Optimization Discussion

In this project, I accelerated a CPU-based Bloom filter by offloading both insertion and membership–query operations to the GPU using CUDA. The baseline implementation performed all hashing and bitwise operations on the host, invoking **SipHash-2-4** for each of the *k* hash functions, then setting or checking bits in a linear array - yielding **O(n·k)** work in a fully serial fashion. The goal was to exploit the massive parallelism and high memory throughput of modern GPUs to achieve orders-of-magnitude speedup, while preserving the same false-positive characteristics of the Bloom filter.

CPU vs. GPU Approaches

- **CPU Implementation**
  Processes each input string one at a time. For each string, the host code calls siphash_cpu() *k* times, computes bit indices, and updates a byte-array representation of the filter. Because updates are serial and memory accesses are not coalesced, performance is limited by both compute and DRAM latency.

- **GPU Implementation**
  Launches one CUDA thread per string, allowing millions of strings to be processed concurrently. Each thread:
  1. Computes *k* 64-bit SipHash values entirely in registers via unrolled SIPROUND macros.
  2. Maps each hash to a bit position, calculating a 32-bit word index and mask.
  3. Uses atomicOr() on that word to safely set bits without global locks.
  4. Reads bits non-atomically for membership checks, exiting early on the first zero bit.

  This design maximizes SM occupancy, hides both compute and global-memory latency through thousands of active warps, and benefits from coalesced memory access patterns.

Key Optimizations

1. **Register-Only Hashing**
   SipHash's round functions are implemented inline, so all state lives in registers. There are no spills to local memory, preserving high instruction throughput.

2. **One-Thread-Per-String**
   A balanced thread-to-data mapping minimizes control overhead and ensures uniform workload distribution across warps.
3. **Fine-Grained Atomics**
   Using 32-bit atomicOr() operations on individual filter words avoids global locks, confines contention to narrow regions, and spreads accesses across DRAM banks.
4. **Coalesced Memory Access**
   Both reads and writes to the filter array and string-position table are arranged so that adjacent threads access adjacent addresses, minimizing DRAM transactions.
5. **Early-Exit in Queries**
   Threads stop hashing as soon as they find a zero bit, reducing unnecessary work for non-member strings.
6. **Tunable Block Size**
   I expose the CUDA block size as a parameter, allowing me to trade off register pressure and shared-resource usage against occupancy.

CUDA Event Timing

To isolate and accurately measure the execution time of GPU kernels - excluding string generation, memory transfers, and CPU-side overhead - I instrumented the code with CUDA events:

1. **Event Creation**
   I create two events on the default stream (cudaEventCreate(&start) and cudaEventCreate(&stop)).
2. **Recording**
   - Record **start** immediately before launching the add_kernel and check_kernel calls.
   - Record **stop** immediately after both kernels have been enqueued.
3. **Synchronization & Measurement**
   After invoking cudaEventRecord(stop), I synchronize on the stop event (cudaEventSynchronize(stop)), then call cudaEventElapsedTime(&elapsed_ms, start, stop) to obtain the elapsed time in milliseconds with sub-millisecond precision.

This method ensures that the reported time reflects only the GPU's work (all SipHash rounds, atomic writes, and bit-tests) and is not skewed by host-device synchronization or driver overhead.

Experimental Setup & Results

To validate these optimizations, the following experiments were conducted:

