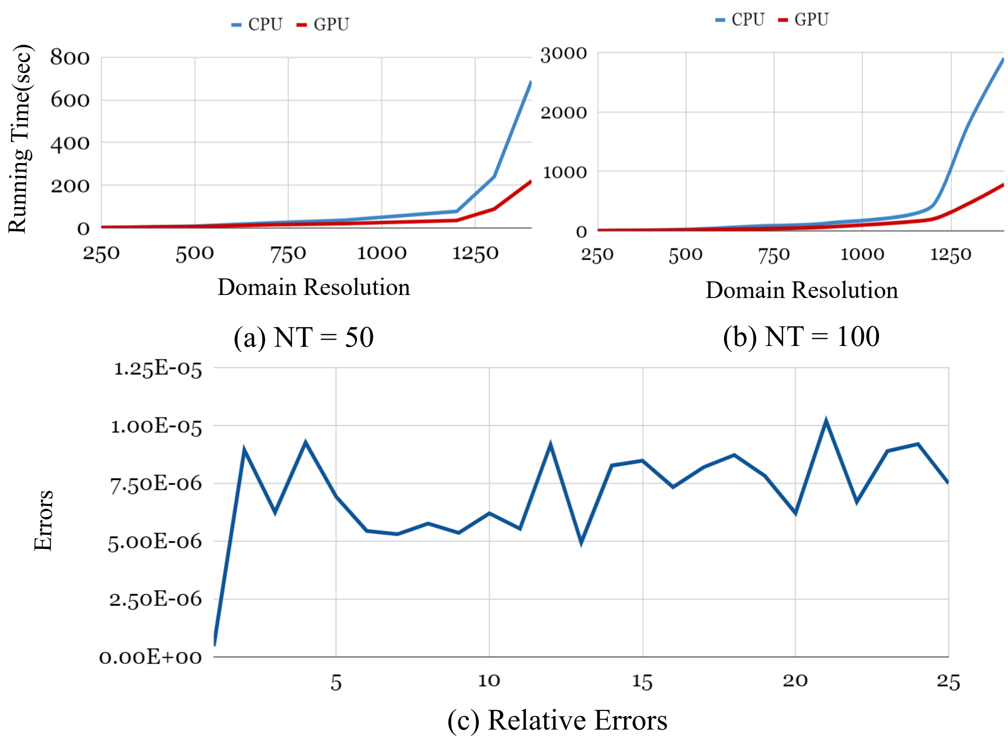


Figure 37: Performance analysis for the GPU-based implementation. (a) reports the running time of CPU and GPU-based implementation on ten artificial 2D datasets with the time windows $NT = 50$. (b) uses similar data, but with the time window $NT = 100$. The data files have spatial resolutions ranging from 250×250 to 1400×1400. The size increases from 17MB to 2.3GB. The bottom graph shows the relative errors of the GPU-based DMD eigenvalues and modes with respect to the CPU-based results. Twenty five test cases are performed in total.

6.3 Large-scale Structure Separation using mrDMD

In this section, we discuss the properties of DMD modes and their eigenvalues which make them suitable for the large-scale coherent structure extraction problem. A coherent structure is characterized by temporal and spatial information. We can use DMD eigenvalues to characterize the temporal persistence of structures. Basically, DMD is a Fourier-based decomposition process. Each DMD mode has time dynamics defined by a single complex eigenvalue. We can reconstruct the original field by taking the summation of all mode values. However, the contribution of each mode at any given time is different, as controlled by the frequency of the eigenmodes. The frequency also indicates the movement speed of the mode: at larger frequencies the DMD mode changes faster. Similar to matrix eigenvectors, each mode is normally normalized to have a unit norm/magnitude.

For the spatial extraction of coherent structures, we can rely on DMD modes as they capture all of the spatial features existing in the flow domain. *If structures evolve linearly, then they will coincide with the DMD modes.* For example, Figure 38 shows two DMD modes generated with a simple Tube simulation. In this simulation, a vortex tube starts from the left side of the domain and moves to the right side, then breaks down. We take a cross section of the original 3D simulation [3] in this example. It can be seen from the two DMD modes that both of them can capture some positions of the main vortex over time. However, the magnitudes of the vortex in each mode are varied. Indeed, a DMD mode works as a projection plane where we can take all the spatial features and project them on the plane, but with different intensity. To reconstruct the features in the original data, we simply add the weighted modes with the weights determined by their temporal magnitude as shown in the bottom plots of Figure 38 and Eq.(8). Note that this reconstruction takes into account both the real and the imaginary parts of the modes and their time evolution characteristics, while in all visualizations of modes only real parts of the modes are shown.

The next step is to determine the modes that have our desired features. As the large-scale structures move slowly as discussed in Section 2.3.1, the corresponding indicators - slow DMD modes (especially the slowest mode) - become an ideal candidate to characterize these large-scale structures. With our fast mrDMD, we can identify the slowest mode from each iteration (or level)

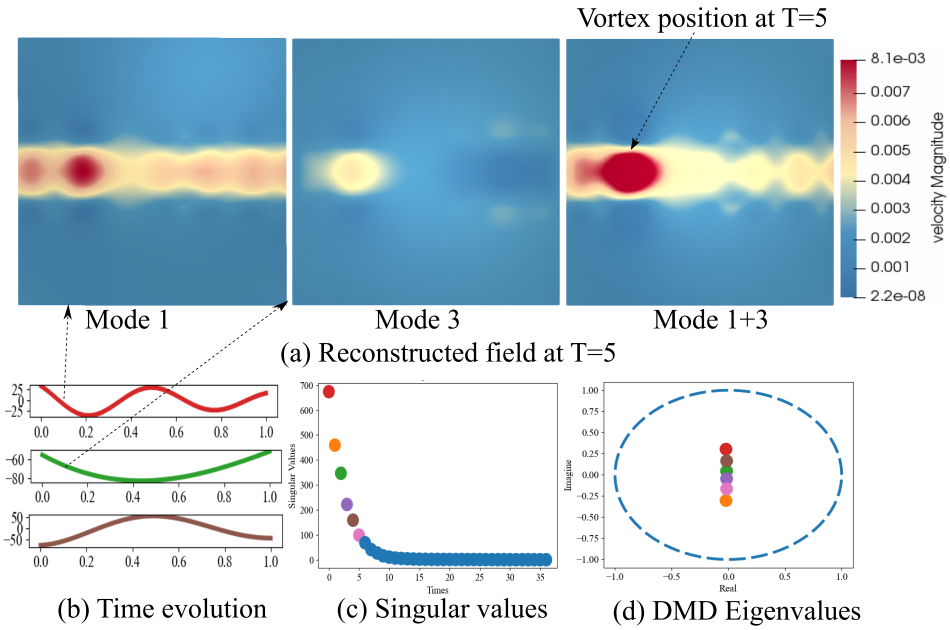


Figure 38: Sample reconstructed field using DMD modes and their time evolution from the Tube simulation. (a) The two DMD reconstructed fields shows different intensity of the main vortex at time $T = 5$. Combining two modes reveals a correct position of the main vortex at the selected time. The last row presents the plots of (b) the time evolution, (c) singular values, (d) DMD eigenvalues and of DMD modes.

of DMD, and use it and its time evolution to reconstruct a flow as the input for the generation of the large-scale structure visualization. In our experiments, we found the slowest mode from the first level of mrDMD most effectively captures the large-scale structures in shear flows (Section 6.4). Once the large-scale structures are separated in the flow reconstructed from the slowest DMD mode, the small-scale structures can be studied in the flow by subtracting the DMD reconstructed flow from the original flow, as shown in Figure 39. The study of the small-scale structures in shear flow is beyond the scope of this work, which we will leave for the future work.

We wish to point out that when utilizing DMD for flow analysis and interpretation, visualizing the reconstructed flow from the selected modes often provides more intuition of the spatio-temporal behaviors of the flow than only visualizing modes and their time evolution that may miss some information (i.e., the imaginary part of the modes and time evolution).

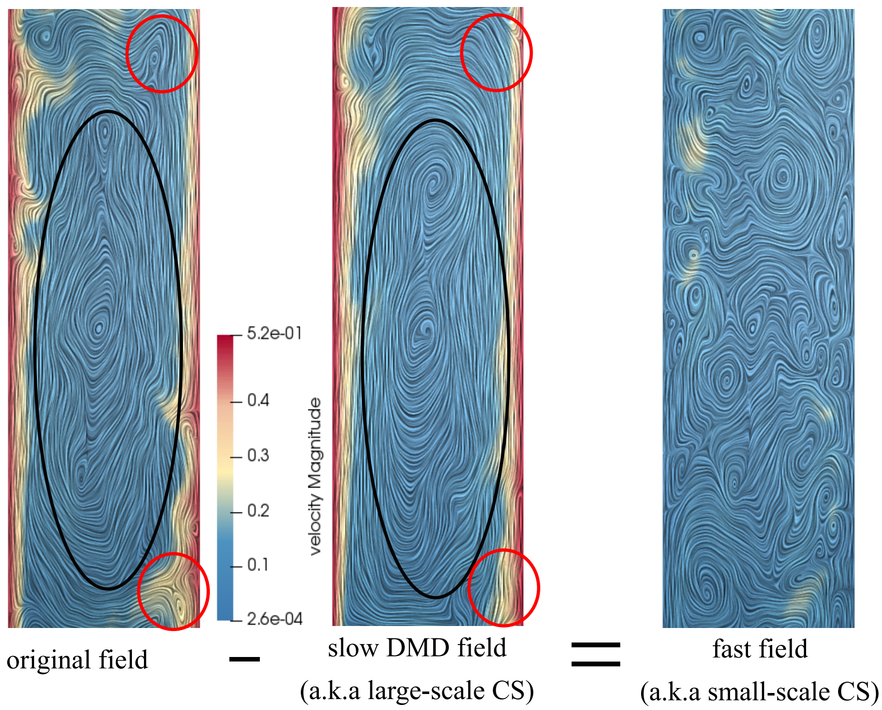


Figure 39: The relation between the original and reconstructed DMD fields in the plane Couette flow with $R_\Omega = 0.1$. The 2D Taylor vortices depicted by black circles are the target large-scale CS in the simulation. The large structures can be observed clearly in the DMD reconstructed field. Small-scale features near the wall are highlighted in the red circles in the original field. The small features are revealed in the subtracted field.

6.4 Applications

We have applied our mrDMD based large-scale structure separation framework to the two main types of shear flows, namely the Couette and Waleffe flows. Their configuration is described in Section 2.2. We look at several 3D Plane Couette simulations that are generated with different anti-cyclonic solid-body rotation ratios R_Ω ranging from 0 to 0.1. Waleffe flow is simulated with a Reynolds number $Re = 3.16 \times 10^3$ and an anti-cyclonic rotation ratio R_Ω =0.63. For these control parameters, Farooq et al. [20] recently used the autocorrelation and other statistic-based methods to prove that large structures exist in the Waleffe flow. We demonstrate that the slow DMD modes can reveal similar well-known large-scale features in both Plane Couette and Waleffe flows. By using the obtained slow DMD modes, we can isolate the large structures without threshold selection.

For each Couette and Waleffe flow, we generate a set of 2D and 3D snapshots. We use 2D datasets to verify and compare our results to the existing works of Taylor Couette and Waleffe flows that are mainly focused on their 2D counterparts. One of the main reasons that looking at 2D cross section of these flows may be sufficient is that the large structures exist along the streamwise direction; thus, fluid mechanic researchers can perform analysis on 2D streamwise-spanwise planes. We also use 2D data to evaluate the effect of time windows on the final DMD modes.

The fast GPU implementation can enable us to process the 3D snapshots of Taylor Couette and Waleffe flows for the first time. However, a large time window is still a challenge for the 3D analysis because the memory required to store the 3D data is beyond the physical memory of our workstation. An out-of-core, streaming implementation of mrDMD is needed, which is beyond the scope of this work. The timing information of our mrDMD when applied to the shear flows used here is provided in Table 3.

