PSCC: Parallel Self-Collision Culling with Spatial Hashing on GPUs

Min Tang
Zhejiang University
Alibaba-Zhejiang University Joint Institute of Frontier Technologies
tang_m@zju.edu.cn

Zhongyuan Liu
Zhejiang University
lzy_work@foxmail.com

Ruofeng Tong
Zhejiang University
trf@zju.edu.cn

Dinesh Manocha
University of North Carolina at Chapel Hill
dm@cs.unc.edu

Figure 1: Our novel GPU-based collision detection algorithm is used to simulate clothes with regular (a-c) or irregular (d-f) shapes, which may correspond to a single layer (b, e) or multiple layers (a, c, d, f). We use a parallel spatial hashing algorithm to perform inter-object and intra-object collision checking on a GPU along with normal cone culling. The yellow line segments highlight all the areas that have no self-collisions and can be culled away. As compared to prior GPU-based collision detection algorithms, we observe $6 - 8\times$ speedup.

ABSTRACT

We present a GPU-based self-collision culling method (PSCC) based on a combination of normal cone culling and spatial hashing techniques. We first describe a normal cone test front (NCTF) based parallel algorithm that maps well to GPU architectures. We use sprouting and shrinking operators to maintain compact NCTFs. Moreover, we use the NCTF nodes to efficiently build an enhanced spatial hashing for triangle meshes and use that for inter-object and intra-object collisions. Compared with conventional spatial hashing, our approach provides higher culling efficiency and reduces the cost of narrow phrase culling. As compared to prior GPU-based parallel collision detection algorithms, our approach demonstrates $6 - 8\times$ speedup. We also present an efficient approach for GPU-based cloth simulation based on PSCC.

KEYWORDS

self-collision culling, normal cone, spatial hashing, GPU, cloth simulation

ACM Reference Format:


1 INTRODUCTION

Fast collision detection between deformable models is a key problem in physically-based simulation. In particular, reliable collision checking is important for many applications including cloth and surgical simulation. In many cases, collision detection is one of the major bottlenecks [Jiang et al. 2017; Selle et al. 2009; Sutherland et al. 2006; Tang et al. 2016].

Most algorithms for collision detection between deformable models are based on bounding volume hierarchies (BVHs) [Heo et al. 2010; Kim et al. 2009; Zhang and Kim 2014] or spatial hashing [Fan et al. 2011; Pabst et al. 2010; Weller et al. 2017]. As the models undergo deformation, these hierarchies/hashing tables are updated or reconstructed and used to cull away non-overlapping pairs. In order to accelerate the computations, many parallel algorithms that utilize multiple cores on a CPU or a GPU have been proposed [Fan et al. 2011; Heo et al. 2010; Kim et al. 2009; Pabst et al. 2010; Tang et al. 2010a; Weller et al. 2017; Zhang and Kim 2014]. However, there are two main challenges in terms of
using these algorithms on complex models related to memory overhead and use of high-level culling methods.

Many BVH-based parallel algorithms maintain or update a front of the underlying BVHs to accelerate the computation—Klosowski et al. [1998] Li and Chen [1998] Tang et al. [2010b] Zhang and Kim [2014]. However, the storage overhead of these fronts can be high and can require several GBs of memory for complex models composed of tens or hundreds of triangles. Moreover, this storage overhead affects the performance on commodity or mobile GPUs, which typically have a few GBs of memory.

One of the key issues in terms is self-collision culling. Many high-level culling methods based on normal cones have been proposed to perform such culling—Provot [1997] Schwartzman et al. [2010] Tang et al. [2009] Wang et al. [2017]. However, they have been limited to sequential collision-detection methods. In fact, most parallel GPU-based collision detection and cloth simulation systems do not use self-collision culling based on normal cones—Tang et al. [2011] [2016] Weller et al. [2017]. As a result, the culling efficiency may not be high and results in a large number false-positive elementary tests.