/apps/GPU_course/runScript.sh proj3_akmal.cu 1000000 0.00001 64

```
[akmal@forest.usf.edu@gaivi-login1 proj3]$ /apps/GPU_course/runScript.sh proj3_akmal.cu 1000000 0.00001 64
----------------- nvcc Info: --------------------
nvcc: NVIDIA (R) Cuda compiler driver
Copyright (c) 2005-2021 NVIDIA Corporation
Built on Wed_Jun__2_19:15:15_PDT_2021
Cuda compilation tools, release 11.4, V11.4.48
Build cuda_11.4.r11.4/compiler.30033411_0
----------------- Compiling --------------------
ptxas info    : 0 bytes gmem
ptxas info    : Compiling entry function '_Z12check_kernelPKjPKcPKiPihym' for 'sm_52'
ptxas info    : Function properties for _Z12check_kernelPKjPKcPKiPihym
    0 bytes stack frame, 0 bytes spill stores, 0 bytes spill loads
ptxas info    : Used 44 registers, 376 bytes cmem[0], 8 bytes cmem[2]
ptxas info    : Compiling entry function '_Z10add_kernelPjPKcPKihym' for 'sm_52'
ptxas info    : Function properties for _Z10add_kernelPjPKcPKihym
    0 bytes stack frame, 0 bytes spill stores, 0 bytes spill loads
ptxas info    : Used 44 registers, 368 bytes cmem[0], 8 bytes cmem[2]
----------------- Executing --------------------
srun: defined options
srun: -------------------- --------------------
srun: exclusive           : exclusive
srun: gres                : gres:gpu:TitanX:8
srun: partition           : ClsParSystems
srun: reservation         : Spring2025Class
srun: time                : 01:00:00
srun: verbose             : 1
srun: -------------------- --------------------
srun: end of defined options
srun: jobid 236620: nodes(1):`GPU15', cpu counts: 16(x1)
srun: launch/slurm: launch_p_step_launch: CpuBindType=(null type)
srun: launching StepId=236620.0 on host GPU15, 1 tasks: 0
srun: route/default: init: route default plugin loaded
srun: launch/slurm: _task_start: Node GPU15, 1 tasks started
Generated 1000000 strings, total size 13504588 bytes.

--- Running CPU Implementation ---
[CPU] Time: 7586.945 ms | False Negatives: 0/1000000

--- Running GPU Implementation ---
[GPU] Time: 21.902 ms (346.41x speedup) | False Negatives: 0/1000000

srun: launch/slurm: _task_finish: Received task exit notification for 1 task of StepId=236620.0 (status=0x0000).
srun: launch/slurm: _task_finish: GPU15: task 0: Completed
[akmal@forest.usf.edu@gaivi-login1 proj3]$
```

/apps/GPU_course/runScript.sh proj3_akmal.cu 1000000 0.00001 128

```
[akmal@forest.usf.edu@gaivi-login1 proj3]$ /apps/GPU_course/runScript.sh proj3_akmal.cu 1000000 0.00001 128
----------------- nvcc Info: --------------------
nvcc: NVIDIA (R) Cuda compiler driver
Copyright (c) 2005-2021 NVIDIA Corporation
Built on Wed_Jun__2_19:15:15_PDT_2021
Cuda compilation tools, release 11.4, V11.4.48
Build cuda_11.4.r11.4/compiler.30033411_0
----------------- Compiling --------------------
ptxas info    : 0 bytes gmem
ptxas info    : Compiling entry function '_Z12check_kernelPKjPKcPKiPihym' for 'sm_52'
ptxas info    : Function properties for _Z12check_kernelPKjPKcPKiPihym
    0 bytes stack frame, 0 bytes spill stores, 0 bytes spill loads
ptxas info    : Used 44 registers, 376 bytes cmem[0], 8 bytes cmem[2]
ptxas info    : Compiling entry function '_Z10add_kernelPjPKcPKihym' for 'sm_52'
ptxas info    : Function properties for _Z10add_kernelPjPKcPKihym
    0 bytes stack frame, 0 bytes spill stores, 0 bytes spill loads
ptxas info    : Used 44 registers, 368 bytes cmem[0], 8 bytes cmem[2]
----------------- Executing --------------------
srun: defined options
srun: -------------------- --------------------
srun: exclusive           : exclusive
srun: gres                : gres:gpu:TitanX:8
srun: partition           : ClsParSystems
srun: reservation         : Spring2025Class
srun: time                : 01:00:00
srun: verbose             : 1
srun: -------------------- --------------------
srun: end of defined options
srun: jobid 236621: nodes(1):`GPU15', cpu counts: 16(x1)
srun: launch/slurm: launch_p_step_launch: CpuBindType=(null type)
srun: launching StepId=236621.0 on host GPU15, 1 tasks: 0
srun: route/default: init: route default plugin loaded
srun: launch/slurm: _task_start: Node GPU15, 1 tasks started
Generated 1000000 strings, total size 13504588 bytes.

--- Running CPU Implementation ---
[CPU] Time: 7578.984 ms | False Negatives: 0/1000000

--- Running GPU Implementation ---
[GPU] Time: 21.874 ms (346.49x speedup) | False Negatives: 0/1000000

srun: launch/slurm: _task_finish: Received task exit notification for 1 task of StepId=236621.0 (status=0x0000).
srun: launch/slurm: _task_finish: GPU15: task 0: Completed
[akmal@forest.usf.edu@gaivi-login1 proj3]$
```