Table 3: Performance of DMD on three datasets

Dataset	**Spatial Resolution**	**Time Steps**	**Running Time (minutes)**
2D plane Couette	1024×512	100	6.45
3D plane Couette	$512 \times 384 \times 256$	40	32.6
3D Waleffe	$512 \times 384 \times 256$	80	124.38

2D-slices of a plane Couette flow- R_Ω=0.1. In the first experiment, we apply mrDMD to 2D slices of a Plane Couette simulation with $R_\Omega = 0.1$. The 2D data is obtained by taking the slice planes parallel to the walls at the mid-gap. As there is a large amount of the overlapping small-scale vortices as shown by the LIC texture of the flow (Figure 40(a2)), extracting the large structures is not a trivial task. The traditional methods using FTLE computation (Figure 40(a3)) or physical attributes (e.g., $Q - Criterion$ and λ_2) are not able to reveal any meaningful large structures. It is important to mention that the Taylor vortices - the prominent large-scale structures - are fully formed with $R_\Omega = 0.1$. Thus, this simulation is an ideal candidate to verify the correctness of the selected DMD modes. Three hundreds time steps are collected and input to our mrDMD program. The DMD eigenvalues are visualized in Figure 40(b3-bottom). The eigenvalue which is nearest to the origin is highlighted in the blue circle. Its corresponding DMD mode is shown in Figure 40(b1-b2). It is easy to recognize some prominent structures spanning along the stream-wise direction in the DMD mode.

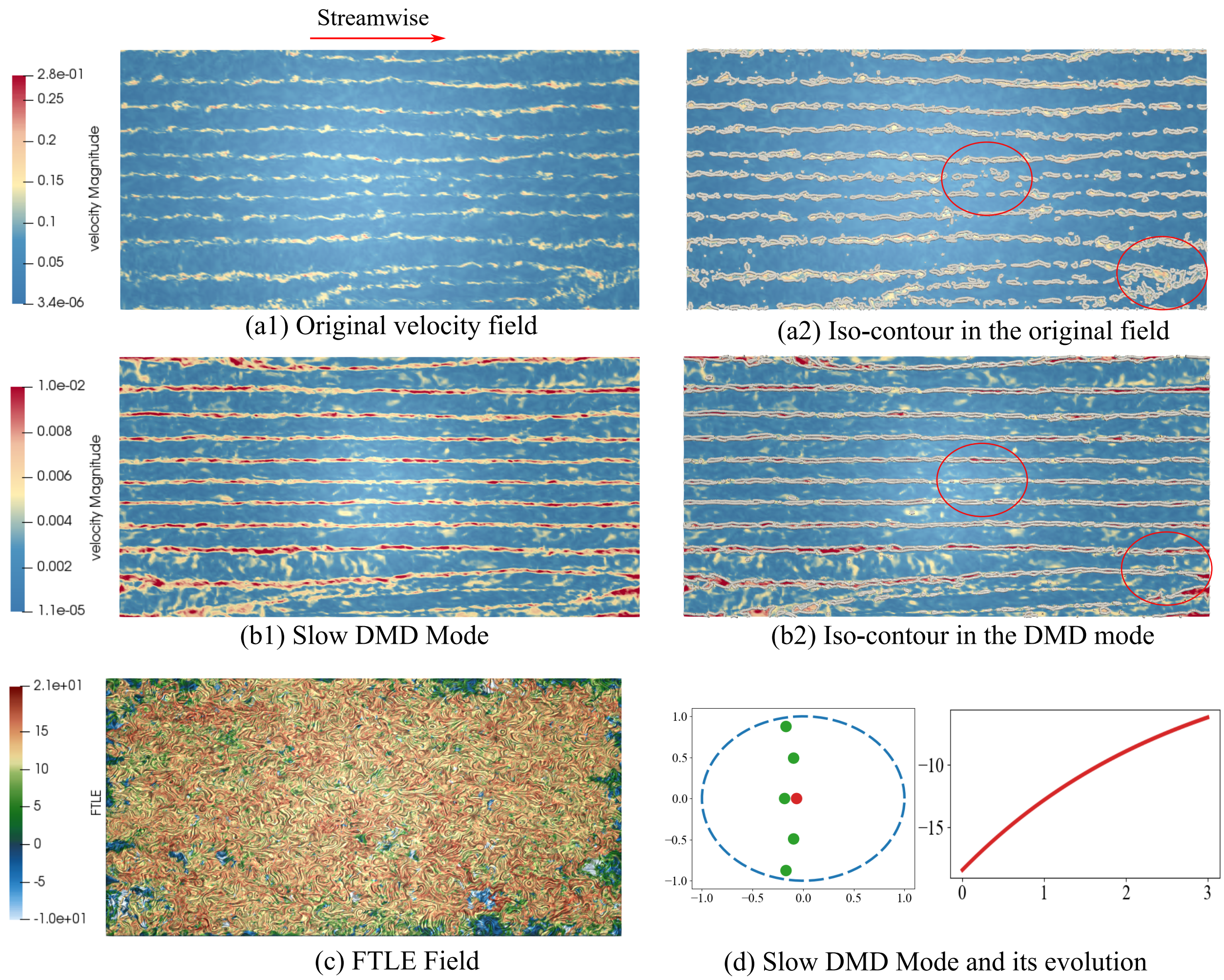
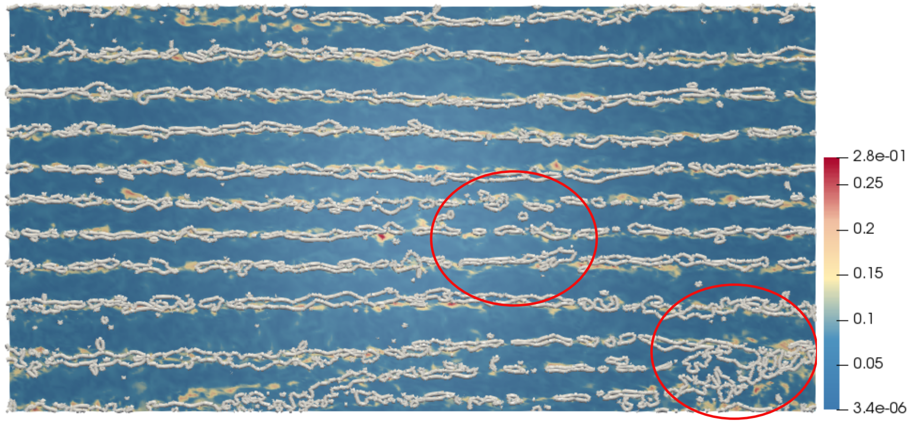
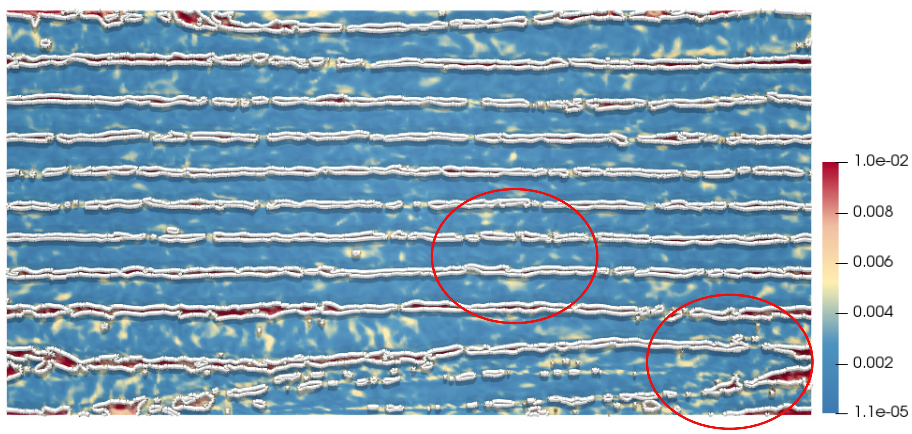


Figure 40: Visualization result on a Plane Couette flow with $R_\Omega = 0.1$. The cut is at the centerline of the wall-normal. The first time step is shown. (a1) Show the original velocity field. The horizontal line structures are the boundaries between two counter-rotating Taylor vortices (a2) The corresponding LIC texture. (b1) The slowest DMD mode helps to better capture the separation of Taylor vortices as their boundaries are clearer than (a1). (b2) its LIC texture. (c) The visualization of FTLE field. (d) shows the plot of DMD Eigenvalues (the red dot). The selected DMD eigenvalue is the closest one to the origin of the unit circle (blue).

The comparison between the iso-contours extracted from the original velocity field and the slowest DMD mode is shown in Figure 41. The highlighted areas indicate several differences between the DMD mode and the original flow. As can be seen, our result produces cleaner iso-contours.



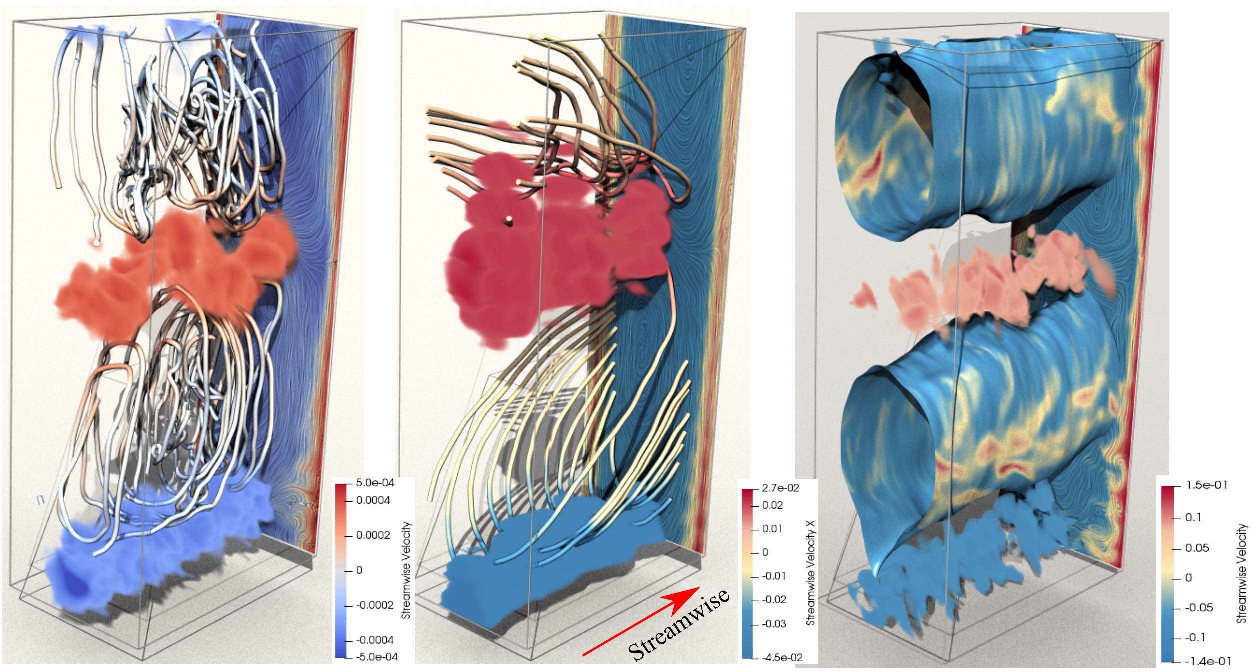
(a) Original field



(b) DMD

Figure 41: Comparison between iso-contours extracted from the velocity in the original flow (a) and the slowest DMD mode (b) for the Plane Couette flow with $R_\Omega = 0.1$. The first time step is shown. Iso-values are 0.12 in (a), and 0.09 in (b) due to different magnitude of the velocity between two fields.