Main Results: In this paper, we present a novel GPU-based collision detection algorithm (PSCC) to address these issues. This includes a parallel normal culling algorithm that uses a combination of shrinking and sprouting operators and has lower memory overhead (Section 3). The resulting hierarchy traversal performs more effective culling with lower runtime overhead. Moreover, we present an enhanced spatial-hashing method that is used for inter-object and intra-object collisions (Section 4). A key aspect of our approach is a novel workload scheme to distribute the computations evenly among different GPU threads. Based on our collision detection algorithm, we present an improved collision-handling scheme for cloth simulation that can reuse broad-phase test results (Section 5).

We have implemented these algorithms on different GPUs and evaluated their performance on many complex benchmarks. Our novel combination of parallel normal cone culling with spatial hashing results in the following benefits:

- **Lower memory overhead:** For scenes with hundreds of thousands of triangles, the storage overhead reduces from several gigabytes—Tang et al. [2011] to about 500M bytes.
- **Faster collision detection:** As compared to prior GPU-based parallel collision detection algorithm—Tang et al. [2014] Weller et al. [2017], our approach demonstrates 6 – 8X speedup.
- **Parallel cloth simulation:** In practice, our GPU-based cloth simulation takes about one second per frame on complex benchmarks with tens or hundreds of thousands of triangles, and is about 4 – 6X faster than prior GPU-based algorithms—Tang et al. [2016].

2 RELATED WORK

There is considerable work on collision detection, cloth simulation, and parallel GPU-based algorithms to accelerate collision detection. In this section, we give a brief overview of prior work on collision culling and parallel algorithms.

**Self-collision culling:** Provot [1997] presented an efficient method using normal cones for self-collision culling. It is used to remove a large number of redundant tests at relatively “flat” areas of the deformable models. Conventionally, the normal cones associated with the internal nodes of a BVH are computed in a bottom-up manner. Self-collision detection can then be performed in a top-down manner. Schwartzman et al. [2010] combined this method with self-collision test trees (SCTT) to accelerate discrete self-collision queries for general deformable models with $O(n)$ complexity. Tang et al. [2009] extended normal cone culling to continuous collision detection. However, its complexity is $O(n^2)$, where $n$ is the number of boundary edges. Wang et al. [2017] presented a conservative normal cone-based culling algorithm with $O(n)$ complexity. Other techniques include energy-based methods—Barbić and James [2010] Zheng and James [2012] and radial-based culling methods—Wong and Cheng [2014] Wong et al. [2013].

**Spatial hashing on GPUs:** Spatial hashing is a classic algorithm for collision detection, and can be easily parallelized on GPUs—Lefebvre and Hoppe [2006]. A spatial hashing algorithm has a constant time complexity on querying all the triangles in proximity. Palst et al. [2010] used a uniform grid to perform the broad phase culling on GPUs for triangle meshes. In order to handle the non-uniform distribution of scene geometry primitives, Faure et al. [2012] extended uniform grids to two-layer grids. Recently, hierarchical grids have been used to further improve the efficiency of spatial subdivision-based CD algorithms—Faure et al. [2012] Weller et al. [2017] Wong et al. [2014]. Although hierarchical grids can address the non-uniform distribution of scene primitives, they cannot cull the triangles that belong to “relatively flat” areas of cloth or deformable models. The previous spatial hashing algorithms—Faure et al. [2012] Weller et al. [2017] Wong et al. [2014] tend to treat soft bodies and rigid bodies uniformly. This can result in a high number of redundant self-collision tests, i.e., false positives.

**Parallel cloth simulation on multi-core/many-core processors:** Selle et al. [2009] designed a parallel cloth simulation algorithm for multi-core platforms. That algorithm can take up to 30 minutes per frame for a cloth mesh with 500K triangles on a 16-core workstation. Tang et al. [2013] proposed a streaming algorithm for regular-shaped high-resolution cloth simulation on GPUs. The algorithm has been extended for cloth simulation with arbitrary topology structures—Tang et al. [2016], and can take up to 35 seconds per frame for cloths with 1M triangles on a commodity GPU. These parallel algorithms maintain a bounding volume traversal tree (BVTT) front for parallel collision checking, which can have a large memory overhead. Recently, Jiang et al. [2017] used a Lagrangian approach to simulate the anisotropic elastoplasticity of cloth, and their implementation takes about 2 minutes per frame for a cloth mesh with 1.8M triangles.
3 PARALLEL NORMAL CONE CULLING

In this section, we present our parallel normal cone culling algorithm. We first introduce the notation used in the rest of the paper.