/apps/GPU_course/runScript.sh proj3_akmal.cu 1000000 0.00001 256

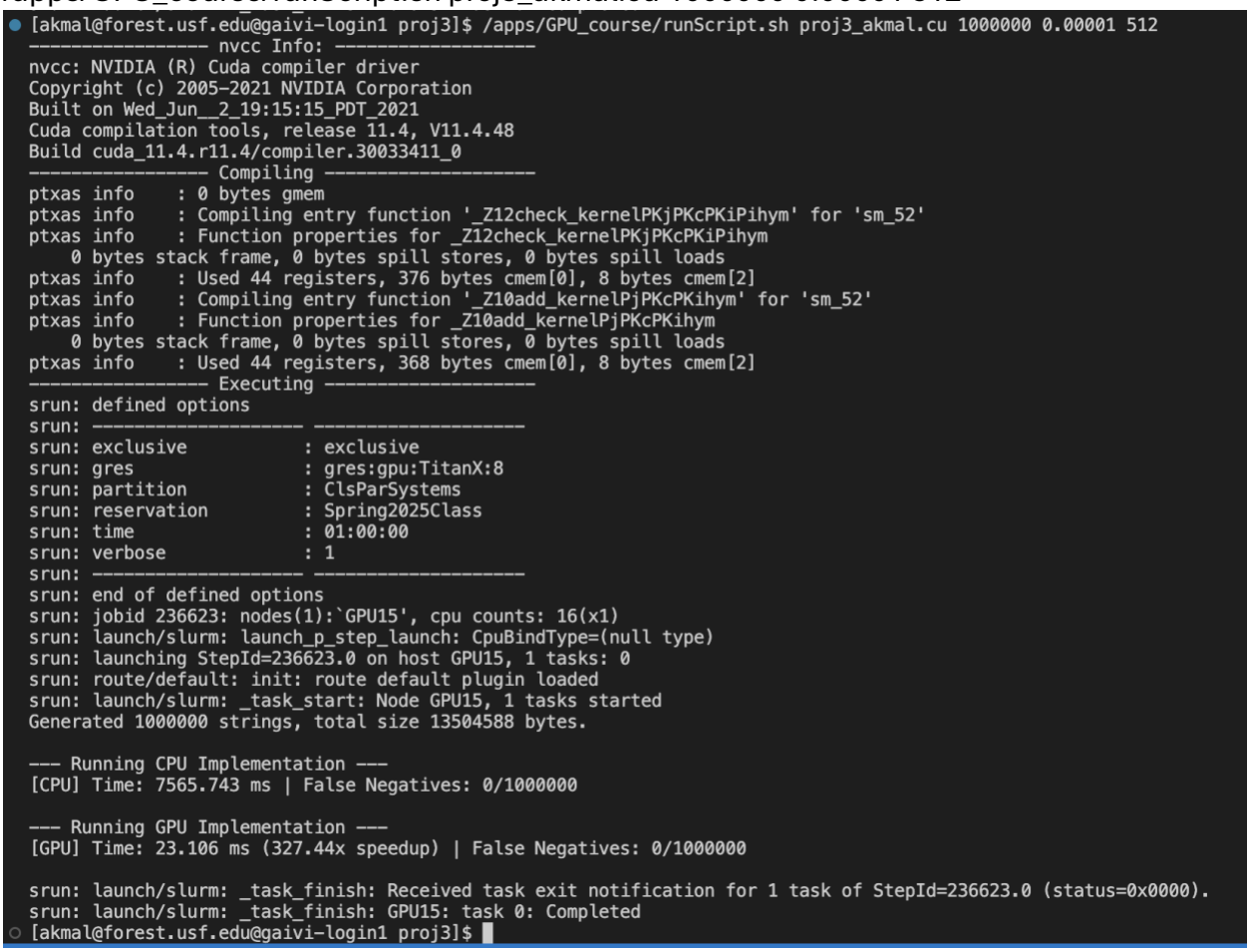
```
[akmal@forest.usf.edu@gaivi-login1 proj3]$ /apps/GPU_course/runScript.sh proj3_akmal.cu 1000000 0.00001 256
----------------- nvcc Info: --------------------
nvcc: NVIDIA (R) Cuda compiler driver
Copyright (c) 2005-2021 NVIDIA Corporation
Built on Wed_Jun__2_19:15:15_PDT_2021
Cuda compilation tools, release 11.4, V11.4.48
Build cuda_11.4.r11.4/compiler.30033411_0
----------------- Compiling --------------------
ptxas info    : 0 bytes gmem
ptxas info    : Compiling entry function '_Z12check_kernelPKjPKcPKiPihym' for 'sm_52'
ptxas info    : Function properties for _Z12check_kernelPKjPKcPKiPihym
    0 bytes stack frame, 0 bytes spill stores, 0 bytes spill loads
ptxas info    : Used 44 registers, 376 bytes cmem[0], 8 bytes cmem[2]
ptxas info    : Compiling entry function '_Z10add_kernelPjPKcPKihym' for 'sm_52'
ptxas info    : Function properties for _Z10add_kernelPjPKcPKihym
    0 bytes stack frame, 0 bytes spill stores, 0 bytes spill loads
ptxas info    : Used 44 registers, 368 bytes cmem[0], 8 bytes cmem[2]
----------------- Executing --------------------
srun: defined options
srun: -------------------- --------------------
srun: exclusive           : exclusive
srun: gres                : gres:gpu:TitanX:8
srun: partition           : ClsParSystems
srun: reservation         : Spring2025Class
srun: time                : 01:00:00
srun: verbose             : 1
srun: -------------------- --------------------
srun: end of defined options
srun: jobid 236622: nodes(1):`GPU15', cpu counts: 16(x1)
srun: launch/slurm: launch_p_step_launch: CpuBindType=(null type)
srun: launching StepId=236622.0 on host GPU15, 1 tasks: 0
srun: route/default: init: route default plugin loaded
srun: launch/slurm: _task_start: Node GPU15, 1 tasks started
Generated 1000000 strings, total size 13504588 bytes.

--- Running CPU Implementation ---
[CPU] Time: 7717.027 ms | False Negatives: 0/1000000

--- Running GPU Implementation ---
[GPU] Time: 22.034 ms (350.23x speedup) | False Negatives: 0/1000000

srun: launch/slurm: _task_finish: Received task exit notification for 1 task of StepId=236622.0 (status=0x0000)
srun: launch/slurm: _task_finish: GPU15: task 0: Completed
[akmal@forest.usf.edu@gaivi-login1 proj3]$
```

