In Situ Magnetic Flux Vortex Visualization in Time-Dependent Ginzburg-Landau Superconductor Simulations

Hanqi Guo 1*, Tom Peterka 1†, Andreas Glatz 2‡

1) Mathematics and Computer Science Division, Argonne National Laboratory, Lemont, IL, USA
2) Materials Science Division, Argonne National Laboratory, Lemont, IL, USA

Abstract

We present an in situ visualization framework to capture comprehensive details of vortex dynamics in superconductor simulations. Vortices, which determine all electromagnetic properties of type-II superconductors, are extracted and tracked at the same time with GPU-based time-dependent Ginzburg-Landau superconductor simulations. The in situ workflow involves three parts: 1) a tightly-coupled GPU-accelerated algorithm that detects primitives for ambiguity-free vortex tracking, 2) a loosely-coupled task-parallel feature-tracking method, and 3) a web-based remote visualization tool for vortex dynamics analysis. Our design minimizes the data movement and storage, maximizes the resource utilization, and reduces the slowdown of the simulation. Our solution captures all vortex dynamics in the simulation, previously impossible with traditional post hoc methods. We also demonstrate in situ visualization cases that help scientists understand how vortices cut each other and recombine into new vortices, which are directly related to energy dissipation of superconducting materials.

Keywords: In situ visualization, superconductivity, feature extraction and tracking, GPU algorithm, web-based visualization.

1 INTRODUCTION

High temperature (type-II) superconductors, which can conduct current with zero electrical resistance at liquid nitrogen temperature, are widely used in power transmission, particle accelerators, and magnetic resonance imaging. A key phenomenon emerging in those systems is magnetic flux vortices, or simply vortices, which determine the energy dissipation and all electromagnetic responses in type-II superconductors. To design superconductors that can sustain higher electric currents for applications, materials scientists must understand and control the behavior of vortices.

An effective way to understand vortex behavior is to simulate the superconductors by using large-scale numerical solvers. Here we consider simulations based on the Ginzburg-Landau theory of superconductivity, which models the superconducting properties in terms of a complex-valued order parameter field \( \psi \in \mathbb{C} \). In this theory, vortices are defined as the singularities in \( \psi \), which are 3D curves that satisfy

\[
|\psi| = 0 \quad \text{and} \quad -\oint_C \nabla \theta \cdot d\mathbf{l} = 2n\pi,
\]

where \( \theta \) and \( |\psi| \) are the phase angle and magnitude of \( \psi \), respectively, \( C \) is a small closed contour that encircles the singularity, and \( n \) is usually \( \pm1 \) and indicates the chirality of the vortex with respect to \( C \). Notice that the definition of magnetic flux vortices is related to but different from that of fluid flow vortices. Vortices in superconductors are 3D curves, which are comparable to vortex core lines or swirling centers in fluid flows. In the rest of this paper, unless otherwise noted, the term vortex means magnetic flux....

* hguo@anl.gov
† tpeterka@mcs.anl.gov
‡ glatz@anl.gov
vortex.

The motivating scientific questions in this paper are when and how exactly vortices recombine with each other [12], specifically, the accurate time of the cuttings and the behaviors of vortices before and after such events. These questions require vortex tracking in high temporal resolution, because vortices move rapidly around events. Unfortunately, in current post hoc visualization workflow, scientists have to dramatically reduce the data written to the storage system because of prohibitive I/O bandwidth and capacity. In practice, $10^3 - 10^6$ timesteps are skipped in the simulation, in order not to slow the simulation and overflow the storage.

We extract and track vortices in situ—as opposed to traditional post hoc processing—to address the precision problem and reduce the I/O demand. As shown in Figure 1, the temporal resolution in traditional post hoc processing is usually insufficient to capture dynamics of interest in the vicinity of topological changes (so-called events). Such a problem can be solved only by accessing all timesteps during the simulation. As an added benefit, the I/O demand of writing vortices and analysis results, instead of raw simulation output, is dramatically reduced.

The entire in situ vortex visualization workflow consists of two parts—online processing and post hoc visualization. The online processing extracts, tracks, analyzes, and reduces vortices in situ with the simulation; the post hoc visualization provides an interactive user interface to explore and analyze the online processing results.

We design and implement the online processing procedure for the time-dependent Ginzburg-Landau (TDGL) simulations based on the characteristics of both simulation and vortex tracking algorithms. The simulation code, GLGPU, is a large-scale GPU-based partial differential equation (PDE) solver [24]. The output order parameter field is available in GPU memory after every iteration of the solver. The vortex tracking algorithm, which has already been used in post hoc analysis, consists of two stages—detecting primitives and transforming the primitives into vortices. The two stages are based on numerical analysis and graph traversal, respectively. In the first stage, we check whether every mesh face or edge is intersected with a vortex, because the intersections indicate the locations and movements of vortices. In the second stage, we transform the primitives into vortices and their trajectories over time that are modeled with connected graphs.