3.1 Notation

We mainly focus on accurate collision detection for general deformable objects, including self-collisions. We mainly focus on continuous collision detection (CCD), while parallel scheme can also be used for discrete collision detection (DCD). Our CCD algorithm assumes linearly interpolating trajectories [Bridson et al. 2002; Provot 1997; Tang et al. 2011] between the mesh vertices corresponding to two successive simulation instances. We use the following acronyms in the rest of the paper: BV (bounding volume), BVH (bounding volume hierarchy), NCTT (normal cone test front), and NCTF (normal cone test front).

3.2 Parallel Culling

Our parallel normal culling algorithm is based on BVH-based culling with BYTT (Bounding volume testing tree) fronts [Tang et al. 2011]. The BVH quality is critical for the culling efficiency of normal cone tests. For cloth simulation, we construct BVH in its material space [Wang et al. 2011] (Figure 2(a)) instead of the geometry space (Figure 2(b)). As a result, we observe much higher culling efficiency for normal cone tests, especially for areas with multiple layers of cloth (see the upper part of the body in Figure 3). Conceptually, any normal cone testing algorithm [Provot 1995; Schwartzman et al. 2010; Tang et al. 2009; Volino and Thalmann 1994] can be used in Algorithm 1. During the construction step, we also compute a normal cone and gather its contour edges for each BVH node. This is useful for the cases where all the underlying triangles are connected with each other. The normal cone is computed by merging all the normal vectors of these triangles.

In this work, we use a recent algorithm by Wang et al. [Wang et al. 2017] based on unprojected contour test with O(n) complexity, due to its support for both CCD and DCD, and its simplicity for implementation.

We use a NCTF to record all the BVH nodes where normal cone tests are terminated during the tree traversal at the last time step (as shown in Figure 2). By utilizing the spatial-temporal coherence between simulation time steps, we perform various normal cone tests in parallel starting from all the nodes recorded in the front. As shown in Algorithm 1, all the nodes of the NCTF can be tested and updated independently. For those nodes that do not pass the normal cone tests, sprouting operators are used to traverse down and add their descendents into the NCTF. In this manner, high-level self-collision culling is performed in parallel on a GPU.

Sprouting operator: In Algorithm 1, we first perform the normal cone test on each node in the NCTF (lines 2-3). For those nodes that fail the test, we use the sprouting operator (Algorithm 2) to traverse to their descendents (line 4-5).

Algorithm 1 SelfCollideParallel(F): Perform self-collision in parallel with a NCTF F.

1: for all \( n^i \in F \) do
2: if IsLeaf\((n^i)\) or NormalConeTest\((n^i)\) == true then
4: end if
5: Delete\((F, n^i)\)
6: Sprouting\((F, n^i \rightarrow \text{left})\)
7: Sprouting\((F, n^i \rightarrow \text{right})\)
8: end for

Algorithm 2 Sprouting\((F, n^i)\): Update the normal cone test front \( F \) by sprouting from a BVH node \( n^i \).

1: if IsLeaf\((n^i)\) or NormalConeTest\((n^i)\) == true then
2: Insert\((F, n^i)\)
3: else
4: Sprouting\((F, f^i \rightarrow \text{left})\)
5: Sprouting\((F, f^i \rightarrow \text{right})\)
6: end if

Ultimately the nodes that pass the normal cone tests are added to the NCTF (line 2).