/apps/GPU_course/runScript.sh proj3_akmal.cu 1000000 0.00001 512



/apps/GPU_course/runScript.sh proj3_akmal.cu 1000000 0.01 64

```
[akmal@forest.usf.edu@gaivi-login1 proj3]$ /apps/GPU_course/runScript.sh proj3_akmal.cu 1000000 0.01 64
----------------- nvcc Info: -------------------
nvcc: NVIDIA (R) Cuda compiler driver
Copyright (c) 2005-2021 NVIDIA Corporation
Built on Wed_Jun__2_19:15:15_PDT_2021
Cuda compilation tools, release 11.4, V11.4.48
Build cuda_11.4.r11.4/compiler.30033411_0
----------------- Compiling --------------------
ptxas info    : 0 bytes gmem
ptxas info    : Compiling entry function '_Z12check_kernelPKjPKcPKiPihym' for 'sm_52'
ptxas info    : Function properties for _Z12check_kernelPKjPKcPKiPihym
    0 bytes stack frame, 0 bytes spill stores, 0 bytes spill loads
ptxas info    : Used 44 registers, 376 bytes cmem[0], 8 bytes cmem[2]
ptxas info    : Compiling entry function '_Z10add_kernelPjPKcPKihym' for 'sm_52'
ptxas info    : Function properties for _Z10add_kernelPjPKcPKihym
    0 bytes stack frame, 0 bytes spill stores, 0 bytes spill loads
ptxas info    : Used 44 registers, 368 bytes cmem[0], 8 bytes cmem[2]
----------------- Executing --------------------
srun: defined options
srun: -------------------- --------------------
srun: exclusive           : exclusive
srun: gres                : gres:gpu:TitanX:8
srun: partition           : ClsParSystems
srun: reservation         : Spring2025Class
srun: time                : 01:00:00
srun: verbose             : 1
srun: -------------------- --------------------
srun: end of defined options
srun: jobid 236624: nodes(1):`GPU15', cpu counts: 16(x1)
srun: launch/slurm: launch_p_step_launch: CpuBindType=(null type)
srun: launching StepId=236624.0 on host GPU15, 1 tasks: 0
srun: route/default: init: route default plugin loaded
srun: launch/slurm: _task_start: Node GPU15, 1 tasks started
Generated 1000000 strings, total size 13504588 bytes.

--- Running CPU Implementation ---
[CPU] Time: 3054.571 ms | False Negatives: 0/1000000

--- Running GPU Implementation ---
[GPU] Time: 9.570 ms (319.19x speedup) | False Negatives: 0/1000000

srun: launch/slurm: _task_finish: Received task exit notification for 1 task of StepId=236624.0 (status=0x0000).
srun: launch/slurm: _task_finish: GPU15: task 0: Completed
[akmal@forest.usf.edu@gaivi-login1 proj3]$
```