The major challenge in our application is that the GPU-based simulation runs much faster than the analysis algorithms do. We must design faster parallel and asynchronous algorithms, in order to minimize the slowdown of the simulation. The online processing hence consists of two components—the tightly-coupled, synchronous, GPU-accelerated primitive detector and the loosely-coupled, asynchronous, task-parallel vortex analyzer. The two components play the role of producer and consumer, respectively. The producer runs at every single iteration of the GPU-based simulation, without any data movement of the order parameter field. The primitives are then transferred to the consumer for vortex reconstruction and analysis. We run the consumer asynchronously as an independent process, in order to avoid slowing the simulation with the time-consuming graph analysis on CPUs. Furthermore, we accelerate the consumer with a novel task-parallel approach, because the vortex analyzer usually runs slower than the simulation and the primitive detector. The output analysis results are written to a database for post hoc visualization. In the performance benchmarks, the amount of output is only $10^3 - 10^6$ smaller than the original order parameter data.

The post hoc visualization enables users to explore and further analyze vortex dynamics that are drawn from the simulation. The tool not only contains traditional 3D visualizations but also provides 2D visualizations of events and distances between vortices, which help users explore when vortices recombine and how they move before and after the events. The web-based tool provides a remote visualization solution and works flexibly on workstations or mobile devices without installing any specific software.

Compared with previous studies on vortex tracking in superconductors [13, 21], our in situ solution enables users to discover more details in the simulations given the fine temporal resolutions. We also improve the performance of vortex tracking in two ways. First, we develop a GPU-based vortex primitive detection algorithm that can run 200 times faster than the CPU implementation. Second, we parallelize the graph-based vortex reconstruction algorithm, which can make full use of all available CPU cores on the node. In summary, the contributions of this paper are fourfold:

- An in situ vortex visualization framework for analyzing and understanding vortex dynamics in superconductors
- A GPU-accelerated algorithm to detect primitives for ambiguity-free vortex extraction and tracking
- A task-parallel feature-tracking method that extracts and tracks vortices asynchronously
- Applications of in situ analysis of vortex dynamics and macro-behaviors of superconductors

The remainder of this paper is organized as follows. Section 2 reviews the background of in situ visualization, vortex visualization, and TDGL simulations. Section 3 presents our in situ workflow, and Sections 4 and 5 detail the primitive detector and the vortex analyzer, respectively. Section 6 describes the post hoc visualization. Implementation details and performance benchmarks are in Sections 7 and 8, respectively. Section 9 demonstrates applications cases, followed by the conclusions in Section 10.

2 Background

We review the background in vortex extraction and tracking, in situ visualization, and TDGL simulations.

2.1 Vortex extraction and tracking

We distinguish the concepts of fluid flow vortices from magnetic flux vortices, which are fundamentally different. In fluid dynamics, vortices can be defined in various ways, such as $\lambda_2 [14, 26]$, to characterize swirling motions. Comprehensive literature reviews on fluid flow vortices are available in [15] and [23]. In general, fluid flow vortices are localized in two ways—regions and core lines. In the former, vortex regions are usually extracted by thresholding the derived scalar field such as $\lambda_2$. In the latter, a vortex core line is defined by a locus of points that satisfy certain criterion, such as extremum of vorticity magnitude. The parallel vectors operator [20] and the feature flow fields [27], respectively, are the established techniques to extract and track vortices in fluid flows.

In the Ginzburg-Landau theory, vortices are well defined as topological defects in a complex-valued order parameter field. The vortex extraction algorithm is based on this theorem in complex scalar fields: there must be an equal number of vortex entry and exit points for any given closed volume, such as a mesh cell in the simulation. As shown in Figure 2(a), one can first localize the punctured points, that is, entries and exits on all mesh faces with the line integral (Equation 1), and then connect the punctured points based on the graph that describes cell/face neighborhoods in the mesh [22]. The vortex tracking algorithm extends the vortex extraction into 4D (space and time), based on the fact that the above theorem still holds in 4D [13, 21]. As shown in Figure 2(b), the movement of a punctured point can be captured by detecting punctured virtual faces of the prism, that is, the space-time extrusion of the mesh face. Vortices and their moving trajectories can then be constructed based on the graph analysis (Figure 2(c)).
We review the two-stage algorithm of the magnetic flux vortex extraction and tracking: primitive detection and graph analysis. In the primitive detection stage, for two timesteps $i$ and $j$, we detect punctured faces (PFs) $F_{i}$ and space-time intersected edges (IEs) $E_{i}$. In the graph analysis stage, we construct vortices in two individual timesteps ($\{V_{i}^{k}\}$ and $\{V_{j}^{k}\}$) based on $F_{i}$ and $F_{j}$, respectively. We then build the association graph $A_{j}^{i}(k, \ell)$ based on $F_{i}$, $F_{j}$, and $E_{i}$. The nodes of the association matrix are $\{V_{i}^{k}\}$ and $\{V_{j}^{k}\}$, and the link $(k, \ell)$ indicates that two vortices $V_{i}^{k}$ and $V_{j}^{\ell}$ are associated. An event happens if a vortex has more than one link.