3D Plane-Couette - R_Ω=0.1. We apply our mrDMD based method to full 3D snapshots of a Plane-Couette simulation with a smaller boundary size than the 2D version. Again, the Taylor vortices are the large-scale structures in this simulation setting. The resolution for the 3D spatial domain $NX \times NY \times NZ \times NT$ is $512 \times 384 \times 256 \times 40$. The total data size is 36GB. It took our framework 32.6 minutes to process this data. As shown in Figures 39, the large-scale structures - Taylor vortices (indicated by black circles) - are hardly observed in the original field. On the other hand, these structures can be revealed clearly in the extracted slow DMD mode. Small-scale features near the wall are highlighted in the red circles in the original field. These small features are considered as noise, and they make the separation of the large structure challenging. Our DMD-based method can filter out these small-scale structures. As shown in Figure 42(a), the slowest DMD mode of this field more effectively reveals the large-scale separation structure that separates the individual Taylor rolls/vortices. Note that the separation layers or the iso-surfaces in 3D correspond to the iso-contours in the 2D results shown in Figure 41.



(a) DMD Mode (b) Convolution kernel (c) Nguyen et al. [61]

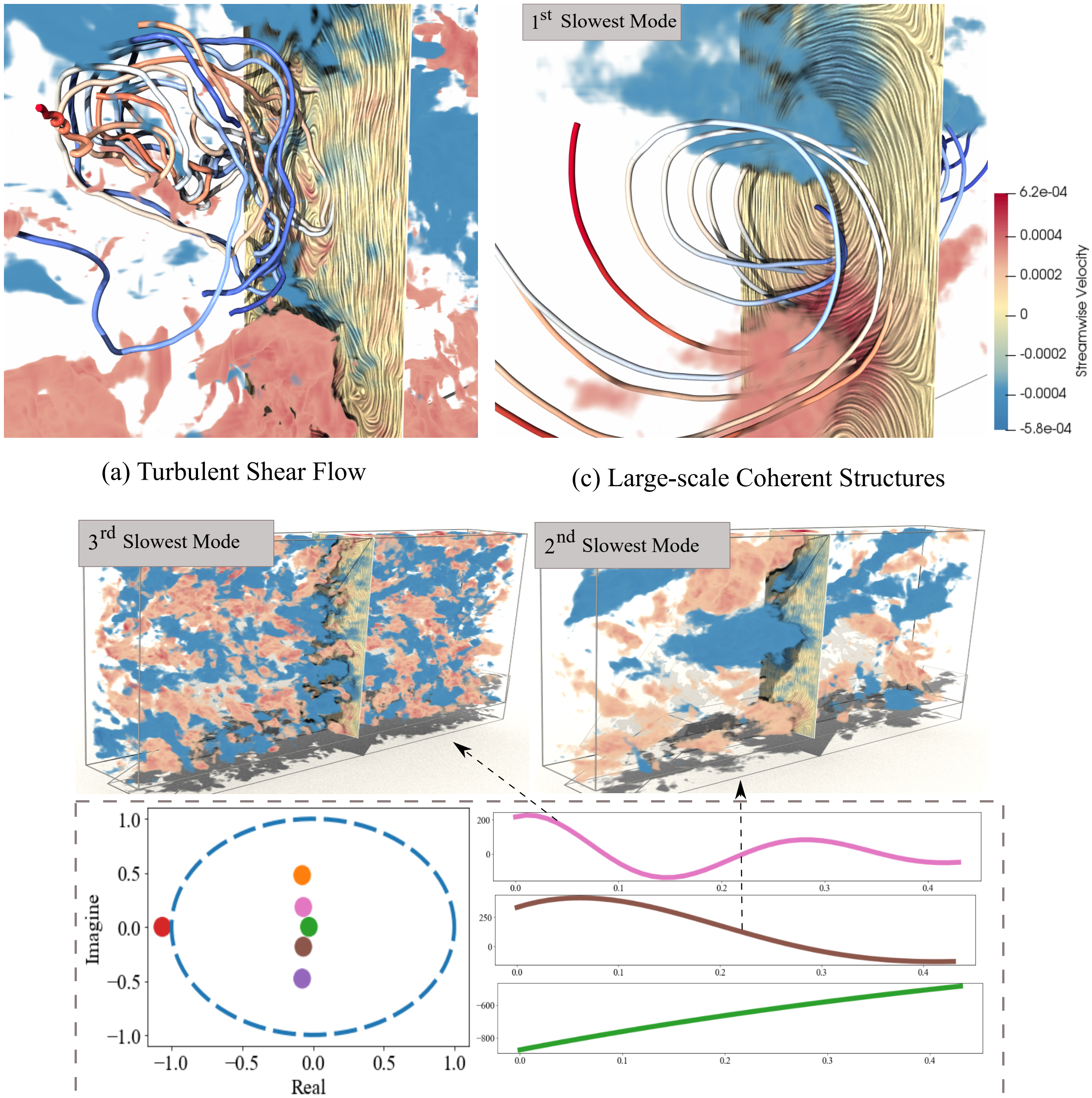
Figure 42: Comparison between DMD, the convolution kernel and Nguyen et al. [61] approaches on the 3D Plane Couette flow with $R_\Omega = 0.1$. Volume rendering show the prominent structures in the DMD mode (a), and the smoothed field using a convolution kernel (b). Streamlines in (a) capture the expected Taylor roll structures better than the ones in (b) as the the top structure is distorted due to the oversmoothing. (c) The approximated surfaces by Nguyen et al.'s method reveal the similar large-scale structures as shown by streamlines in (a)

3D Waleffe - R_Ω=0.63. The Waleffe flow is simulated with the Reynolds number $Re = 3.16 \times 10^3$ and the rotation ration $R_\Omega = 0.63$. A recent work [20] has demonstrated that the large-scale structure emerges with this optimum rotation ratio. We collect 80 continuous snapshots with a spatial resolution of $512 \times 384 \times 256$. The three slowest DMD modes are shown in Figure 43. Again, we select the slowest mode which is nearest to the origin of the unit circle. The second, third, fourth and fifth modes are provided as the references to compare with the slowest mode. The faster the modes, the more small-scale features they can reveal. It is interesting to point out that their eigenvalues have the identical real parts, but they have different imaginary parts. As the result, their corresponding time evolution plots are similar.

Figure 44 shows the difference between the original field and the extracted slow mode. The original field is turbulent, including many small-scale features. It is difficult to observe any prominent large-scale features. In contrast, the extracted slow DMD mode helps to reveal the structures similar to the Taylor vortices seen in the above 3D Taylor Couette flow. In the most recent work [20], these structures are reported as the large-scale coherent structures for the Waleffe flow. We use the 2D slice cuts at different positions along the streamwise direction to highlight the shape of the prominent structures.

After the filtering process, we can extract iso-surfaces based on the streamwise velocity component of the DMD mode as shown in Figure 45. These iso-surfaces form a layer to separate two neighboring large-scale structures, which is similar to 3D Taylor Couette flow, as indicated by the seeding streamlines that are circulating around some common curves. However, the large-scale structures revealed in the Waleffe flow are not always invariant along the streamwise direction in contrast to the Taylor rolls seen in the Taylor-Couette flow. This result solidifies the intuition already hinted by the data and visualizations in [20]: the Waleffe flow large-scale structures are not exactly analogous to those of Plane Couette flow but present some oscillations in the spanwise direction. This state-of-the-art fully three-dimensional visualization will be useful to precisely characterize the difference between the large-scale structures in Waleffe and plane Couette flow in the future.

Comparison with POD. We apply POD to the Waleffe flow and extract the first ranked mode or the most energetic mode. The results are shown in Figure 46.



(b) Multi-resolution Dynamic Mode Decomposition

Figure 43: We apply the multi-resolution Dynamic Mode Decomposition (mrDMD) to extract large-scale coherent structure for turbulent shear flows. (a) shows a closeup view of the input Waleffe flow. The colors are mapped to the streamwise velocity, i.e., the x component of the velocity vector with red means positive blue for negative. (b) demonstrates a few DMD modes that correspond to coherent structures of different scales for this flow. Each mode is sufficiently characterized by their eigenvalues shown in the unit circle plot and their respective time evolution plots. (c) the slowest DMD mode characterized by the slowest change in its time evolution plot is used to represent the large-scale structure of this flow that also has a slow dynamic. The three modes of the entire flow domain can be found in the supplemental document.

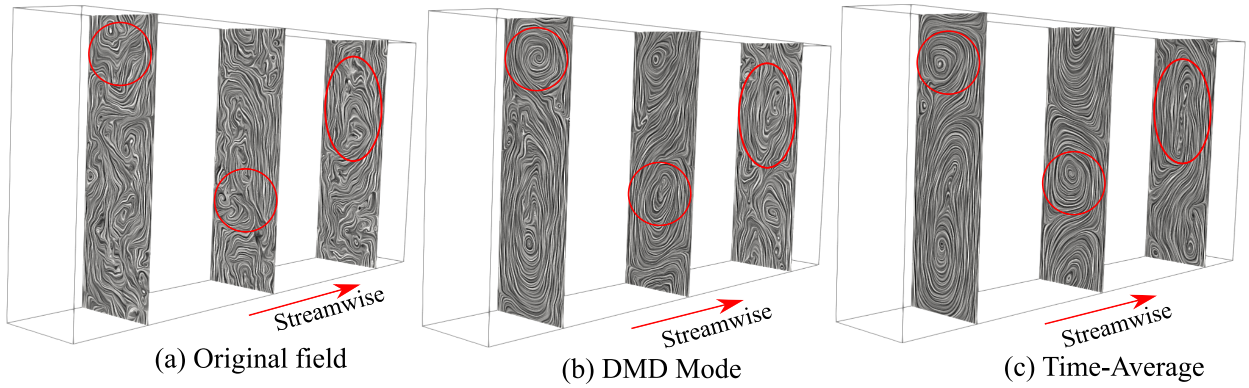
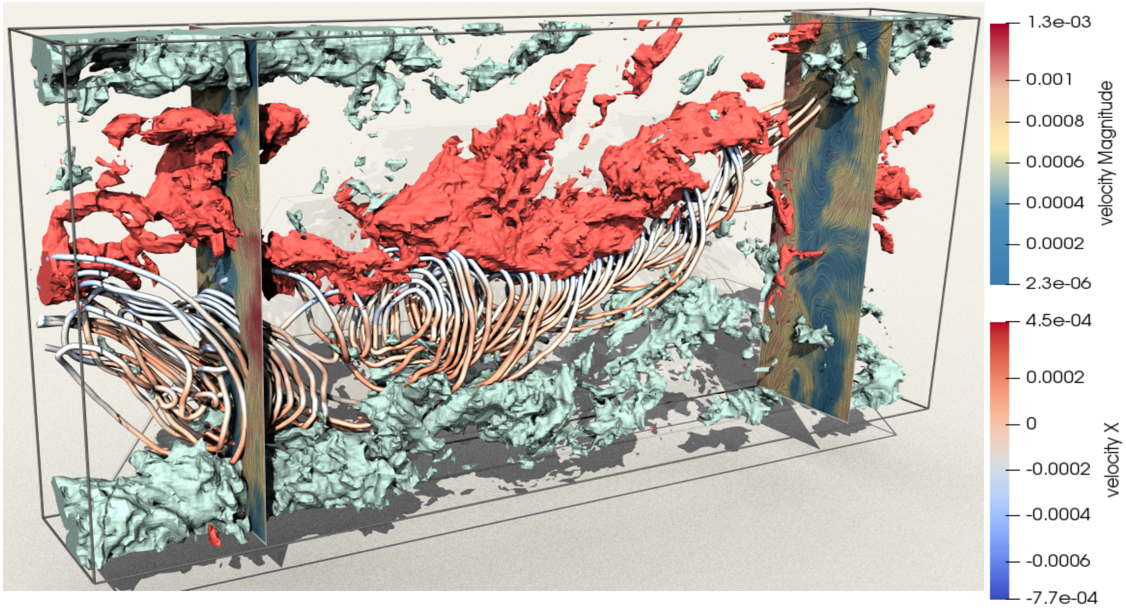
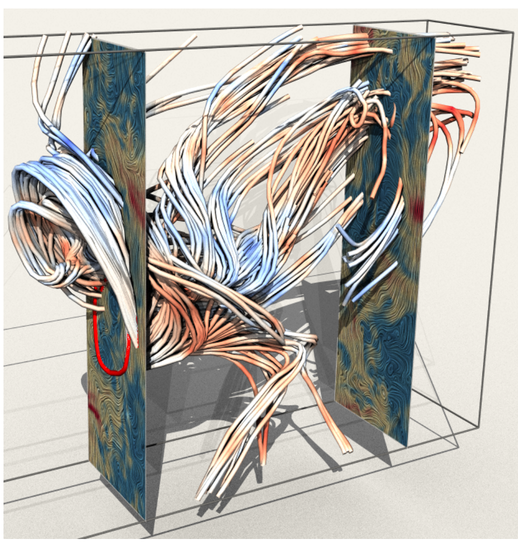


Figure 44: Results of a 3D Waleffe flow with $R_\Omega = 0.63$. The first time step is shown. Three slice cuts along the streamwise direction at $x = 1$, 3.14, and 5 are shown, respectively. (a) The original field is very turbulent. It is impossible to visually distinguish the large-scale feature. (b) The large-scale coherent vortices appear clearly in the DMD mode. The extracted vortices behave similarly to Taylor vortices. This observation is verified by the result of the recent work [20]. (c) Time average velocity field also reveals similar structures, but their temporal information is discarded.