/apps/GPU_course/runScript.sh proj3_akmal.cu 1000000 0.01 128

```
[akmal@forest.usf.edu@gaivi-login1 proj3]$ /apps/GPU_course/runScript.sh proj3_akmal.cu 1000000 0.01 128
---------------- nvcc Info: -------------------
nvcc: NVIDIA (R) Cuda compiler driver
Copyright (c) 2005-2021 NVIDIA Corporation
Built on Wed_Jun__2_19:15:15_PDT_2021
Cuda compilation tools, release 11.4, V11.4.48
Build cuda_11.4.r11.4/compiler.30033411_0
---------------- Compiling --------------------
ptxas info    : 0 bytes gmem
ptxas info    : Compiling entry function '_Z12check_kernelPKjPKcPKiPihym' for 'sm_52'
ptxas info    : Function properties for _Z12check_kernelPKjPKcPKiPihym
    0 bytes stack frame, 0 bytes spill stores, 0 bytes spill loads
ptxas info    : Used 44 registers, 376 bytes cmem[0], 8 bytes cmem[2]
ptxas info    : Compiling entry function '_Z10add_kernelPjPKcPKihym' for 'sm_52'
ptxas info    : Function properties for _Z10add_kernelPjPKcPKihym
    0 bytes stack frame, 0 bytes spill stores, 0 bytes spill loads
ptxas info    : Used 44 registers, 368 bytes cmem[0], 8 bytes cmem[2]
---------------- Executing --------------------
srun: defined options
srun: -------------------- --------------------
srun: exclusive           : exclusive
srun: gres                : gres:gpu:TitanX:8
srun: partition           : ClsParSystems
srun: reservation         : Spring2025Class
srun: time                : 01:00:00
srun: verbose             : 1
srun: -------------------- --------------------
srun: end of defined options
srun: jobid 236626: nodes(1):`GPU15', cpu counts: 16(x1)
srun: launch/slurm: launch_p_step_launch: CpuBindType=(null type)
srun: launching StepId=236626.0 on host GPU15, 1 tasks: 0
srun: route/default: init: route default plugin loaded
srun: launch/slurm: _task_start: Node GPU15, 1 tasks started
Generated 1000000 strings, total size 13504588 bytes.

--- Running CPU Implementation ---
[CPU] Time: 3108.260 ms | False Negatives: 0/1000000

--- Running GPU Implementation ---
[GPU] Time: 9.534 ms (326.03x speedup) | False Negatives: 0/1000000

srun: launch/slurm: _task_finish: Received task exit notification for 1 task of StepId=236626.0 (status=0x0000).
srun: launch/slurm: _task_finish: GPU15: task 0: Completed
[akmal@forest.usf.edu@gaivi-login1 proj3]$
```

/apps/GPU_course/runScript.sh proj3_akmal.cu 1000000 0.01 256

```
[akmal@forest.usf.edu@gaivi-login1 proj3]$ /apps/GPU_course/runScript.sh proj3_akmal.cu 1000000 0.01 256
------------------ nvcc Info: --------------------
nvcc: NVIDIA (R) Cuda compiler driver
Copyright (c) 2005-2021 NVIDIA Corporation
Built on Wed_Jun__2_19:15:15_PDT_2021
Cuda compilation tools, release 11.4, V11.4.48
Build cuda_11.4.r11.4/compiler.30033411_0
------------------ Compiling ---------------------
ptxas info    : 0 bytes gmem
ptxas info    : Compiling entry function '_Z12check_kernelPKjPKcPKiPihym' for 'sm_52'
ptxas info    : Function properties for _Z12check_kernelPKjPKcPKiPihym
    0 bytes stack frame, 0 bytes spill stores, 0 bytes spill loads
ptxas info    : Used 44 registers, 376 bytes cmem[0], 8 bytes cmem[2]
ptxas info    : Compiling entry function '_Z10add_kernelPjPKcPKihym' for 'sm_52'
ptxas info    : Function properties for _Z10add_kernelPjPKcPKihym
    0 bytes stack frame, 0 bytes spill stores, 0 bytes spill loads
ptxas info    : Used 44 registers, 368 bytes cmem[0], 8 bytes cmem[2]
------------------ Executing ---------------------
srun: defined options
srun: -------------------- --------------------
srun: exclusive           : exclusive
srun: gres                : gres:gpu:TitanX:8
srun: partition           : ClsParSystems
srun: reservation         : Spring2025Class
srun: time                : 01:00:00
srun: verbose             : 1
srun: -------------------- --------------------
srun: end of defined options
srun: jobid 236627: nodes(1):`GPU15', cpu counts: 16(x1)
srun: launch/slurm: launch_p_step_launch: CpuBindType=(null type)
srun: launching StepId=236627.0 on host GPU15, 1 tasks: 0
srun: route/default: init: route default plugin loaded
srun: launch/slurm: _task_start: Node GPU15, 1 tasks started
Generated 1000000 strings, total size 13504588 bytes.

--- Running CPU Implementation ---
[CPU] Time: 3057.033 ms | False Negatives: 0/1000000

--- Running GPU Implementation ---
[GPU] Time: 9.530 ms (320.78x speedup) | False Negatives: 0/1000000

srun: launch/slurm: _task_finish: Received task exit notification for 1 task of StepId=236627.0 (status=0x0000).
srun: launch/slurm: _task_finish: GPU15: task 0: Completed
[akmal@forest.usf.edu@gaivi-login1 proj3]$
```