Shrinking operator: In Algorithm 1, we only use the sprouting operator to update the front. In practice, the underlying deformable mesh or the cloth may change from a wrinkling status to a flat status. In order maintain a compact NCTF, we use a shrinking operator to re-scan the front nodes after several time steps. During the simulation, some tangled areas may become flat again. So the front needs to shrink upward. If two sibling front nodes are in the front, we remove
Front B
Sprout
Shrink
Front A

Figure 3: Front Updating: The NCTF remains compact by updating with sprouting and shrinking operators. During different time steps, the regions that can be culled with normal cone tests change. Therefore, the front needs to be updated from Front A to Front B. Both A and B are compact. Sprouting and shrinking operators are used to maintain the compact fronts (i.e. A, B).

Figure 4: The tightness of the BVH is important for NC culling. Here we highlight the culling efficiency with different BVH construction algorithms. Compared with the BVH constructed in geometry space (b), the BVH constructed in material space (a) results in better normal cone culling efficiency, especially for multi-layer clothes (see the upper part of the body).

all these sibling nodes and insert their parent into the front. In this way, the front will shrink upward in various areas which become flat again. As shown in Figure 3, the front can be fully updated in parallel with sprouting (Algorithm 2) and shrinking operators.

Memory overhead: The number of nodes in NCTF is bounded by $2 \times N + 1$, where $N$ is the number of triangles. This $O(N)$ memory requirement is much less than the memory overhead of the BVTT front [Tang et al. 2011], which can be $O(N^2)$. As a result, we observe reduced memory usage and better caching behavior on complex benchmarks.

Self-collision culling efficiency: NCTF is designed to perform parallel normal cone culling, while BVTT front is used for parallel collision detection between BVH nodes. For self-collision detection, NCTF is more efficient than BVTT due to its ability to remove large ”flat” areas. On the other hand, the BVTT needs to be traversed to leaf level, and the resulting algorithm needs to check all the triangle pairs with overlapping bounding volumes for exact collisions. Most of these overlap tests tend to be false positives, as the adjacent triangle pairs cannot be cut out with bounding volume tests. As highlighted in Figure 1, all these areas bounded by the yellow line segments are skipped for self-collision checking with NCTF. If we use BVTT, all these triangle pairs have to be checked. Therefore, NCTF tends to be more efficient than BVTT for self-collision detection.

4 SPATIAL HASHING WITH NORMAL CONE FRONT

The high-level culling algorithm for self-collisions needs to be combined with low-level collision culling methods to reduce the number of false positive elementary tests. We combine parallel normal culling with extended spatial hashing for improved performance. A key aspect of this integrated scheme is the use of paired hashing key. We split the bounding volume of the entire simulation scene into cells and assign each cell a unique Cell ID. For each triangle, we keep track of the corresponding Cell IDs for those cells that overlap with its BV. Each triangle also has a unique NCTF node ID, which is computed from the parallel normal cone tests (e.g., a, b, c, ..., n in the Fig. 2). We use both, spatial information (Cell IDs) and NCTF node IDs, as the hashing keys. Logically, the triangles of a mesh can be organized into a workload distribute table, as shown in Figure 7. All the triangles with the same $\{\text{CellID, NodeID}\}$ are skipped in terms of performing
intersecting tests, because they are in same flat region and do not need to be tested for intra-object collisions.

With conventional spatial hashing, only \( \{ \text{CellID} \} \) are used as the hashing keys. The intersecting number in one spatial cell, \( N_{\text{inter}} \), is:

\[
N_{\text{inter}} = n_t \cdot (n_t - 1)/2,
\]

where \( n_t \) is the number of triangles in that cell. With the new hashing key \( \{ \text{CellID}, \text{NodeID} \} \), the new number, \( N'_{\text{inter}} \), becomes:

\[
N'_{\text{inter}} = \sum_{i \neq j} n_i \cdot n_j,
\]

where \( n_i \) and \( n_j \) are the numbers of triangles with the same Node IDs in the cell, and \( \sum n_i = n_t \). As long as \( n_i \geq 1 \), we have:

\[
N_{\text{inter}} - N'_{\text{inter}} = n_t \cdot (n_t - 1)/2 - \sum_{i \neq j} n_i \cdot n_j = [\sum n_i]^2 - \sum n_i - 2 \cdot \sum_{i \neq j} n_i \cdot n_j]/2 = (\sum n_i^2 - \sum n_i)/2 \geq 0
\]

Based on the paired hashing key, we do not need to perform self-collision detection between the small group of triangles that have the same NodeID. Due to our formulation, the number of candidate triangle pairs is reduced.