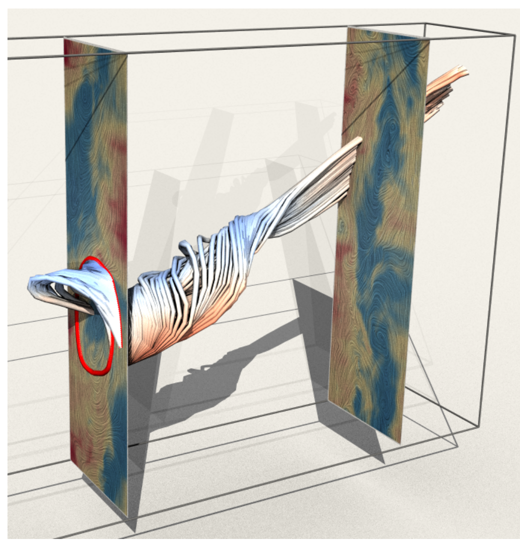
Although the extracted POD mode has similar spatial patterns as the DMD mode and the large-scale structures are observed in both modes, their time coefficients do not encode meaningful time evolution information of the spatial features. As demonstrated in Figure 46(a), the reconstructed flow field from POD suffers from a temporally non-coherent characteristics (i.e., the velocity magnitude changes arbitrarily over time), which is not physically plausible. In contrast, the flow reconstructed from the slowest DMD mode and its corresponding time evolution has smooth transition of its pattern and velocity magnitude, which is more physically meaningful and accurate. POD and DMD use different criteria to compute the modes. The POD seeks for the optimal approximation via a principal component analysis (PCA), while the DMD tries to obtain the best linear dynamical system describing the input data. POD is a statistical decomposition technique that does not tell us much more about the modes that it finds, and further processing must be done on them. This is unlike the information which can be obtained from a DMD decomposition: time evolution of the spatial features, where we know the modes evolve linearly and with a certain dynamics given by the characteristic frequencies. The advantages of DMD over POD have also reported in previous works [47].



(a) Streamwise velocity iso-surfaces and Streamlines



(b) Pathlines in the original field



(c) Pathlines in the DMD field

Figure 45: Results on the 3D Waleffe flow with $R_\Omega = 0.63$. (a) The iso-surfaces of the velocity streamwise component in the slowest DMD mode. The roll-like structures act as the separation layer which separate two neighboring large-scale structures. The streamlines depict the shapes of the large structures along the streamwise direction. (b)(c) Pathlines in the original and the DMD reconstructed field are generated with a same seeding curve (the red line). The pathlines in DMD reveal a vortex-alike structure similar to (a) due to the almost stationary characteristic of the structure.

Comparison with the simple time average of the original flow. It is shown that a simple time average (e.g., average all velocity values at the individual spatial locations over time) may

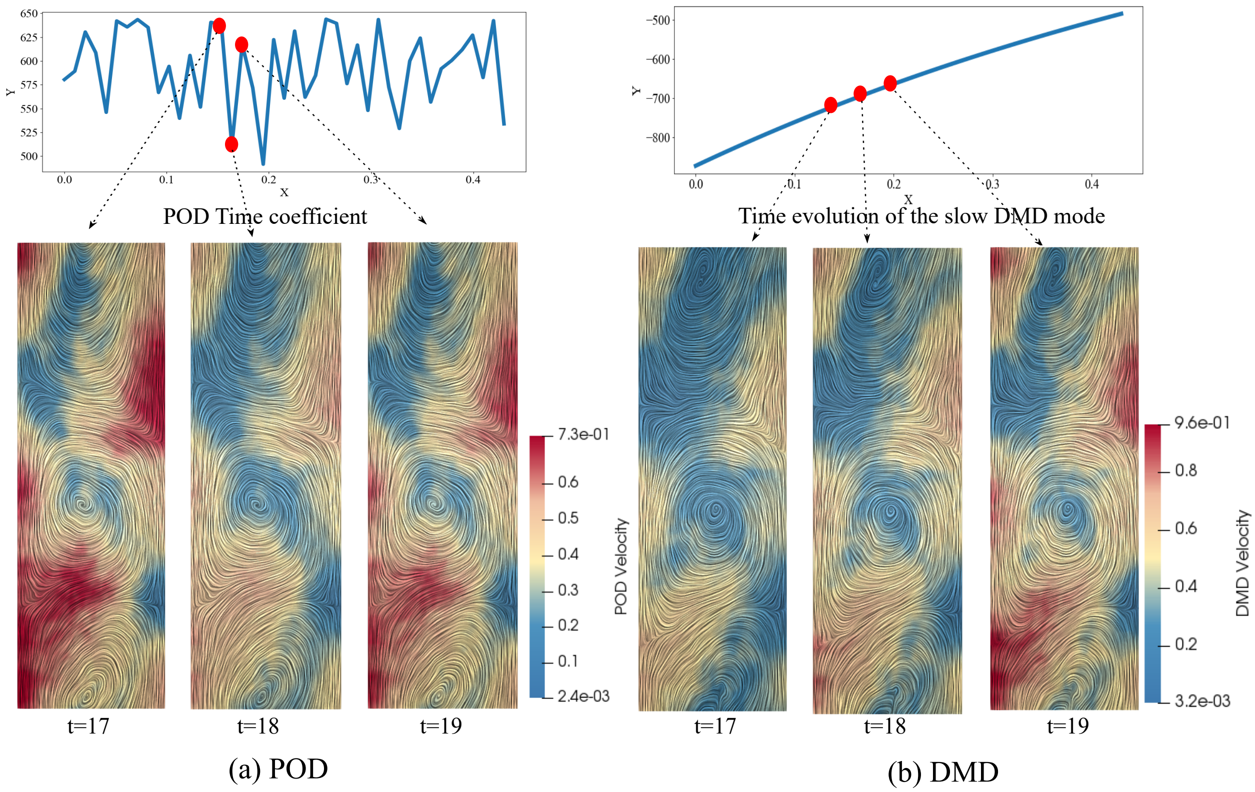


Figure 46: Comparison between POD and DMD on the Waleffe flow. The cuts are at the center of the streamwise direction for the Waleffe flow. Although the spatial representation of large-scale structures are observable in both the slowest (for DMD) and the first (for POD) modes, POD cannot characterize the temporal characteristic of the structures. In the three consecutive time steps selected (red dots), the velocity magnitude varies significantly in POD contradicting to the slow movement characteristic of the large-scale structures. In contrast, the slow motion of the structure is captured accurately by the DMD method.

reveal the existence of certain structure with slow change over time in the flow. In particular, large-scale structures that move slowly would be emphasized in the average flow due to their continuous contribution to the average computation, as shown in Figure 44 (c). Nonetheless, the average flow loses the temporal information of the flow features. In contrast, even though the slow DMD mode is similar to a single snapshot of the flow, it is accompanied by the information of its contribution to the flow over time (characterized by the time evolution plot); thus, it naturally encodes more temporal information of the flow than a simple time average. In the example shown in Figure 44, both the time average (c) and the slowest DMD mode (d) reveal similar large-scale structure of the 3D Waleffe flow. However, by multiplying the time evolution plot, the DMD mode reveals certain temporal behavior of the extracted structure as shown in Figure 47, while the time average cannot.