/apps/GPU_course/runScript.sh proj3_akmal.cu 1000000 0.01 512

```
[akmal@forest.usf.edu@gaivi-login1 proj3]$ /apps/GPU_course/runScript.sh proj3_akmal.cu 1000000 0.01 512
----------------- nvcc Info: -------------------
nvcc: NVIDIA (R) Cuda compiler driver
Copyright (c) 2005-2021 NVIDIA Corporation
Built on Wed_Jun__2_19:15:15_PDT_2021
Cuda compilation tools, release 11.4, V11.4.48
Build cuda_11.4.r11.4/compiler.30033411_0
----------------- Compiling --------------------
ptxas info    : 0 bytes gmem
ptxas info    : Compiling entry function '_Z12check_kernelPKjPKcPKiPihym' for 'sm_52'
ptxas info    : Function properties for _Z12check_kernelPKjPKcPKiPihym
    0 bytes stack frame, 0 bytes spill stores, 0 bytes spill loads
ptxas info    : Used 44 registers, 376 bytes cmem[0], 8 bytes cmem[2]
ptxas info    : Compiling entry function '_Z10add_kernelPjPKcPKihym' for 'sm_52'
ptxas info    : Function properties for _Z10add_kernelPjPKcPKihym
    0 bytes stack frame, 0 bytes spill stores, 0 bytes spill loads
ptxas info    : Used 44 registers, 368 bytes cmem[0], 8 bytes cmem[2]
----------------- Executing --------------------
srun: defined options
srun: -------------------- --------------------
srun: exclusive           : exclusive
srun: gres                : gres:gpu:TitanX:8
srun: partition           : ClsParSystems
srun: reservation         : Spring2025Class
srun: time                : 01:00:00
srun: verbose             : 1
srun: -------------------- --------------------
srun: end of defined options
srun: jobid 236628: nodes(1):`GPU15', cpu counts: 16(x1)
srun: launch/slurm: launch_p_step_launch: CpuBindType=(null type)
srun: launching StepId=236628.0 on host GPU15, 1 tasks: 0
srun: route/default: init: route default plugin loaded
srun: launch/slurm: _task_start: Node GPU15, 1 tasks started
Generated 1000000 strings, total size 13504588 bytes.

--- Running CPU Implementation ---
[CPU] Time: 3080.663 ms | False Negatives: 0/1000000

--- Running GPU Implementation ---
[GPU] Time: 10.025 ms (307.29x speedup) | False Negatives: 0/1000000

srun: launch/slurm: _task_finish: Received task exit notification for 1 task of StepId=236628.0 (status=0x0000).
srun: launch/slurm: _task_finish: GPU15: task 0: Completed
[akmal@forest.usf.edu@gaivi-login1 proj3]$
```