Self-culling with spatial hashing decreases the number of threads, as each test is performed by a separate thread and PSCC results in fewer triangle-triangle tests (see Figure 6). This is useful in the narrow phase testing of the collision detection pipeline. We use two hash tables to build our workload-distribute table, and modify the workload distribute algorithm [Fan et al. 2011] for our approach to balance the computational load of GPU threads. Finally, a token position technique [Fan et al. 2011] is used to avoid an output of duplicate triangle pairs after the broad phase.

### 4.1 Workload Distribution

Corresponding to the paired hash keys, we use a novel workload distribution to distribute the computations evenly among GPU threads. Before workload distribution, we need two pieces of information. For each spatial cell, we need to collect all the triangles that overlap with that cell. We separate all the triangles that lie into the same cell into different sets by NodeID. As shown in Algorithm 5, two hash tables are used to implement this approach. In our algorithm, for all triangles that fall into the same cell and have the same NodeID, no extra intersection testing is needed. Therefore, triangleHashTable is used to separate the triangles with different CellIDs and NodeIDs into many different groups. Next, workLoadHashTable collects all the groups indexed with the same CellID. Finally, through these two hash tables, we can compute all the triangle pairs that are needed to perform the BV overlap tests (Figure 7). In this way, we can perform workload distribution among all the GPU threads.

We represent our spatial-hashing table as a row-major sparse matrix. We can insert a \{key, value\} pair by hashing its key to find the corresponding row index, and put the value to a value array on this row. The hash table construction algorithm is a two stage process very similar to the sparse matrix assembly method used in [Tang et al. 2016].
We use NCTF node IDs and cell IDs to make a uuid as the key to insert triangle index into the triangleHashTable.

```plaintext
10: for all row\textsuperscript{i} ∈ triangleHashTable do
11: Insert(workLoadHashTable, cellId\textsuperscript{i}, i)
12: end for
13: // Count the total amount of BV tests.
14: CountTotalThreadNum(workLoadHashTable, triangleHashTable)
15: // For every BV test, a thread is used to balance the computation load.
16: for all threadIdx\textsuperscript{j} ∈ totalThread do
17: DecodePairsAndTest(threadIdx\textsuperscript{j}, workLoadHashTable, triangleHashTable, T)
18: end for
```

### 4.2 Hashing Function

We use NCTF node IDs and cell IDs to make a uuid as the hash key. The \( \text{Key} \) is calculated by:

\[
\text{Key} = \text{uuid} \mod \text{hashTableSize} = (\text{CellId} + \text{NodeId} \times \text{CellIdMax}) \mod \text{hashTableSize}
\]

and

\[
\text{CellId} = 1d_x \times N_x + 1d_y \times N_y + 1d_z, \quad \text{NodeId} = x, y, z, \quad N_x, N_y, N_z \text{ are the number of cells and the ID of a current cell in the x,y,z axis of the spatial grid, respectively.}
\]

Our spatial hashing algorithm can easily be extended for collision detection between rigid bodies by simply assigning a unique NodeID for each rigid body. In this manner, all self-collision tests among the triangles belonging to the same rigid body can be avoided.

### 4.3 Overall PSCC Algorithm

Our collision detection algorithm is a combination of Algorithm 1 (high level culling) and Algorithm 3 (low level culling). We first execute Algorithm 1 to compute the NCTF node IDs, then execute Algorithm 3 for inter-object and intra-object collision culling. The paired hash keys in Algorithm 3 are based on the NCTF node IDs computed by Algorithm 1.