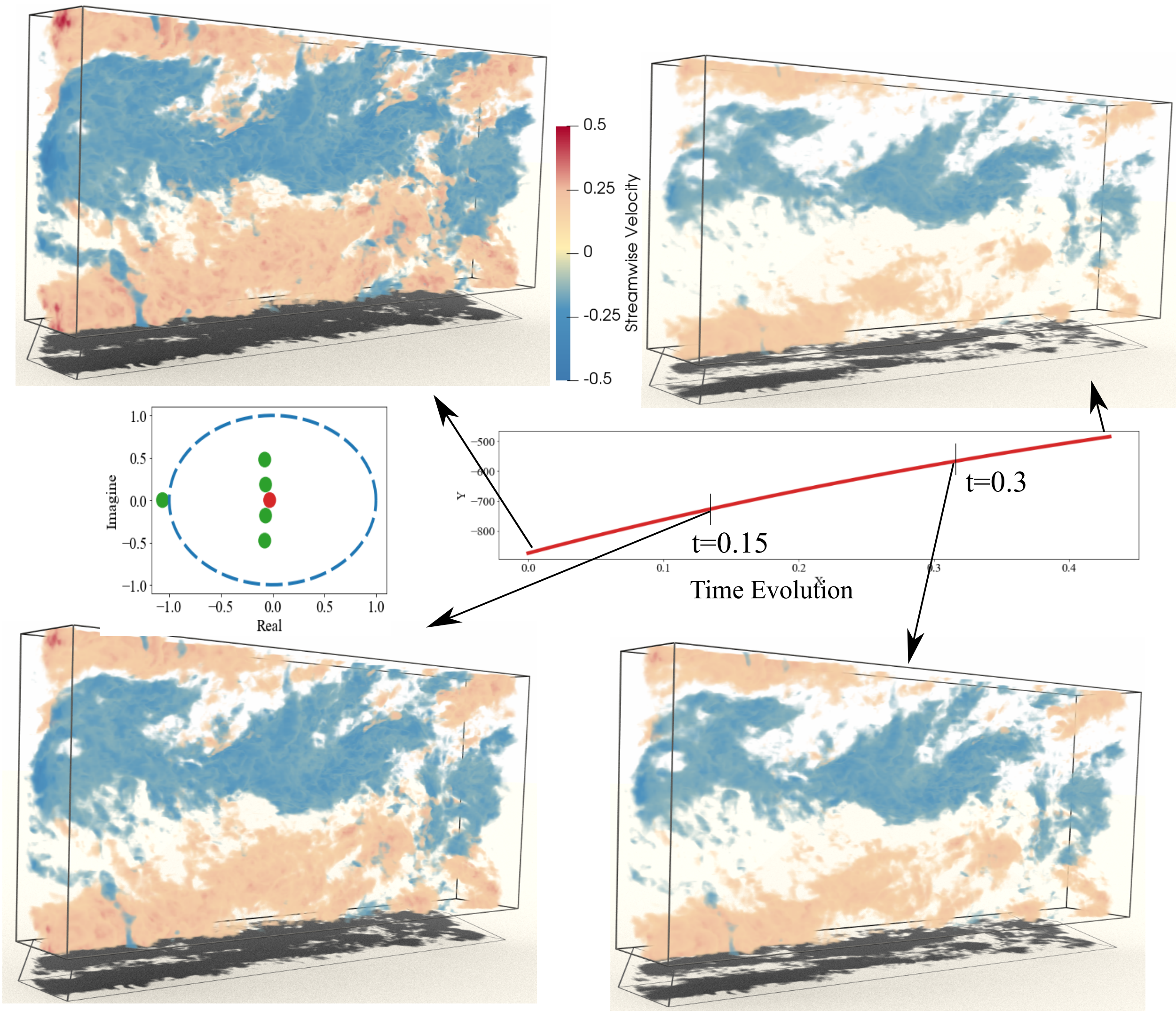


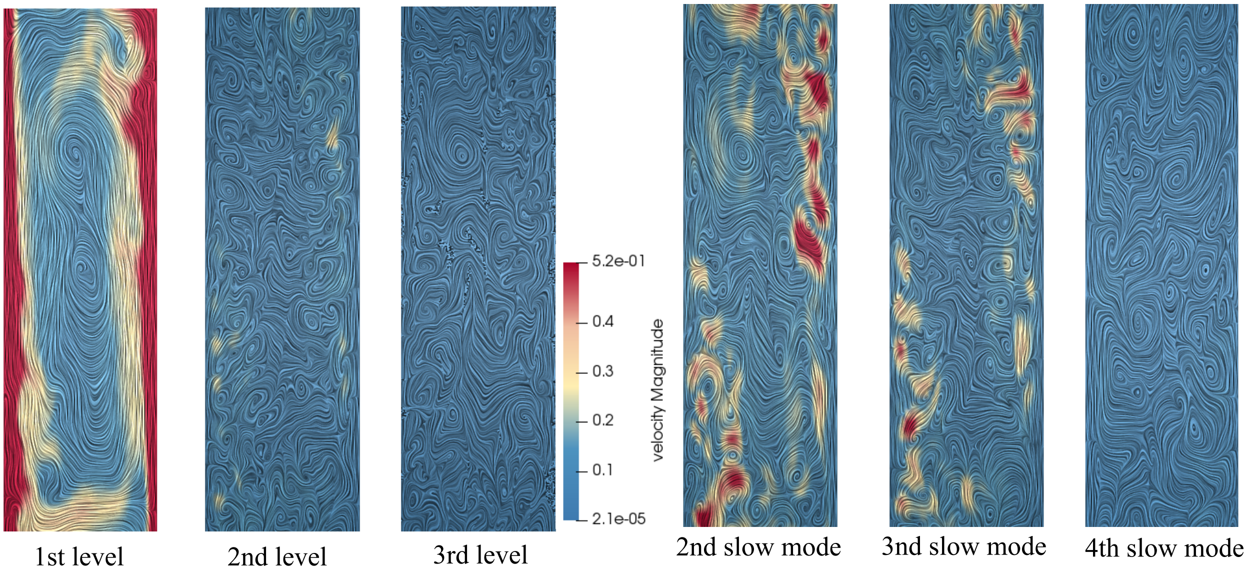
Figure 47: The volume rendering of the streamwise velocity of the reconstructed field using the slowest DMD mode at a number of sampled times. Although the time evolution is increasing, their values are negative; thus, the structures become weaker.

Comparison with convolution kernel based method [105]. Another popular approach for revealing large-scale structure is to perform a convolution kernel based filtering. We compare our DMD-based method with the convolution kernel based filtering in Figure 42. In particular, the extracted large-scale structure using our method is shown in Figure 42 (a), while the result of the convolution kernel based is in (b). From the comparison, we see that the convolution kernel-based method can also reveal the positions of the large structures, but it is threshold dependent and may significantly alter the velocity field as shown in (b). In contrast, the slow DMD mode gives the most meaningful result which retains the main features from the original field, while the small vortices near the walls are removed. More importantly, our DMD-based method does not require any thresholding trials.

6.4.1 Other Modes Returned by MR-DMD on Shear flows

Slow modes in higher-level mrDMD. mrDMD is an iterative algorithm that can provide multiple levels of slow modes and their corresponding temporal evolution. After the first level, the lowest-frequency or slowest modes are removed from the data. The remaining data having more dynamic behavior is used for the next level. It means that the slow modes in the second level are faster than the ones in the first level. Since the large-scale structures of interest are mostly stationary or move very slowly in shear flows, the slowest mode in the first level is sufficient. This observation is demonstrated in Figure 48(a) which shows three slow modes in the first levels, and the slowest modes in the other levels. The large-scale structures are only observable in the first level. The higher level modes are useful for capturing transient phenomena, handling the translational and rotational invariant in the data [47]. These properties do not exist in the large-scale coherent structures of shear flows. In the next section, we discuss the application of mrDMD for non-shear flows in which the higher level modes can be used to extract coherent structures that have translational property. Note that selecting the single slowest mode in the first level is parameter-free. However, choosing the number of levels and a sufficient number of slow modes in each levels is a parameter dependent process. It is still an open problem and beyond the scope of our work.

Other slow modes in the first level mrDMD. Figure 48 (b) show the other slow modes in the first level mrDMD. Compared to the slowest mode, these slow modes do not encode the expected large-scale structures of this flow.



(a) The slowest modes in the different mrDMD levels

(b) Modes in the first mrDMD Level

Figure 48: Slow modes in different levels of mrDMD. The large-scale coherent structures are only able to be observed in the slowest mode of the first level.

6.5 MR-DMD on Non-Shear Flows

To evaluate the effectiveness of the proposed method for the coherent structure extraction task in non-shear flows, we perform an experiment with a vortex ring simulation which simulates a vortex ring hitting a wall with a Reynolds number of 2000. We take the 2D cross section of this simulation as input [3]. The flow motion is illustrated in Figure 49(a). During the interaction with the wall, the primary vortex (i.e., the cross section of the vortex ring) approaches the wall and induces a boundary shear layer. As the vortex slides against the wall, the boundary layer becomes unstable and is lifted up as a secondary vortex, which in turn lifts up the primary vortex. This data set helps us analyze the role of coherent structures interacting with boundaries, and the generation of turbulence in wall-bounded flows. The flow is a transient flow, and contains several phenomena within, rather than being a purely shear flow. Because of the high-Reynolds number, we do not expect that at any level this flow has structures that behave quasi-linearly. Our mrDMD result is shown in Figure 49 (b–d). There are two pairs of symmetric eigenvalues (b). The two slow modes shown in Figure 49(d1)(d2) are selected from one eigenvalue of each of the

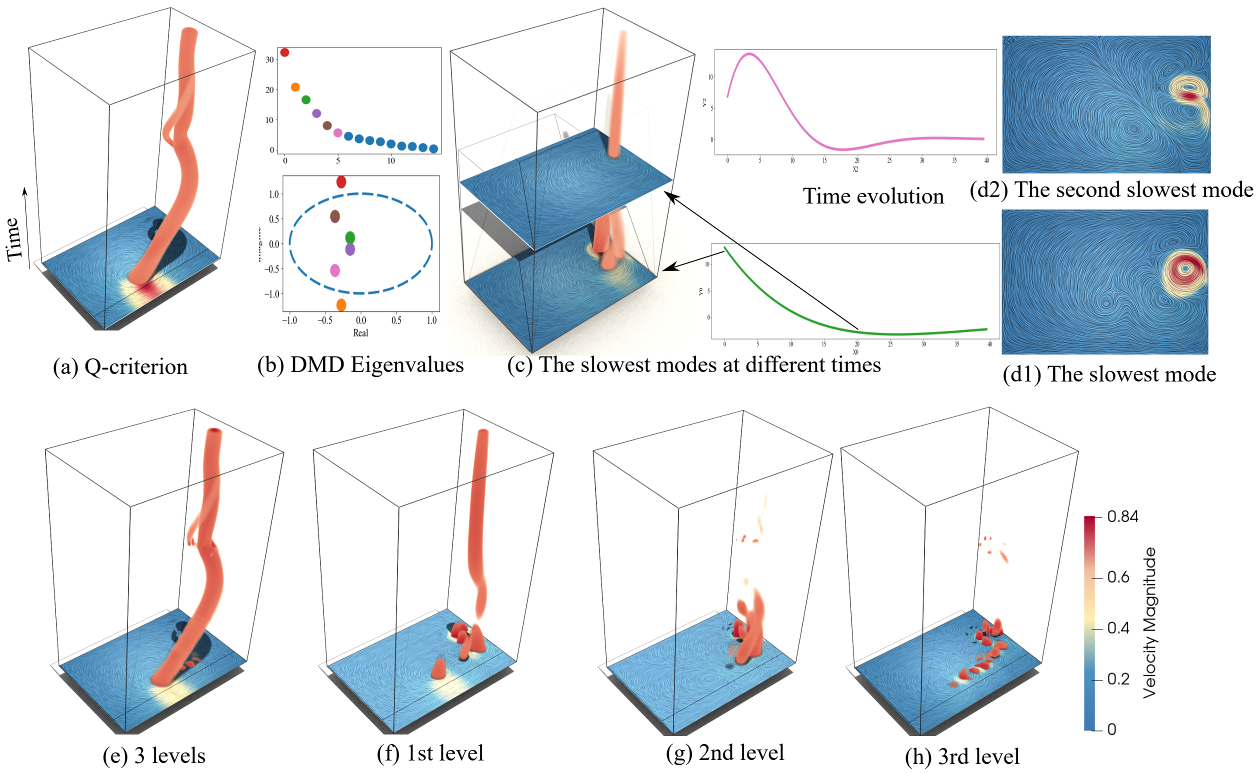


Figure 49: The mrDMD results of the 2D vortex ring simulation. (a) The volume rendering of the Q-criterion field illustrates the behavior of the original flow where a primary vortex starts near the center of the domain and gradually moves toward and eventually hits the wall, which induces a shear layer at the wall that lifts a secondary vortex. (b) The plot of the singular (top) and DMD eigenvalues(bottom). Note that the ranks of the singular values do not correspond to the position of the DMD eigenvalues in a unit circle. (c) Volume rendering of the Q-criterion derived from the reconstructed field by using the slowest modes. The two time slice cuts show the evolution of the slow mode over time. The slowest mode only captures the behavior of the flow after the primary vortex hits the wall. Two slow modes are shown in (d1)(d2). The volume rendering of Q-criterion derived from the reconstructed field by using (e) slow modes in three levels, (f) three slow modes in the first level, (g) two slow modes in the second level, and (g) one mode in the third level. We need to combine multiple levels to capture the behavior of the main vortex structure.

two pairs. The reconstructed field using these two modes according to their time evolution plots is shown in (d). As can be seen, the primary vortex before impacting the wall is completely lost in the reconstructed flow. This indicates that while the slow DMD modes provide a mathematically correct decomposition, they do not capture any relevant physical modes in this case. Specifically, they fail to capture the complete behavior of the primary vortex which is the dominant feature in the flow. This simple example shows that DMD (a linear approximation of unsteady flows) may not work well with flows that are not quasi-linear. However, how to determine whether an unsteady flow is quasi-linear or not is not a trivial task, which is beyond the scope of this work yet is important to achieve in order to provide a more accurate guideline for the application of DMD framework.