Compared with previous studies, in this paper we develop a GPU-accelerated primitive detection algorithm and a task-parallel graph analysis for vortex analysis, in order to enable efficient in situ processing with the simulation. We also propose a method to remove ambiguity cases in previous studies (see discussions in Section 4).

2.2 In situ visualization

As summarized by a recent comprehensive review [2], in situ visualization is motivated by multiple factors—the disparity between scientific computing and I/O rate, the increased temporal resolution needed for accurate data analysis, and the utilization of all computing resources. Our in situ workflow benefits the analysis of vortex dynamics in all three aspects. We categorize the related work on in situ visualization into applications, algorithms, and infrastructures.

In situ visualization techniques have been used in various scientific applications. For example, Yu et al. [33] visualized volume and particle data in combustion simulations. Topologies in combustion data can be extracted by using segmented merge trees in situ [16]. Woodring et al. [31] developed an in situ workflow to analyze eddies in the study of ocean-climate models. Fabian [10] used in situ processing to detect fragments in explosion simulations. All these studies aim to achieve higher temporal resolution of the data in order to extract time-critical features. In our application, an additional challenge is that the TDGL simulation runs much faster than the vortex analysis algorithms. We must redesign the workflow and accelerate the algorithms in our in situ processing.

Various algorithms are used for efficient in situ visualization. For example, the performance of in situ volume rendering depends on image compositing algorithms, such as binary swap [18] and 2-3 swap [34]. Explorable images [28], which can generate new volume rendering results with a small number of rendered images, can be used in in situ visualization. Similarly, path tubes can be visualized in situ with explorable images [32]. Ahrens et al. presented an image-based in situ visualization framework for interactive exploration [1]. In addition to rendering technologies, algorithms are proposed to select optimal numbers of time steps for in situ visualization [19]. GoldRush [35] uses idle resources for in situ processing. Bennett et al. [3] explored a method to offload the computations of merge trees into secondary resources. In our study, we decouple the heavyweight vortex analysis in an independent process for parallel and asynchronous processing, in order to reduce the slowdown of the simulation.

In situ infrastructures are bridges between the simulation code and visualization methods. Typical examples include ParaView Catalyst [11] and VisIt Libsim [30], which can help scientists and visualization practitioners couple the simulation with production visualization tools. Because visualization algorithms are usually I/O intensive, various I/O solutions are proposed for in situ processing, such as ADIOS [17], FlexIO [36], and DataSpaces [7]. Instead of using existing frameworks and I/O solutions, based on the characteristics of both our simulation and analysis algorithms, we customize our in situ workflow and store the output data in high performance databases.

Recently, a group of visualization researchers started the “in situ terminology project” [6], in order to formalize the description of in situ methods with proper terminologies. We follow this project to characterize our methods in several axes. Both integration type and data access are direct, because the primitive detection code is directly plugged into the TDGL simulations and shares the same address space. The proximity is twofold: the primitive detector shares the same GPU cores as the simulation, and the vortex analyzer uses CPUs on or off the node. The synchronization is hybrid: the primitive detector runs synchronously with the simulation, and the vortex analyzer runs asynchronously with the simulation. The operation controls are not applicable in our workflow, and the output type is explorable.

2.3 TDGL simulations

Until recently, TDGL simulations, which are computationally expensive, were limited to 2D [5] or small-scale 3D problems [9]. To study electromagnetic properties of type-II superconductors at larger scales, scientists have developed a TDGL implementation, GLGPU [24]. It is a CUDA-based finite-difference PDE solver that runs on a single node. The size of the model is currently limited by the GPU memory, but the time-varying output can be arbitrarily large.

In our study, we design in situ visualization workflows for GLGPU. The input parameters include the size of the model, configuration of material inclusions, external magnetic field, and external current. At every iteration, the output order parameter field is
available on the GPU memory in the format of real and imaginary parts. The simulation also generates other measurements such as voltage—the indicator of energy dissipation of the material. Because the simulation is GPU-based, we use GPU-accelerated algorithms to detect primitives and use idle CPU resources for complex vortex analysis.

3 In situ vortex visualization workflow
Our in situ visualization workflow (Figure 2) consists of two major components: the online processing and post hoc visualization.

The online processing is pipelined with the primitive detector and vortex analyzer, which play the role of a producer and consumer, respectively. The primitive detector, which is tightly-coupled with the simulation process, detects primitives (\(F \) and \(E\)) right after every iteration of the simulation. The vortex analyzer, which is loosely-coupled with the simulation, transforms the primitives into vortices, tracks the vortices over time, and analyzes the properties of vortices on the fly. The tracked vortices along with the analysis results are written to the storage for post hoc visualization.