### 5 COLLISION HANDLING FOR CLOTH SIMULATION

We use our novel parallel collision detection algorithm to improve the performance of GPU-based cloth simulation. The original pipeline for a fast GPU-based algorithms is shown in Figure 8(a), which first performs proximity checking with DCD, collects all the proximity constraints for implicit time integration, performs penetration handling with CCD to ensure there are no penetrations between triangle pairs. In [Tang et al. 2016], BVTT front tracking is used as broad phrase culling, and is performed twice for DCD and CCD.

We notice that the vertices used for CCD in the current time step are the same as those for DCD at the next time step. This is due to the fact that the penetration-free mesh output from last time step is directly used as the input mesh for the next simulation time step, i.e. they have the same vertices. Therefore, the results of broad-phrase culling for last time step can be reused for next time step (as shown in Fig. 8(b)). We therefore perform spatial hashing culling...
Figure 9: Benefits of the new collision handling pipeline: For this frame of Benchmark Sphere, with the new collision handling pipeline, we are able to reduce the number of BV tests and the running time of broad phrase culling to 51.1% and 53.3%, respectively.

<table>
<thead>
<tr>
<th>Resolution (triangles)</th>
<th>Benchmark</th>
<th>Time Steps(s)</th>
<th>CAMA (K40c)</th>
<th>Our (K40c)</th>
<th>Our (1080)</th>
<th>Our (1080 TI)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Sphere</td>
<td>1/200</td>
<td>2.84</td>
<td>1.62</td>
<td>0.94</td>
<td>0.82</td>
<td></td>
</tr>
<tr>
<td>200k</td>
<td>1/200</td>
<td>2.92</td>
<td>1.72</td>
<td>0.97</td>
<td>0.87</td>
<td></td>
</tr>
<tr>
<td>Flag</td>
<td>1/100</td>
<td>3.39</td>
<td>0.81</td>
<td>0.35</td>
<td>0.20</td>
<td></td>
</tr>
<tr>
<td>Falling</td>
<td>1/30</td>
<td>3.49</td>
<td>0.98</td>
<td>0.51</td>
<td>0.39</td>
<td></td>
</tr>
<tr>
<td>Bishop</td>
<td>1/30</td>
<td>3.19</td>
<td>1.82</td>
<td>0.94</td>
<td>0.82</td>
<td></td>
</tr>
</tbody>
</table>

Figure 10: Performance: This figure shows the average running time for a single frame of our algorithm on four different configurations (i.e., CAMA on Tesla K40C, our system on Tesla K40C, our system on GTX 1080, and our system on GTX 1080 Ti, respectively). We observe significant speedups (up to 16X) over CAMA.

during the broad phrase culling, and execute it only once per time step. The culling results are used for the narrow phrase (both for DCD and CCD). During the narrow phrase with CCD, we detect all the penetrations between Vertex-Face/Edge-Edge pairs. The CCD computation is performed until all penetrations are handled. After CCD computation, another pass of narrow phrase testing with DCD is performed to collect proximity constraints for the next time step. In this way, we save the running time in terms of performing another pass during broad phrase culling by reusing its output.

In the Benchmark Sphere (Figure 9), we are able to reduce the number of BV tests and the running time of broad phrase culling by 51.1% and 53.3%, respectively.

6 IMPLEMENTATION AND RESULTS

Implementation: We have implemented our PSCC algorithm on three commodity GPUs, an NVIDIA Tesla K40c (with 2880 cores at 745MHz and 12G memory), an NVIDIA GeForce GTX 1080 (with 2560 cores at 1.6 GHz and 8G memory), and an NVIDIA GeForce GTX 1080 Ti (with 3584 cores at 1.58 GHz and 11G memory), respectively. We use these GPUs with varying number of cores to test the parallel performance of our approach. Our implementation is based on CUDA toolkit 8.0 and Visual Studio 2013 as the underlying development environment. We used a standard PC (Windows 7 Ultimate 64 bits/Intel i7 CPU@3.5GHz 8G RAM) as the testing environment and perform single-precision floating-point arithmetic for all the computations performed on a GPU. Moreover, we used Thrust for the prefix-sum operator and cuBLAS/cuSPARSE for linear system solving.