6.6 Discussion

In this work, we introduced to the visualization community a large-scale coherent structure identification framework for shear flows based on the recently popular dynamic mode decomposition (DMD). Our method is based on an observation between the slowest DMD mode and the large-scale structures seen in the shear flows. In particular, we show that the slow DMD mode characterizes the coherent structure that has a slow temporal evolution, which shares some similarity to the large-scale structures in the shear flows that also change slowly. Based on this observation, we propose to use the slowest DMD mode to help identify the large-scale structures in the shear flows. To address the issue of the standard DMD in handling transient time events, we resort to multi-resolution DMD (mrDMD) to identify the slow mode. To address the slow computation of mrDMD, we provide a new CUDA implementation. We demonstrated that our DMD based strategy can help better reveal the large-scale structures from a number of 2D and 3D unsteady shear flows than existing methods. Finally, we apply our method to some non-shear flows that do not behave quasi-linearly. Our results show that DMD fails to capture the essential behavior in those flows, which further suggests that *DMD should be applied primarily to quasi-linear flows in order to reveal physically meaningful structures.*

Although our work shows the promising use of DMD in identifying the large-scale structures from shear flows, there are a few limitations that we aim to improve. First, the structure revealed by the DMD modes need not be the structure of interest to the experts. In particular, DMD can be biased to structures that are better characterized by the attributes used for DMD computation (i.e., velocity field in our work). For example, in the 3D Plane Couette flow, while the large-scale structures of interest are the Taylor rolls, our DMD method reveals the separations structures between the individual rolls. This is because the flow inside the Taylor roll has small velocity in comparison with that at the boundary of two rolls. Second, the information encoded by the slowest DMD mode along with its time evolution plot need not fully represent the large-scale structures. This is apparent because the patterns in the reconstructed field using a DMD mode are fixed with only varying amplitude over time. This cannot capture the movement or deformation of the structure over time. A possible way to address this is rather than using just one slow mode to characterize the large-scale structure, we choose a number of slow modes that complement each other over time to represent the large-scale structures. mrDMD also provides a hierarchical mechanism with multiple levels of modes. In this version, we only utilize the slow mode in the first level which is suitable for the large-scale coherent extraction task. Investigating the sub-modes are also promising to characterize the hierarchical behavior of flow features. Finally, even though we implement a fast mrDMD using CUDA, it does not address the memory constraint for large-scale data. In the future, an out-of-core implementation will be needed for it to process the large-scale 3D turbulence flows.

7 Conclusion

This dissertation has addressed two main challenges in the coherent structure analysis for turbulent flow study. The first challenge is the gap between the geometric visualization and physical characteristics. The geometric representation does not always capture the relevant physical features that are interesting to the domain experts. To address that, we have proposed to encode relevant physical attributes into the geometric representation of flow (e.g., pathlines) through their respective time-activity curves (TACs). The second challenge comes from the multi-scale coherent structure extraction task. Two new analysis and visualization frameworks have been developed to enable the clear separation of large- and small- scale structures in the wall-bounded shear flows.

For the physics-aware selection of geometric representation of flows, we have devised an interactive visualization system for the analysis and exploration of unsteady flows based on TACs. Given a vector field, we first computed the TACs over the entire flow domain and applied time interval segmentation to all TACs. To describe the behavior of a TAC, a sub-TAC extraction method has been introduced to identify one or more interesting temporal trends. To measure the similarity of two TACs properly, we have introduced a new distance metric, called the TAC Similarity Metric (TSM) to calculate the dissimilarity of TACs based on their events.

We have implemented the Agglomerative Hierarchical Clustering algorithm with the new TSM distance metric for the clustering of Lagrangian TACs on different temporal intervals. The clustering results provide different levels of details for flow behaviors in both space and time, which facilitate the exploration of different flow behaviors over time. We have also improved the edge-bundling technique to better represent the general behavior of TACs in a cluster and the connection of clusters among different time intervals. We have introduced a 2D stack plot to visualize the TAC clusters without occlusion. This TAC-based framework has been evaluated with multiple unsteady flow simulations to help domain experts analyze vortex structures and other flow features.

To extract multi-scale coherent structures for the Taylor-Couette flows, we have presented a two-stage framework. First, we derived physical attributes and found a combination among these attributes via an improved feature level-set technique to better capture the difference between the large- and small- scale features. Second, we extracted the iso-surface from the kernel density estimation of the distance field obtained from the feature level-set computation. We have also provided a 2D abstract representation through the plane projection along the streamwise direction to highlight the concentrated positions of structures with different scales. This method is simple yet effective, enabling the separation of the large-scale structures from the smaller ones. It provides a clear 3D visualization of Taylor-Couette flows for the first time, facilitating its analysis by the experts. The main drawbacks of the method is that it is still threshold (or parameter) dependent.

To overcome the parameter dependent issue of the second work, we have proposed to use Multi-Resolution Dynamic Mode Decomposition (mrDMD) to extract large-scale structures and track their temporal behavior in shear flows. We have demonstrated that the slow motion DMD modes extracted by mrDMD possess the relevant characteristics to the large-scale structures in shear flows. Thus, the slow modes can be used to capture the large-scale features in shear flows. To speed up the computation of mrDMD, a fast GPU-based implementation has been provided. We have introduced a visualization framework built on the fast mrDMD computation which can help experts effectively analyze the obtained DMD modes and their time dynamics for the study of the large-scale structures in shear flows. We have also performed experiments with other types of flows than shear flows using the proposed DMD-based framework to demonstrate its potential extension. The outcome of this work not only leads to an effective framework for the large-scale coherent structure study for shear flows but also provides a first guideline on the use of DMD technique for multi-scale analysis of general turbulent flows for the visualization community.

Limitations & Future Work. There are a number of limitations of the presented TAC-based analysis and visualization framework. First, the clustering computation is slow as the AHC computation has not been fully optimized, thus, it limits the framework's scalability for large-scale data. Second, the current TAC-based framework concentrates on scalar attributes. However, it may be extended to other attribute types, such as vector-valued and tensor-valued attributes. The extension to the vector-valued attribute has been demonstrated. However, the visualization of the clustering results based on vector-valued (and later tensor-valued) attributes in the TAC space needs to be addressed. Third, the 2D stack plots can provide a summary view of the global clustering results, but they cannot properly visualize the transition between neighboring temporal clusters, which we plan to address. Finally, this framework has been evaluated via the vortex structure analysis and understanding and other relevant features such as shearing layers and symmetric behavior. In the future, it is important to extend it for the study of energy transport in turbulent flow.

For the coherent structure extraction work for the Taylor-Couette turbulent flows, despite the fact that we can separate regions with small-scale features from those with large-scale ones, the presented approach still has some limitations. First, the quality of the extracted surface representation for large-scale structures still depends on the proper selection of the value ranges of certain relevant attributes. Second, the extracted surface representation of the large-scale structures only provides an approximation on the regions where those structures may reside. In the meantime, the expert wishes to see the more accurate transport barriers of this flow, i.e., the boundaries of Taylor rolls, to quantify their dynamics. Using DMD can overcome the issue with the threshold selection. However, only one slowest mode is selected for the coherent structure analysis task. The use of higher level and other DMD modes is still an open question. Also, the extraction of the coherent structures from the selected DMD modes is then to be completed. We plan to address these limitations in the future work.

References

[1] Avsarkisov, V., Hoyas, S., Oberlack, M., and García-Galache, J. Turbulent plane Couette flow at moderately high Reynolds number. *Journal of Fluid Mechanics 751* (2014), R1.

[2] Batchelor, G. K. *The Theory of Homogeneous Turbulence.* Cambridge University Press, Cambridge, England, 1982, ch. 3, pp. 34–38.

[3] Berenjkoub, M., Monico, R. O., Laramee, R. S., and Chen, G. Visual analysis of spatia-temporal relations of pairwise attributes in unsteady flow. *IEEE Transactions on Visualization and Computer Graphics 25*, 1 (2019), 1246–1256.

[4] Bhatia, H., Norgard, G., Pascucci, V., and Bremer, P. The Helmholtz-Hodge decomposition—a survey. *IEEE Transactions on Visualization and Computer Graphics 19*, 8 (2013), 1386–1404.

[5] Bodenschatz, E., Pesch, W., and Ahlers, G. Recent developments in Rayleigh-Bénard convection. *Annual Review of Fluid Mechanics 32*, 1 (2000), 709–778.

[6] Cabral, B., and Leedom, L. C. Imaging vector fields using line integral convolution. In *Proceedings of the 20th Annual Conference on Computer Graphics and Interactive Techniques* (New York, USA, 1993), SIGGRAPH '93, Association for Computing Machinery, p. 263–270.

[7] Carnecky, R., Brunner, T., Born, S., Waser, J., Heine, C., and Peikert, R. Vortex detection in 4D MRI data: Using the proper orthogonal decomposition for improved noise-robustness. In *Proceedings of the 35th Annual Conference of the European Association for Computer Graphics* (Strasbourg, France, 2014), EuroVis - Short Papers, The Eurographics Association, pp. 127–131.

[8] Chen, G., Deng, Q., Szymczak, A., Laramee, R. S., and Zhang, E. Morse set classification and hierarchical refinement using Conley index. *IEEE Transactions on Visualization and Computer Graphics 18*, 5 (2012), 767–782.

[9] Chen, G., Mischaikow, K., Laramee, R. S., Pilarczyk, P., and Zhang, E. Vector field editing and periodic orbit extraction using Morse decomposition. *IEEE Transactions on Visualization and Computer Graphics 13*, 4 (2007), 769–785.

[10] Chen, G., Mischaikow, K., Laramee, R. S., and Zhang, E. Efficient Morse decompositions of vector fields. *IEEE Transactions on Visualization and Computer Graphics 14*, 4 (2008), 848–862.

[11] Chong, M. S., Perry, A. E., and Cantwell, B. J. A general classification of three-dimensional flow fields. *Physics of Fluids A: Fluid Dynamics 2*, 5 (1990), 765–777.

[12] de Leeuw, W., and van Liere, R. Collapsing flow topology using area metrics. In *Proceedings of the Conference on Visualization '99: Celebrating Ten Years* (Washington, DC, USA, 1999), VIS '99, IEEE Computer Society Press, p. 349–354.

[13] Defays, D. An efficient algorithm for a complete link method. *The Computer Journal 20*, 4 (1977), 364–366.

[14] Edmunds, M., Laramee, R. S., Chen, G., Max, N., Zhang, E., and Ware, C. Surface-based flow visualization. *Computers & Graphics 36*, 8 (2012), 974–990.

[15] Esling, P., and Agon, C. Time-series data mining. *ACM Computing Surveys (CSUR) 45*, 1 (2012), 12.

[16] Fang, Z., Möller, T., Hamarneh, G., and Celler, A. Visualization and exploration of time-varying medical image data sets. In *Proceedings of Graphics Interface 2007* (New York, NY, USA, 2007), GI '07, ACM, pp. 281–288.

[17] Farge, M. Wavelet transforms and their applications to turbulence. *Annual Review of Fluid Mechanics 24*, 1 (1992), 395–458.

[18] Farge, M., Kevlahan, N., Perrier, V., and Goirand, E. Wavelets and turbulence. *Proceedings of the IEEE 84*, 4 (1996), 639–669.

[19] Farge, M., Pellegrino, G., and Schneider, K. Coherent vortex extraction in 3D turbulent flows using orthogonal wavelets. *Physical Review Letters 87*, 5 (2001), 054501.

[20] Farooq, S., Huarte-Espinosa, M., and Ostilla-Mónico, R. Large-scale structures in high-Reynolds-number rotating Waleffe flow. *Journal of Fluid Mechanics 884* (2019).

[21] Ferrari, S., Hu, Y., and Martinuzzi, R. J. Evolution surfaces for spatiotemporal visualization of vortex features. *Canadian Journal of Electrical and Computer Engineering 43*, 1 (2020), 30–42.

[22] Ferstl, F., Kanzler, M., Rautenhaus, M., and Westermann, R. Time-hierarchical clustering and visualization of weather forecast ensembles. *IEEE Transactions on Visualization and Computer Graphics 23*, 1 (2017), 831–840.

[23] Frisch, U. *Turbulence: The Legacy of A. N. Kolmogorov*, 1 ed. Cambridge University Press, Cambridge, England, 2010, ch. 2, pp. 86–87.

[24] Fuchs, R., Kemmler, J., Schindler, B., Waser, J., Sadlo, F., Hauser, H., and Peikert, R. Toward a Lagrangian vector field topology. *Computer Graphics Forum 29* (2010), 1163–1172.