The rationale of the producer-consumer design is threefold: reducing data movement, minimizing the slowdown of the simulation, and avoiding dependencies in the simulation code. First, the producer, or the primitive detector, reduces data movement between GPU and main memory. The size of the primitives is much smaller than that of the order parameter field; thus we copy back the primitives instead of order parameters and conserve scarce bandwidth. Second, the consumer, or the vortex analyzer, minimizes the slowdown of the simulation. At every time step, the simulation continues after an asynchronous request is posted to transmit the primitives. Further analysis will not interfere with the simulation at all. Third, our implementation, especially the vortex analyzer, is independent of the simulation code. Embedding visualization and analysis into simulation code causes extra overhead on management and communication. Our design finds a balance to couple highly optimized kernels with the GPU-based simulation and do the rest of the analysis independently. More details on the primitive detector and the vortex analyzer are in Sections 4 and 5, respectively.

The post hoc visualization enables users to explore and further analyze vortex dynamics that are drawn from the simulation. The tool not only contains traditional 3D visualizations but also provides 2D visualizations of events and distances between vortices, which help users explore when vortices recombine and how they move before and after the events. The web-based tool provides a remote visualization solution and works flexibly on workstations or mobile devices without installing any specific software.

4 Tightly-coupled and GPU-accelerated primitive detection
We develop a GPU-accelerated algorithm for the tightly-coupled vortex primitive detection in TDGL simulations. We also propose a method to use tetrahedral subdivision to eliminate ambiguities of vortex extraction in a hexahedral mesh.

4.1 Mesh subdivision
We detect primitives in tetrahedra instead of hexahedra to prevent ambiguities in vortex extraction and tracking. In previous studies that extract vortices in a hexahedral mesh, ambiguous cases could appear if two vortices penetrates the same hexahedron at the same time. For example, in Figure 3(a), four punctured faces are detected, but we cannot determine their correspondence. Because such ambiguities are guaranteed to not exist in tetrahedral meshes, we subdivide the mesh into tetrahedra and detect primitives in the new mesh. The theoretical foundation is Lemma 2 given in the Appendix.

We subdivide each hexahedron into six tetrahedra, as illustrated in Figures 3(b), (c), and (d). Other subdivision schemes are possible, but our subdivision is conformal, which means that the tetrahedra share edges with neighbor cells. In the subdivided mesh, we index the elements (cells, faces, edges) by the Cartesian coordinates \((i, j, k)\) of the corner node (A in Figure 3(b)) and the type. After removing the duplicated cases, there are 6 types of cells, 12 types of faces, and 7 types of edges in the mesh. Notice that special treatments are necessary for mesh boundaries. The mesh graph that records the cell-face face-edge adjacency can also be built implicitly in further analysis.

4.2 GPU-accelerated primitive detection
We use the GPU to detect primitives in the tetrahedral mesh for vortex extraction. The inputs are the order parameter field of the simulation, and the outputs are the punctured faces and intersected space-time edges for vortex extraction and tracking. Each GPU thread is in charge of a single mesh element (a face or a space-time edge), which can be indexed by a unique ID.

In CUDA, threads are executed in a thread block that typically consists of tens or hundreds of threads. Threads in the same block share a small piece of high speed memory, so-called shared memory. All threads can access the global memory on the GPU, which is much larger but slower than the shared memory. Outside the GPU is the host memory, and the bandwidth between the host memory and the global memory is limited. Our algorithm takes advantage of the GPU thread model and the memory hierarchy for accelerated tightly-coupled primitive detection.

4.2.1 Punctured face detection
The input data of punctured face detection—real and imaginary parts of the order parameter field—are available in the GPU global memory after every iteration of the simulation. We examine every mesh face if punctured by a vortex by computing the phase jump over the mesh boundary by the contour integral defined in Equation 1. If the contour integral is \(\pm 2\pi\), the face is punctured; the punctures located where both real and imaginary parts of the order parameter are zero.

In the parallel execution, each GPU thread is in charge of a single mesh face. A list is used to collect detection results. We use a memory buffer in the global memory to store the list. An atomic integer records the end location of the list. To reduce memory contention, for each CUDA thread block, we first store the results in shared memory and then merge them into the list in global memory after all threads are finished in the thread block. The list of punctured faces, each of which consists of the face ID, chirality, and the puncture location, is copied back to the host memory for further analysis.

We must carefully manage the memory footprint of our analysis, because the GPU memory budget is tight in the simulation. First, we use a GPU memory buffer that is allocated by the simulation in a time-sharing way. The GPU memory buffer is not used by the simulation during our analysis and has the same capacity as the order parameter field. In rare cases, such as the very early stage of the simulation that is not of scientific interest, the punctured face list might overflow the buffer, so we have to drop the current timestep in this case. Second, we store the list of punctured faces in a compact way. We use only 12 or 16 bytes for a punctured face, depending on the total number of faces in the mesh. If the mesh has fewer than \(2^{31}\) faces, we use 31 bits for the face ID and 1 bit for chirality, otherwise 63 bits for the face ID and 1 bit for chirality. The 63 bits suffice because there are not likely to be more than \(2^{63}\) faces in the current GLGPU implementation. We store the puncture location—the barycentric coordinates on the face—with two 32-bit floats, so a punctured face takes 12 or 16 bytes in total.
4.2.2 Intersected space-time edge detection

The input data of the intersected space-time edge detection are the phase field of the two adjacent iterations in the simulation, and the outputs are a list of intersected space-time edges. The detection of space-time edges is similar to that of punctured faces. For each edge in the mesh, we extrude the edge into the time dimension and compute the integral over the space-time contour. The space-time edge is intersected if the contour integral equals $±2\pi$.