Benchmarks: We used three different benchmarks for regular-shaped cloth simulation:

- **Twisting**: Three pieces of cloth with a total of 200K triangles twist severely as the underlying ball rotates (Fig. 1(a)).
- **Flag**: A waving flag with 80K triangles is blowing in the wind (Fig. 1(b)).
- **Sphere**: Three pieces of cloth with a total of 200K triangles are hit by a forward/backward moving sphere (Fig. 1(c)).

These benchmarks contain many inter- and intra-object collisions. We used three other benchmarks for garment simulation:

- **Andy**: A boy wearing three pieces of clothing (with 127K triangles) is practicing Kung-Fu (Fig. 1(d)).
- **Falling**: A man wearing a robe (with 172K triangles) falls down rapidly under strikes (Fig. 1(e)).
- **Bishop**: A swing dancer wears three pieces of clothing (with 124K triangles) (Fig. 1(f)).

These are complex benchmarks with multiple pieces, layers, and wrinkles, which result in a high number of collisions. Our algorithm can handle inter- and intra-object collisions reliably (see video).

Performance: Figure 10 shows the resolutions and time steps for different benchmarks, and highlights the performance of our algorithm on these benchmarks. This includes the average frame time of our GPU-based algorithm on three commodity GPUs with different numbers of cores. These results demonstrate that our cloth simulation algorithm works well on different GPU architectures and that the performance is proportional to the number of cores. Compared with the CAMA performance [Tang et al. 2016] on an NVIDIA Tesla K40C, we observe significant speedups (up to 16X) on an NVIDIA GeForce GTX 1080 Ti.

Running time ratios: Figure 11 shows the running time ratios of different computing stages: time integration, broad phrase testing, narrow phrase testing, and penetration handling. These data are collected by running our system for Benchmark Sphere on the NVIDIA GeForce GTX 1080. As shown in the figure, time integration takes almost constant running time for all the time steps. Collision detection (broad phrase and narrow phrase) and penetration handling appear to be the most computationally expensive parts, especially when the cloths are tangled.

Memory overhead: Figure 12 compares the memory overhead of our algorithm (spatial hashing-based) and CAMA (BVH-based). We list all the memory breakdown for the Benchmark Sphere running using two different algorithms.
Figure 11: Ratios of different stages of our system: This figure shows the running time ratios of different computing stages: time integration, broad phrase testing, narrow phrase testing, and penetration handling, respectively. These data are collected by running our system for the Benchmark Andy on the NVIDIA GeForce GTX 1080.

Compared with our approach, CAMA \cite{Tang16} uses much more memory as the storage for BVTT front and triangle pairs (passed broad phrase culling). Our approach requires much less memory for spatial hashing and normal cone front. Due to the culling effects of normal cones, our approach also needs much less memory for the triangles pairs. CAMA \cite{Tang11, Tang16} needs more than 2G GPU memory for this benchmark, while our system needs approximately 500M GPU memory. We also observe similar memory reduction in other benchmarks using PSCC.

7 COMPARISON AND ANALYSIS

In this section, we compare the features and performance of our approach with prior methods.

GPU-based collision detection algorithm and cloth simulation system: As compared with prior GPU-based cloth simulation systems \cite{Tang13, Tang16}, the main benefits of our algorithm include much less memory overhead (Figure 12) and improved runtime performance on the same GPU (Figure 10). With the enhanced spatial hashing data structure, our approach is capable of performing large-area self-collision culling. Figure 16 shows the running time ratios of different computing stages of the CAMA system \cite{Tang16}. This performance data is collected by running CAMA for Benchmark Sphere on the NVIDIA GeForce GTX 1080. As shown in the figure, CAMA needs much more running time than our system for collision detection and handling. All the speedup are due to the improvements in the collision detection algorithm (broad phrase testing, narrow phrase testing, and revised pipeline), reduced memory overhead and the improved cloth simulation pipeline. On the same GPU platform (NVIDIA GeForce GTX 1080), our collision detection algorithm provides 6−8X speedup over prior GPU-based algorithms \cite{Tang11}, and results in 4−6X speedup over prior cloth simulation algorithm \cite{Tang16}.