[25] Fuchs, R., Peikert, R., Hauser, H., Sadlo, F., and Muigg, P. Parallel vectors criteria for unsteady flow vortices. *IEEE Transactions on Visualization and Computer Graphics 14*, 3 (2008), 615–626.

[26] G. Gan, C. M., and Wu, J. *Data Clustering: Theory, Algorithms, and Applications.* Society for Industrial and Applied Mathematics, Philadelphia, Pennsylvania, USA, 2007, ch. 3, pp. 128–137.

[27] Garth, C., Wiebel, A., Tricoche, X., Joy, K. I., and Scheuermann, G. Lagrangian visualization of flow-embedded surface structures. *Computer Graphics Forum 27*, 3 (2008), 1007–1014.

[28] Gilka, G., Luchtenburg, D., Thiele, F., and Morzynski, M. Dynamic characterization of an actuated bluff body wake. In *Proceedings of the Fifth European Conference on Computational Fluid Dynamics* (Lisbon, Portugal, 2010), ECCOMAS CFD, APMTAC, p. 1557.

[29] Guennebaud, G., Jacob, B., et al. Eigen v3. http://eigen.tuxfamily.org, 2010.

[30] Gunther, T., and Theisel, H. Hyper-objective vortices. *IEEE Transactions on Visualization and Computer Graphics 26*, 3 (2020), 1532 – 1547.

[31] Guo, H., Renaut, R., Chen, K., and Reiman, E. Clustering huge data sets for parametric pet imaging. *BioSystems 71*, 1 (2003), 81–92.

[32] Günther, T., and Theisel, H. The state of the art in vortex extraction. *Computer Graphics Forum 37*, 6 (2018), 149–173.

[33] Haller, G. Lagrangian coherent structures and the rate of strain in two-dimensional turbulence. *Physics of Fluids 13*, 11 (2001), 3365–3385.

[34] Haller, G. An objective definition of a vortex. *Journal of Fluid Mechanics 525* (2005), 1–26.

[35] Haller, G., Hadjighasem, A., Farazmand, M., and Huhn, F. Defining coherent vortices objectively from the vorticity. *Journal of Fluid Mechanics 795* (2016), 136–173.

[36] Helman, J. L., and Hesselink, L. Representation and display of vector field topology in fluid flow data sets. *IEEE Computer 22*, 8 (1989), 27–36.

[37] Huisman, S. G., van der Veen, R. C. A., Sun, C., and Lohse, D. Multiple states in highly turbulent Taylor-Couette flow. *Nature Communications 5*, 1 (2014), 3820.

[38] Hunt, J. Vorticity and Vortex Dynamics in Complex Turbulent Flows. *Transactions of the Canadian Society for Mechanical Engineering 11*, 1 (1987), 21–35.

[39] Hunt, J. C., Wray, A., and Moin, P. Eddies, streams, and convergence zones in turbulent flows. In *Proceedings of the 1988 Summer Program - Studying Turbulence Using Numerical Simulation Databases* (Stanford, CA, 1988), Ames Research Center, pp. 193–208.

[40] Jankowai, J., and Hotz, I. Feature level-sets: Generalizing iso-surfaces to multi-variate data. *IEEE Transactions on Visualization and Computer Graphics 26*, 2 (2020), 1308–1319.

[41] Jeong, J., and Hussain, F. On the identification of a vortex. *Journal of Fluid Mechanics 285* (1995), 69–94.

[42] Jovanović, M. R., Schmid, P. J., and Nichols, J. W. Sparsity-promoting dynamic mode decomposition. *Physics of Fluids 26*, 2 (2014), 024103.

[43] Kasten, J., Reininghaus, J., Hotz, I., and Hege, H.-C. Two-dimensional time-dependent vortex regions based on the acceleration magnitude. *IEEE Transactions on Visualization and Computer Graphics 17*, 12 (2011), 2080–2087.

[44] Kasten, J., Zoufahl, A., Hege, H. C., and Hotz, I. Analysis of vortex merge graphs. In *Proceedings of the Conference on Vision, Modeling and Visualization* (Magdeburg, Germany, 2012), VMV, pp. 111–118.

[45] Kerschen, G., Golinval, J.-c., Vakakis, A. F., and Bergman, L. A. The method of proper orthogonal decomposition for dynamical characterization and order reduction of mechanical systems: An overview. *Nonlinear Dynamics 41*, 1 (2005), 147–169.

[46] Krug, D., Yang, X. I. A., de Silva, C. M., Ostilla-Mónico, R., Verzicco, R., Marusic, I., and Lohse, D. Statistics of turbulence in the energy-containing range of taylor–couette compared to canonical wall-bounded flows. *Journal of Fluid Mechanics 830*, 10 (2017), 797–819.

[47] Kutz, J. N., Brunton, S. L., Brunton, B. W., and Proctor, J. L. *Dynamic Mode Decomposition: Data-Driven Modeling of Complex Systems*. Society for Industrial and Applied Mathematics, Philadelphia, PA, 2016, ch. 1, p. 22.

[48] Kutz, J. N., Fu, X., and Brunton, S. L. Multi-resolution dynamic mode decomposition. *SIAM Journal on Applied Dynamical Systems 15*, 2 (2016), 713–735.

[49] Laramee, R. S., Hauser, H., Doleisch, H., Post, F. H., Vrolijk, B., and Weiskopf, D. The state of the art in flow visualization: dense and texture-based techniques. *Computer Graphics Forum 23*, 2 (2004), 203–221.

[50] Laramee, R. S., Hauser, H., Zhao, L., and Post, F. H. Topology-based flow visualization, the state of the art. In *Topology-based Methods in Visualization* (Berlin, Heidelberg, 2007), Springer, pp. 1–19.

[51] Lee, T., and Shen, H. Visualization and exploration of temporal trend relationships in multivariate time-varying data. *IEEE Transactions on Visualization and Computer Graphics 15*, 6 (2009), 1359–1366.

[52] Lee, T.-Y., and Shen, H.-W. Visualizing time-varying features with tac-based distance fields. In *Proceedings of the 2009 IEEE Pacific Visualization Symposium* (Beijing, China, 2009), PacificVis, IEEE, pp. 1–8.

[53] Leeuw, W. d., and Liere, R. v. Visualization of global flow structures using multiple levels of topology. In *Proceedings of the Joint EUROGRAPHICS and IEEE TCVG Symposium on Visualization* (1999), Springer and The Eurographics Association, pp. 102–134.

[54] Leishman, J. G. *Principles of Helicopter Aerodynamics*, 2 ed. Cambridge University Press, Cambridge, UK, 2016, ch. 4, pp. 167–168.

[55] Liao, T. W. Clustering of time series data — a survey. *Pattern Recognition 38*, 11 (2005), 1857–1874.

[56] Lugt, H. J. *The Dilemma of Defining a Vortex*. Springer, Berlin, Heidelberg, 1979, pp. 309–321.

[57] Marston, J. B., Chini, G. P., and Tobias, S. M. The generalized quasilinear approximation: Application to Zonal jets. *Physical Review Letters 116*, 21 (2016), 214501.

[58] McLoughlin, T., Jones, M. W., Laramee, R. S., Malki, R., Masters, I., and Hansen, C. D. Similarity measures for enhancing interactive streamline seeding. *IEEE Transactions on Visualization and Computer Graphics 19*, 8 (2013), 1342–1353.

[59] McLoughlin, T., Laramee, R. S., Peikert, R., Post, F. H., and Chen, M. Over two decades of integration-based, geometric flow visualization. *Computer Graphics Forum 29*, 6 (2010), 1807–1829.

[60] Muzy, J.-F., Bacry, E., and Arneodo, A. Wavelets and multifractal formalism for singular signals: Application to turbulence data. *Physical Review Letters 67*, 25 (1991), 3515.

[61] Nguyen, D. B., Monico, R. O., and Chen, G. A Visualization Framework for Multi-scale Coherent Structures in Taylor-Couette Turbulence. *IEEE Transactions on Visualization and Computer Graphics 27* (2020), 1–1.

[62] Nvidia. Cuda toolkit documentation. https://docs.nvidia.com/cuda/.

[63] Oesterling, P., Heine, C., Janicke, H., Scheuermann, G., and Heyer, G. Visualization of high-dimensional point clouds using their density distribution's topology. *IEEE Transactions on Visualization and Computer Graphics 17*, 11 (2011), 1547–1559.

[64] Orlandi, P., and Verzicco, R. Vortex rings impinging on walls: axisymmetric and three-dimensional simulations. *Journal of Fluid Mechanics 256* (1993), 615–646.

[65] Ostilla Monico, R. *Highly turbulent Taylor-Couette flow: direct numerical simulations.* PhD thesis, University of Twente, 2 2015.

[66] Ostilla-Mónico, R., Verzicco, R., Grossmann, S., and Lohse, D. The near-wall region of highly turbulent Taylor–Couette flow. *Journal of Fluid Mechanics 788* (2016), 95–117.

[67] Ostilla-Mónico, R., Verzicco, R., and Lohse, D. Effects of the computational domain size on direct numerical simulations of Taylor-Couette turbulence with stationary outer cylinder. *Physics of Fluids 27*, 2 (2015), 110–115.

[68] Palmas, G., Bachynskyi, M., Oulasvirta, A., Seidel, H. P., and Weinkauf, T. An edge-bundling layout for interactive parallel coordinates. In *Proceedings of the 2014 IEEE Pacific Visualization Symposium* (Yokohama, Japan, 2014), PacificVis '14, IEEE Computer Society, p. 57–64.

[69] Parzen, E. On estimation of a probability density function and mode. *The Annals of Mathematical Statistics 33*, 3 (1962), 1065–1076.

[70] Pascarella, G., Kokkinakis, I., and Fossati, M. Analysis of transition for a flow in a channel via reduced basis methods. *Fluids 4*, 4 (2019), 202.

[71] Peikert, R., and Roth, M. The "parallel vectors" operator: a vector field visualization primitive. In *Proceedings of the 10th IEEE Visualization 1999 Conference* (Los Alamitos, CA, USA, 1999), VIS '99, IEEE Computer Society Press, pp. 263–270.

[72] Pendergrass, S. D., Kutz, J. N., and Brunton, S. L. Streaming GPU singular value and dynamic mode decompositions. *arXiv e-prints* (Dec. 2016), arXiv:1612.07875.

[73] Pobitzer, A., Lez, A., Matkovic, K., and Hauser, H. A statistics-based dimension reduction of the space of path line attributes for interactive visual flow analysis. In *Proceedings of the 2012 IEEE Pacific Visualization Symposium* (Songdo, Korea, 2012), PacificVis, IEEE, pp. 113–120.

[74] Pobitzer, A., Peikert, R., Fuchs, R., Schindler, B., Kuhn, A., Theisel, H., Matkovic, K., and Hauser, H. The state of the art in topology-based visualization of unsteady flow. *Computer Graphics Forum 30*, 6 (2011), 1789–1811.

[75] Pobitzer, A., Tutkun, M., Andreassen, , Fuchs, R., Peikert, R., and Hauser, H. Energy-scale aware feature extraction for flow visualization. *Computer Graphics Forum 30*, 3 (2011), 771–780.

[76] Polthier, K., and Preuss, E. Identifying vector field singularities using a discrete Hodge decomposition. In *Visualization and Mathematics III* (Berlin, Heidelberg, 2003), Springer, pp. 113–134.

[77] Robert, T. Multi-resolution dmd. http://www.pyrunner.com/weblog/2016/08/05/mrdmd-python/, 08 2016.

[78] Robertson, E. D., Wang, Y., Pant, K., Grismer, M. J., and Camberos, J. A. A flow feature detection framework for large-scale computational data based on incremental proper orthogonal decomposition and data mining. *International Journal of Computational Fluid Dynamics 32*, 6-7 (2018), 261–277.