Similar to the punctured face detection, each GPU thread is in charge of an edge, and we use the shared memory to avoid contentions in the global memory on the GPU. Each intersected space-time edge is a tuple of the edge ID and chirality, which can be encoded with 4 or 8 bytes depending on the total number of edges in the mesh. Upon completing the detection, the results are transferred back to the host memory. The amount of the output is usually smaller than that of punctured faces. The intersected space-time edge detection takes about 20% of the simulation time, which is less than for the punctured face detection.

Overall, the primitive detection greatly reduces the amount of data for further analysis. Typically, the amount of output is only $10^{-1}$ to $10^{-2}$ of the order parameter field, depending on the data complexity. In our benchmark, the GPU-accelerated detection is more than 200 times faster than a CPU-based implementation. We further avoid unnecessary frames by adaptive strategies as described in the in following sections.

4.3 Simulation coupling and adaptive detection

The pseudocode of the simulation coupling is in Algorithm 1. We tightly couple the GPU-accelerated primitive detection with the simulation; thus there is no data movement for the punctured face detection. The intersected edge detection needs to copy the phase field in the previous timestep, but the cost to copy data on the same GPU is minor. The time of copying results back to the main memory is also negligible compared with the detection.

We further reduce the overall online processing time by adapting the primitive detection. If we detect the primitives in every iteration, the primitive detection takes about 50% of the computation time in the simulation. We instead skip a frame if there are no intersected space-time edges. The rationale of the adaptive detection is based on the vortex tracking algorithm. If there are no intersected space-time edges for timesteps $\mathit{it}$ and $\mathit{i}$ ($|E'_{it}| = 0$), the IDs of punctured faces remain the same, but the puncture locations may change within the bounds of those faces. We regard the movement of vortices as negligible in this case. Depending on the stability of simulations, we can skip up to 95% of the timesteps in our experiments.

Algorithm 1 Main loop of the tightly-coupled vortex primitive detection. TDGL is the simulation; comm is the communicator, and isend() is the non-blocking send. $\psi$ and $\psi'$ are the order parameter fields of the previous and the current time step, respectively; $\mathit{it}$ and $\mathit{i}$ are the time step of the previous and the current time step, respectively. PFs and IEs are abbreviations for punctured faces and intersected space-time edges, respectively.

$$i ← 0$$

while !TDGL.\text{finished}() do
  $\psi ← $TDGL.\text{iterate}() ❌ Get $\psi$ from the simulation
  if $i = 0$ then
    $\mathcal{F}_i ← $detect_PFs_GPU($\psi$) ❌ Detect PFs
    comm.isend(tag='F', $\{i, \mathcal{F}_i\}$) ❌ Send PFs
  else
    $E'_{it} ← $detect_IEs_GPU($\psi', \psi$) ❌ Detect IEs
    if $|E'_{it}| = 0$ then
      continue ❌ If vortices do not move, skip the current time step
    else
      comm.isend(tag='E', $\{\mathit{it}, E'_{it}\}$) ❌ Send IEs
      $\mathcal{F}_i ← $detect_PFs_GPU($\psi$)
      comm.isend(tag='F', $\{i, \mathcal{F}_i\}$)
      $\{\mathit{it}, \psi'\} ← \{i, \psi\}$ ❌ Copy the current time step
  end if
  $i ← i + 1$
end while

5 LOOSELY-COUPLED AND TASK-PARALLEL VORTEX ANALYSIS

The vortex analyzer extracts, tracks, analyzes, and reduces vortices in the in situ workflow. The main challenge in the vortex analyzer is the unbalanced producer (simulation/primitive detector) and consumer (vortex analyzer) rate. Hence, we design and implement a task-parallel graph analysis scheme. Because the producer can generate vortex primitives much faster than the consumer can, the consumer could overflow memory. We schedule the graph analysis tasks on different threads in parallel, in order to improve the consuming rate.

5.1 Parallel task execution

As shown in Figure 4, our task-parallel model is an acyclic directed graph, where the nodes are data and tasks and the links define their dependencies. There are two types of data nodes: punctured faces $\mathcal{F}_i$ and intersected space-time edges $E'_{it}$, where $\mathit{it}$ and $\mathit{i}$ are the in-

Figure 3: Mesh subdivision for ambiguity-free vortex extraction and tracking: (a) the ambiguity case that a hexahedron intersected by two vortices; (b) the unambiguous results that vortex A punctures cells DEHG and DEFG while vortex B punctures ABCF; (c) and (d) show how a hexahedron is subdivided into six tetrahedra.
We compute the minimum distance between vortices at every time step. The output of the distances is stored in a matrix. Scientists are interested in how distance between vortices changes over time, especially for vortices that recombine. In the following sections we also use the matrix for distance projection in order to provide an occlusion-free 2D visualization.