CPU-based collision culling algorithms: The recent work of Wang et al. \cite{Wang17} combines normal cone culling
with BVTT front tracking. However, it is a serial algorithm designed for single thread implementation, it is hard to parallelize and compare the performance with a GPU-based parallel approach. Another drawback of this algorithm is that it needs a lot of CPU memory to store the BVTT front. Spatial Hashing Only: As a comparison, we also implemented the hierarchical spatial hashing algorithm KDet [Weller et al. 2017]. Figure 15 shows the total amount of BV-tests and output triangle pairs of the broad phase. The BV-tests’ numbers indicate the spatial hashing method’s efficiency and the triangle pairs’ output indicates the workload of the narrow phase. Our approach achieves better performance at both the broad phase and narrow phase.

Performance on high-resolution cloth simulation: Figure 14 shows the average running time for a single frame of our algorithm on a benchmark with 1M triangles. We observe 5X speedups (on a GTX 1080) over CAMA (on Tesla K40c). The recent work of Jiang et al. [Jiang et al. 2017] takes about 2 minutes per frame for a similar benchmark with 1.8M triangles, while our system takes about 13s per frame on an NVIDIA GeForce GTX 1080.

8 CONCLUSION AND LIMITATIONS

We present a GPU-based self-collision culling method, PSCC, based on a combination of normal cone culling (Algorithm 1) and spatial hashing techniques (Algorithm 3). We use a NCTF-based parallel algorithm, along with sprouting and shrinking operators to maintain compact NCTFs on GPU. This algorithm perform high-level self-collision culling and compute the NCTF node IDs. We use the NCTF node IDs and spatial location information to build an enhanced spatial hashing for low-level culling between triangles. This enhanced spatial hashing technique can perform inter-object collisions and also used for self-collision checking based on normal cone tests. We also redesign the collision handling pipeline of a GPU-based cloth simulation system by reusing the computation of broad phrase culling. We have demonstrated its performance on many complexly layered cloth
benchmarks containing 80 – 300K triangles. We observe significant speedups (4 – 6X) and much less memory overhead, as compared to prior GPU-based systems.

Our approach has several limitations. For tangled, collision detection and penetration handling still remain a major bottleneck in terms of the overall running time (Figure 11). We need better techniques to exploit frame-to-frame coherence since only a few of the vertices involved in the penetrations tend to move between the frame. Also, our approach uses normal cone culling for self-collision culling. For meshes undergoing topological changes, the normal cones and their associated contour edges need to be updated on-the-fly.

There are many avenues for future research. In addition to overcoming the limitations, we feel that it is possible to further improve the performance by exploiting the memory hierarchy and cache of modern GPUs (e.g. combined with [Wang et al 2018]). Also, it will be interesting to explore the computation potential of multiple GPUs by performing data/task partitioning wisely and perform interactive cloth and deformable simulations.

ACKNOWLEDGEMENTS

This research is supported in part by the National Key R&D Program of China (2017YFB1002703), NSFC (61732015, 61572423, 61572424), the Science and Technology Project of Zhejiang Province (2018C01080), and Zhejiang Provincial NSFC (LZ16F020003). Dinesh Manocha is supported in part by the 1000 National Scholar Program of China and NSFC (61732015).

REFERENCES


Duksu Kim, Jae-Pil Heo, Jaehyuk Huh, John Kim, and Sung-eui Yoon. 2009. HPCCD: hybrid parallel continuous collision detection using CPUs and GPUs. 28 (10 2009), 1791–1800.