[79] Robinson, S. K. Coherent motions in the turbulent boundary layer. *Annual Review of Fluid Mechanics 23*, 1 (1991), 601–639.

[80] Sacco, F., Verzicco, R., and Ostilla-Mónico, R. Dynamics and evolution of turbulent Taylor rolls. *Journal of Fluid Mechanics 870* (2019), 970–987.

[81] Sacco, F., Verzicco, R., and Ostilla-Mónico, R. Dynamics and evolution of turbulent taylor rolls. *Journal of Fluid Mechanics 870* (2019), 970–987.

[82] Sadarjoen, I. A., and Post, F. H. Geometric methods for vortex extraction. In *Proceedings of the Joint EUROGRAPHICS and IEEE TCVG Symposium on Visualization* (Vienna, 1999), Springer, pp. 53–62.

[83] Sadarjoen, I. A., and Post, F. H. Detection, quantification, and tracking of vortices using streamline geometry. *Computers & Graphics 24*, 3 (2000), 333–341.

[84] Sadlo, F., and Peikert, R. Efficient visualization of Lagrangian coherent structures by filtered AMR ridge extraction. *IEEE Transactions on Visualization and Computer Graphics 13*, 6 (2007), 1456–1463.

[85] Sadlo, F., and Weiskopf, D. Time-dependent 2D vector field topology: An approach inspired by Lagrangian coherent structures. *Computer Graphics Forum 29*, 1 (2010), 88–100.

[86] Sakurai, T., Yoshimatsu, K., Schneider, K., Farge, M., Morishita, K., and Ishihara, T. Coherent structure extraction in turbulent channel flow using boundary adapted wavelets. *Journal of Turbulence 18*, 4 (2017), 352–372.

[87] Salzbrunn, T., Garth, C., Scheuermann, G., and Meyer, J. Pathline predicates and unsteady flow structures. *The Visual Computer 24*, 12 (2008), 1039–1051.

[88] Salzbrunn, T., and Scheuermann, G. Streamline predicates. *IEEE Transactions on Visualization and Computer Graphics 12*, 6 (2006), 1601–1612.

[89] Salzbrunn, T., Wischgoll, T., Jänicke, H., and Scheuermann, G. The state of the art in flow visualization: Partition-based techniques. In *In Simulation and Visualization 2008 Proceedings* (Magdeburg, Germany, 2008), SimVis, pp. 75–92.

[90] Sayadi, T., and Schmid, P. J. Parallel data-driven decomposition algorithm for large-scale datasets: with application to transitional boundary layers. *Theoretical and Computational Fluid Dynamics 30*, 5 (2016), 415–428.

[91] Schaeffer, N., and Le Dizès, S. Nonlinear dynamics of the elliptic instability. *Journal of Fluid Mechanics 646* (2010), 471–480.

[92] Schmid, P. J. Dynamic mode decomposition of numerical and experimental data. *Journal of Fluid Mechanics 656*, 8 (2010), 5–28.

[93] Shi, K., Theisel, H., Hauser, H., Weinkauf, T., Matkovic, K., Hege, H.-C., and Seidel, H.-P. Path line attributes - an information visualization approach to analyzing the dynamic behavior of 3d time-dependent flow fields. In *Topology-Based Methods in Visualization II* (Grimma, Germany, 2009), Mathematics and Visualization, Springer, pp. 75–88.

[94] Sillero, J. A., Jiménez, J., and Moser, R. D. Two-point statistics for turbulent boundary layers and channels at Reynolds numbers up to 2000. *Physics of Fluids 26*, 10 (2014), 105–109.

[95] Silverman, B. *Density Estimation for Statistics and Data Analysis.* Taylor & Francis, Oxfordshire UK, 1986, ch. 5, pp. 107–108.

[96] Skraba, P., Rosen, P., Wang, B., Chen, G., Bhatia, H., and Pascucci, V. Critical point cancellation in 3d vector fields: Robustness and discussion. *IEEE Transactions on Visualization and Computer Graphics 22*, 6 (2016), 1683–1693.

[97] Skraba, P., Wang, B., Chen, G., and Rosen, P. Robustness-based simplification of 2D steady and unsteady vector fields. *IEEE Transactions on Visualization and Computer Graphics 21*, 8 (2014), 930–944.

[98] Strogatz, S. H. *Nonlinear Dynamics And Chaos: With Applications To Physics, Biology, Chemistry, And Engineering*, 1 ed. CRC Press, Boca Raton, Florida, USA, 2000, pp. 46–64.

[99] Swinney, H. L., and Gollub, J. P. *Hydrodynamic instabilities and the transition to turbulence.* Springer-Verlag, Berlin, Germany, 1985, pp. 139–180.

[100] Szymczak, A., and Zhang, E. Robust Morse decompositions of piecewise constant vector fields. *IEEE Transactions on Visualization and Computer Graphics 18*, 6 (2012), 938–951.

[101] Theisel, H., Rössl, C., and Seidel, H.-P. Compression of 2d vector fields under guaranteed topology preservation. *Computer Graphics Forum 22*, 3 (2003), 333–342.

[102] Theisel, H., Weinkauf, T., christian Hege, H., and peter Seidel, H. Grid-independent detection of closed stream lines in 2D vector fields. In *Proceedings of the Conference on Vision, Modeling and Visualization* (Stanford, California, USA, 2004), VMV, pp. 204–212.

[103] Tobias, S. M., and Marston, J. B. Three-dimensional rotating couette flow via the generalised quasilinear approximation. *Journal of Fluid Mechanics 810* (2017), 412–428.

[104] Treib, M., Bürger, K., Reichl, F., Meneveau, C., Szalay, A., and Westermann, R. Turbulence visualization at the terascale on desktop PCs. *IEEE Transactions on Visualization and Computer Graphics 18*, 12 (2012), 2169–2177.

[105] Treib, M., Bürger, K., Reichl, F., Meneveau, C., Szalay, A., and Westermann, R. Turbulence Visualization at the Terascale on Desktop PCs. *IEEE Transactions on Visualization and Computer Graphics 18*, 12 (2012), 2169–2177.

[106] Tricoche, X., Scheuermann, G., and Hagen, H. A topology simplification method for 2D vector fields. In *Proceedings of IEEE Visualization* (Los Alamitos, CA, USA, 2000), IEEE Computer Society Press, pp. 359–366.

[107] Tricoche, X., Scheuermann, G., and Hagen, H. Continuous topology simplification of planar vector fields. In *Proceedings of the Conference on Visualization* (San Diego, California, USA, 2001), VIS '01, IEEE Computer Society, p. 159–166.

[108] Tu, H. J., Rowley, W. C., Luchtenburg, M. D., Brunton, L. S., and Kutz, J. N. On dynamic mode decomposition: Theory and applications. *Journal of Computational Dynamics 1*, 2 (2014), 391–421.

[109] Üffinger, M., Sadlo, F., and Ertl, T. A time-dependent vector field topology based on streak surfaces. *IEEE Transactions on Visualization and Computer Graphics 19*, 3 (2013), 379–392.

[110] Vamsi, S., Rodolfo, O.-M., Detlef, L., and Roberto, V. Identifying coherent structures and vortex clusters in Taylor-Couette turbulence. *Journal of Physics: Conference Series 708* (2016), 012006.

[111] van der Poel, E. P., Ostilla-Mónico, R., Donners, J., and Verzicco, R. A pencil distributed finite difference code for strongly turbulent wall-bounded flows. *Computers Fluids 116* (2015), 10 – 16.

[112] von Funck, W., Weinkauf, T., Theisel, H., and Seidel, H.-P. Smoke surfaces: An interactive flow visualization technique inspired by real-world flow experiments. *IEEE Transactions on Visualization and Computer Graphics 14*, 6 (2008), 1396–1403.

[113] Walter, B., Bala, K., Kulkarni, M., and Pingali, K. Fast agglomerative clustering for rendering. In *Proceedings of the IEEE Symposium on Interactive Ray Tracing* (Los Angeles, CA, USA, 2008), IEEE, pp. 81 – 86.

[114] Wang, C., Yu, H., and Ma, K.-L. Importance-driven time-varying data visualization. *IEEE Transactions on Visualization and Computer Graphics 14*, 6 (2008), 1547–1554.

[115] Wang, Y., Qian, J., Song, H., Pant, K., Yang, H. Q., Li, X., Grismer, M. J., Camberos, J. A., and Fahroo, F. Feature extraction from massive, dynamic computational data based on proper orthogonal decomposition and feature mining. *Journal of Visualization 17*, 4 (Nov 2014), 363–372.

[116] Wei, J., Yu, H., Grout, R. W., Chen, J. H., and Ma, K.-L. Dual space analysis of turbulent combustion particle data. In *Proceedings of the 2011 IEEE Pacific Visualization Symposium* (Hong Kong, 2011), PacificVis, IEEE, pp. 91–98.

[117] Weinkauf, T., and Theisel, H. Streak lines as tangent curves of a derived vector field. *IEEE Transactions on Visualization and Computer Graphics 16*, 6 (2010), 1225–1234.

[118] Weinkauf, T., Theisel, H., Shi, K., Hege, H.-C., and Seidel, H.-P. Extracting higher order critical points and topological simplification of 3D vector fields. In *Proceedings of IEEE Visualization* (Minneapolis, MN, USA, 2005), IEEE Computer Society Press, pp. 559–566.

[119] Wiebel, A., Chan, R., Wolf, C., Robitzki, A., Stevens, A., and Scheuermann, G. *Topological Flow Structures in a Mathematical Model for Rotation-Mediated Cell Aggregation.* Springer, Berlin, Heidelberg, 2011, pp. 193–204.

[120] Wischgoll, T., and Scheuermann, G. Detection and visualization of closed streamlines in planar fields. *IEEE Transactions on Visualization and Computer Graphics 7*, 2 (2001), 165–172.

[121] Wong, K.-P., Feng, D., Meikle, S., and Fulham, M. Segmentation of dynamic PET images using cluster analysis. *IEEE Nuclear Science Symposium 3* (200), 126 – 130.

[122] Woodring, J., and Shen, H.-W. Multiscale time activity data exploration via temporal clustering visualization spreadsheet. *IEEE Transactions on Visualization and Computer Graphics 15*, 1 (2009), 123–137.

[123] Woodring, J., and Shen, H.-W. Semi-automatic time-series transfer functions via temporal clustering and sequencing. In *Proceedings of the 11th Eurographics / IEEE - VGTC Conference on Visualization* (Chichester, UK, 2009), EuroVis '09, pp. 791–798.

[124] Wu, Z., Laurence, D., Utyuzhnikov, S., and Afgan, I. Proper orthogonal decomposition and dynamic mode decomposition of jet in channel crossflow. *Nuclear Engineering and Design 344* (2019), 54–68.

[125] Yaglom, A., and Monin, A. *Statistical Fluid Mechanics, Volume II: Mechanics of Turbulence*, vol. 2. Dover Publications, Mineola, NY, USA, 2007, ch. 3, pp. 186–187.

[126] Zhang, E., Mischaikow, K., and Turk, G. Vector field design on surfaces. *ACM Transactions on Graphics 25*, 4 (2006), 1294–1326.

[127] Zhang, L., Chen, G., Laramee, R. S., Thompson, D., and Sescu, A. Flow visualization based on A derived rotation fields. In *Visualization and Data Analysis* (San Francisco, California, USA, 2016), Society for Imaging Science and Technology, pp. 1–10.

[128] Zhang, L., Nguyen, D., Thompson, D., Laramee, R. S., and Chen, G. Enhanced vector field visualization via Lagrangian accumulation. *Computer and Graphics (special issue of CAD/Graphics 70* (2018), 224–234.

[129] Zhang, Q., Liu, Y., and Wang, S. The identification of coherent structures using Proper Orthogonal Decomposition and Dynamic Mode Decomposition. *Journal of Fluids and Structures 49* (2014), 53 – 72.