**Moving speed.** The moving speed $V$ of a vortex, which is directly related to the energy dissipation of superconducting material, is defined as the area swept by the vortex in unit time. The swept area of an interval is computed by first tessellating the vortices in the two time steps into triangles and then summing up the areas of the triangles. The results are stored with the output vortices.

The data reduction of vortices is based on the work of Phillips et al. [22], which reduces the data size by geometry simplification and parameterization. First, the vortices are simplified with the Ramer-Douglas-Peucker [8] algorithm, which approximates an input curve by a series of points. Second, we use Bezier curves to dynamically fit the simplified curve [25]. Both data reduction algorithms are lossy but error bounded. Notice that errors in the former algorithm are measured by the maximum distance between the simplified curve and the input curve, and the errors in the latter algorithm are measured by the squared error between the input and fitted curve. Given a reasonable error bound (0.1 and 0.01 in our experiments), the output vortices can reduce up to 90% in storage size.

## 6 Post hoc visualization

We provide post hoc visualization for users to explore the online analysis results. The visualization involves four linked views: spatial, timeline, distance projection, and distance. The design goals are to help scientists explore spatial distributions of vortices and analyze the changes of vortices over time.

The spatial view visualizes the 3D geometries of vortices and material inclusions in the simulation. Users can identify and explore important spatial structures of vortices. In this view, the vortex curves are extruded into tubes for better shading and perception than possible with plain line geometries. The color indicates the ID of vortices. The color of a vortex remains the same over time unless the vortex has a topological change.

The timeline view visualizes the events over time with glyphs. Users can navigate the time-varying vortices by moving the cursor on the timeline view.

The distance view (Figure 5) visualizes how distances between vortices change before and after recombination in the *Crossing* simulation. The legends show the timestep of recombination and the involved vortices of the event. The dark blue line corresponds to the event in Figure 10.
at time 0. We design this view because scientists would like to know how the minimum distance changes before and after a recombination. Users can also select vortices to see how they morph over time, which is the key in the study of vortex dynamics.

The distance projection view (Figure 6) maps the minimum distances between vortex lines into 2D space for individual time steps. We use the multidimensional scaling algorithm to transform the input distance matrix into a set of points in 2D space. The distance between points in the 2D space is preserved as much as possible. The rationales for the projection view is twofold. First, the projection view provides a 2D visualization of vortices without any occlusions. Although the geometric information in the projection is lost, the distance information is still kept in the 2D projections. Second, the projection view can visualize the change of distances between time steps by animation. The projection provides a visual clue about the distance changes and helps users select vortices of interest.

7 IMPLEMENTATION

We use various tools and APIs to build our in situ visualization framework. The software stack of our implementation is illustrated in Figure 7.

The tightly-coupled primitive detector is implemented with CUDA and C, without any third-party libraries. The output primitives are sent to the loosely-coupled vortex analyzer asynchronously via UNIX pipes. The primitives can also be transferred to the vortex analyzer running on a remote machine by redirecting the pipe with the netcat command in UNIX.

The loosely-coupled vortex analyzer is written in C++11. The Intel Threading Building Blocks (TBB)¹ library provides APIs to schedule the task-parallel execution and lock-free data structures to achieve high concurrency. The output data, including vortex lines, association matrices, events, and all other analysis results, are serialized with Google protocol buffers² and then written to storage by Facebook RocksDB.³ We store data in a database instead of files because RocksDB provides a high performance persistent key-value store optimized for multithreaded writes. We use bzip2 to further compress the output in the database. Notice that the vortex analyzer can run on either the same node as the simulation or a standalone node with network connections. We usually run both the primitive detector and the vortex analyzer on the same node to fully utilize the hardware resources of the GPU and all CPU cores.

The web server is hosted with Node.js,⁴ which is a server-side JavaScript runtime. We build bindings of our C++11 code for Node.js, in order to enable access to the output data in the database. The web server and clients exchange queries and results via WebSockets, which provides full-duplex communication.

The web client is implemented with WebGL and D3.js [4]. The 3D interactive visualization is based on Three.js,⁵ which manages the scene graph and wraps WebGL. The 2D visualizations are rendered in SVG and overlaid on top of the 3D view.

8 PERFORMANCE BENCHMARK

The performance benchmarks are shown in Table 1, Figure 8, and Figure 9. The benchmark was conducted on a workstation equipped with two quad-core Intel Xeon E5620 CPUs, 12 GB main memory, and an NVidia K40c GPU. Each CPU supports up to 8 hardware threads, and the GPU has 2,880 stream cores and 12 GB memory. The simulation process that manages the GPU simulation and primitive detection is single threaded, and the vortex tracker can use up to all eight CPU cores on the machine. The output data are written to the database on a local hard disk drive. We analyze the performance of our in situ workflow in four aspects: speedup of the GPU-accelerated primitive detection, performance of task-parallel vortex analyzer, and I/O cost.

First, we measure the speedup of the GPU-accelerated primitive detection algorithms in Section 4. The timings with respect to the different numbers of primitives are shown in Figure 8. The baseline

²http://threejs.org/
³http://nodejs.org/
⁴https://developers.google.com/protocol-buffers/
⁵http://rocksdb.org/
approach is a software implementation that can run on either single or multiple (eight in the experiments) CPU threads. The timings of GPU algorithms include the data transfer, that is, the copy back of primitives from the GPU memory. The performance of the primitive detection does not change significantly with the number of primitives. On average, the punctured face detection is slower than the single-threaded and multithreaded CPU implementation, respectively. Likewise, the intersected space-time edge detection on the GPU is faster than the single-threaded and multithreaded CPU code, respectively.

Second, we evaluate the efficiency of the adaptive primitive detection strategy in Section 4.3. The criterion to skip a time step, namely, the number of intersected space-time edges, is zero, because they imply the movement of vortices and we ignore a time step that has minimal change to the previous one. As shown in Table 1, the percentage of skipped time steps ranges from 1% to 95%; thus the efficiency highly depends on the simulation. In general, if the simulation is more stable, the adaptive strategy can save more time in the in situ workflows.

Third, we benchmark the loosely coupled, asynchronous, and task-parallel vortex tracker in Section 5. Figure 8 shows the timings of extraction tasks and tracking tasks with respect to different numbers of primitives. We can see that the execution time of an individual task grows linearly with the number of primitives. In most cases, the execution time is much longer than that of the producer (simulation and primitive detection), which confirms that we must use parallelism to boost the performance of the consumer (vortex detector). Figure 9 visualizes the CPU utilization of the in situ workflow as the simulation time elapses. The CPU utilization reflects how many CPU cores are used. Notice that the online data analysis and I/O also run in the experiment. The CPU utilization of the producer remains almost constant (100%), because the GPU is managed by a single CPU thread. The CPU utilization of the consumer, which is proportional to the number of primitives, varies from 10% to 1500%. At the early stage of the simulation, the system is highly unstable; thus there are usually more vortices, and the workload of the vortex tracker is higher. After the simulation is stabilized, the workload is also stabilized—the workload of the consumer is about three times higher than that of the producer.

Fourth, we measure the I/O cost of the in situ workflow. The output data written in the database is only about $10^{-5}$ to $10^{-2}$ of the order parameter field data in the simulation. As recorded by the database log, the peak output I/O bandwidth is about 30 MB/s, which is much more affordable than storing the raw simulation outputs.

Overall, the cost of the whole in situ workflow (($T_f + T_e$)/$T_{total}$) is 75%--88% of the simulation, depending on the data complexity. Notice that $T_f$ and $T_e$ are already much less than they used to be with CPUs, and we also hide $T_{analysis}$ in $T_{total}$ by the asynchronous task-parallel execution. The in situ workflow also greatly reduces the amount of output while obtaining the most precise analysis results.

9 Application Results

We demonstrate application cases of in situ vortex visualization. The specifications of the simulations are in Table 1.

9.1 Unstable BX Simulation

The Unstable BX simulation is conducted in order to understand the periodic energy dissipative states in a superconductor with several inclusions. The magnetic field is aligned with the external current, which would result in no vortex dynamics (due to the absence of Lorentz forces). However, thermal fluctuations and material defects, modeled as spherical inclusions, lead to interesting behavior. Scientists would like to study when, and how vortices recombine with each other while performing periodic motion.

Our in situ workflow enables the detailed visualization and analysis of vortex dynamics. Figure 1 compares the visualization results from both traditional post hoc methods and in situ methods. Previously, scientists saved the order parameter field for every 100 timesteps and then analyzed the data in low temporal resolution after the simulation. As shown in Figures 1(c) and (d), a “compound” event happens, which involves vortices #588, #489, and #485 at timestep 6990, two vortices #588 and #489 cross each other and swap parts. Two new vortices #590 and #591 are generated after the recombination at timestep 6991. Right after this event, #590 and #485 recombine into #592 and #593. Another example is shown in Figure 6, which visualizes a recombination event with distance projection. We can observe how vortices deform, bend, attract each other, and then repel each other after the events.

9.2 Crossing Simulation

Scientists study vortex crossings in superconducting slabs that are induced by tilting the applied magnetic field [29]. The way vortices in superconductors cross each other has been studied for many years, but most studies were performed either under artificial conditions or on very small lengths scales. The GLGPU simulation makes it possible to reveal realistic situations. The in situ visualization of vortices and their distances gives a good impression of the
physics of these crossing events, which typically consist of merge, split, and recombination.

As shown in Figure 10, the early stage of the simulation is unstable and chaotic. Vortices are mostly stable in the later stage, yet a few recombinations happen. A traditional post hoc workflow is difficult to capture the rapidly changing vortices near recombinations. Now we can find the exact time of the events and observe how vortices morph before and after the events.

We generate the distance plot (Figure 5) in situ, recently was used to study various vortex recombinations scenarios [12]. Previously in the post hoc workflow, scientists had to run the simulation twice to generate the distance plot. In the first “full” run, the order parameter field is stored every 250 iterations. After the first run, scientists choose an interval of interest that involves a recombination event. In the second “detail” run, every iteration in this interval (2,000 timesteps) is stored for post hoc analysis. Such a workflow requires a manual process and computation resources, and it may miss important features in the simulation.

With the in situ workflow, the distance plot can be automatically generated with no manual operations or additional simulation runs. The recombination events and vortex curves are extracted and stored based on every single iteration in situ. Compared with the traditional post hoc workflow, the in situ workflow enables scientists to visualize and analyze the vortex dynamics in greater detail and with less work than previously possible.

10 CONCLUSIONS AND FUTURE WORK

In this paper, we present an in situ vortex visualization framework for GPU-based TDGL simulations. The system consists of tightly-coupled GPU-accelerated primitive detection for ambiguity-free vortex analysis, loosely-coupled task-parallel feature tracking, and web-based visualizations. Applications show that the only solution to our driving scientific problem—finding vortex recombinations in the rapidly changing simulations—is the in situ workflow.

We would like to extend our study in four directions. First, we are going to design the in situ workflow for the next-generation TDGL simulation, which is based on an unstructured mesh in distributed memory. Simulation steering is also going to be a necessary component in the future system. Second, we would like to apply our task-parallel design to other feature tracking problems. Third, we are going to develop a scalable solution to visualize events in long time sequences for in situ visualization. Fourth, we plan to use our in situ workflow in other complex-valued Ginzburg-Landau simulations such as Bose-Einstein condensation and superfluidity.

APPENDIX

In this appendix, we prove that a tetrahedron in our mesh subdivision cannot be punctured by two vortices and therefore ambiguity-free vortex extraction is guaranteed. In the discrete \( \psi \) field, we calculate the phase jump over a mesh edge \( AB \) by

\[
\Delta_{AB} = \int_{AB} \nabla \theta \cdot d\mathbf{l} = \mod (\theta_B - \theta_A + \pi, 2\pi) - \pi, \tag{2}
\]

where \( \theta_A \) and \( \theta_B \) are the phase angle on nodes \( A \) and \( B \), respectively. Notice that the modulo maps \( \Delta_{AB} \) into the range of \([ -\pi, \pi )\), so we cannot measure any larger phase jumps over mesh edges than this range.

Lemma 1. If \( x + y + z = 2\pi \), \( x, y, z \in [ -\pi, \pi ] \) then \( x, y, z > 0 \).

Proof. Assume \( z \leq 0 \). Then \( x + y \geq 2\pi \). A contradiction occurs because \( x < \pi \) and \( y < \pi \).

Lemma 2. There is at most one pair of punctured points on a tetrahedron, if the phase jump over every edge of this tetrahedron is in the range of \([ -\pi, \pi )\).
As shown in the right figure, in tetrahedron $ABCD$, the phase jumps over every edge $\Delta AB, \Delta BC, \Delta CA, \Delta CD, \Delta DA$, and $\Delta BD$ are in the range of $[-\pi, \pi]$. We prove this lemma by contradiction.

Proof. Without loss of generality, we assume there are two pairs of punctured faces, so that $\Delta AB + \Delta BC + \Delta CA = 2\pi$, $\Delta AC + \Delta CD + \Delta DA = 2\pi$, $\Delta AD + \Delta DB + \Delta DA = -2\pi$, and $\Delta BD + \Delta DC + \Delta CB = -2\pi$. Based on Lemma 1, we have $\Delta CA \neq 0$ and $\Delta AC > 0$. A contradiction occurs because $\Delta CA = -\Delta AC$. The lemma is true because relabeling the tetrahedron can cover all cases (any pair of faces have common edge, which is $AC$ in our case).

ACNOWLEDGMENTS

This material is based upon work supported by the U.S. Department of Energy, Office of Science, under contract number DE-AC02-06CH11357. This work was supported by the Scientific Discovery through Advanced Computing (SciDAC) program funded by the U.S. Department of Energy, Office of Science, under contract number DE-AC02-76CH03000. This work was supported by the Scientific Discovery through Advanced Computing (SciDAC) program funded by the Office of Science, under contract number DE-AC02-76CH03000.

REFERENCES


The submitted manuscript has been created by UChicago Argonne, LLC, Operator of Argonne National Laboratory (“Argonne”). Argonne, a U.S. Department of Energy Office of Science laboratory, is operated under Contract No. DE-AC02-06CH11357. The U.S. Government retains for itself, and others acting on its behalf, a paid-up nonexclusive, irrevocable worldwide license in said article to reproduce, prepare derivative works, distribute copies to the public, and perform publicly and display publicly, by or on behalf of the Government.