/apps/GPU_course/runScript.sh proj3_akmal.cu 10000 0.001 256

```
[akmal@forest.usf.edu@gaivi-login1 proj3]$ /apps/GPU_course/runScript.sh proj3_akmal.cu 10000 0.001 256
----------------- nvcc Info: --------------------
nvcc: NVIDIA (R) Cuda compiler driver
Copyright (c) 2005-2021 NVIDIA Corporation
Built on Wed_Jun__2_19:15:15_PDT_2021
Cuda compilation tools, release 11.4, V11.4.48
Build cuda_11.4.r11.4/compiler.30033411_0
----------------- Compiling --------------------
ptxas info    : 0 bytes gmem
ptxas info    : Compiling entry function '_Z12check_kernelPKjPKcPKiPihym' for 'sm_52'
ptxas info    : Function properties for _Z12check_kernelPKjPKcPKiPihym
    0 bytes stack frame, 0 bytes spill stores, 0 bytes spill loads
ptxas info    : Used 44 registers, 376 bytes cmem[0], 8 bytes cmem[2]
ptxas info    : Compiling entry function '_Z10add_kernelPjPKcPKihym' for 'sm_52'
ptxas info    : Function properties for _Z10add_kernelPjPKcPKihym
    0 bytes stack frame, 0 bytes spill stores, 0 bytes spill loads
ptxas info    : Used 44 registers, 368 bytes cmem[0], 8 bytes cmem[2]
----------------- Executing --------------------
srun: defined options
srun: -------------------- --------------------
srun: exclusive           : exclusive
srun: gres                : gres:gpu:TitanX:8
srun: partition           : ClsParSystems
srun: reservation         : Spring2025Class
srun: time                : 01:00:00
srun: verbose             : 1
srun: -------------------- --------------------
srun: end of defined options
srun: jobid 236629: nodes(1):`GPU15', cpu counts: 16(x1)
srun: launch/slurm: launch_p_step_launch: CpuBindType=(null type)
srun: launching StepId=236629.0 on host GPU15, 1 tasks: 0
srun: route/default: init: route default plugin loaded
srun: launch/slurm: _task_start: Node GPU15, 1 tasks started
Generated 10000 strings, total size 135309 bytes.

--- Running CPU Implementation ---
[CPU] Time: 42.012 ms | False Negatives: 0/10000

--- Running GPU Implementation ---
[GPU] Time: 0.220 ms (190.91x speedup) | False Negatives: 0/10000

srun: launch/slurm: _task_finish: Received task exit notification for 1 task of StepId=236629.0 (status=0x0000).
srun: launch/slurm: _task_finish: GPU15: task 0: Completed
[akmal@forest.usf.edu@gaivi-login1 proj3]$
```

/apps/GPU_course/runScript.sh proj3_akmal.cu 1000000 0.001 256

```
[akmal@forest.usf.edu@gaivi-login1 proj3]$ /apps/GPU_course/runScript.sh proj3_akmal.cu 1000000 0.001 256
----------------- nvcc Info: -------------------
nvcc: NVIDIA (R) Cuda compiler driver
Copyright (c) 2005-2021 NVIDIA Corporation
Built on Wed_Jun__2_19:15:15_PDT_2021
Cuda compilation tools, release 11.4, V11.4.48
Build cuda_11.4.r11.4/compiler.30033411_0
----------------- Compiling --------------------
ptxas info    : 0 bytes gmem
ptxas info    : Compiling entry function '_Z12check_kernelPKjPKcPKiPihym' for 'sm_52'
ptxas info    : Function properties for _Z12check_kernelPKjPKcPKiPihym
    0 bytes stack frame, 0 bytes spill stores, 0 bytes spill loads
ptxas info    : Used 44 registers, 376 bytes cmem[0], 8 bytes cmem[2]
ptxas info    : Compiling entry function '_Z10add_kernelPjPKcPKihym' for 'sm_52'
ptxas info    : Function properties for _Z10add_kernelPjPKcPKihym
    0 bytes stack frame, 0 bytes spill stores, 0 bytes spill loads
ptxas info    : Used 44 registers, 368 bytes cmem[0], 8 bytes cmem[2]
----------------- Executing --------------------
srun: defined options
srun: -------------------- --------------------
srun: exclusive           : exclusive
srun: gres                : gres:gpu:TitanX:8
srun: partition           : ClsParSystems
srun: reservation         : Spring2025Class
srun: time                : 01:00:00
srun: verbose             : 1
srun: -------------------- --------------------
srun: end of defined options
srun: jobid 236630: nodes(1):`GPU15', cpu counts: 16(x1)
srun: launch/slurm: launch_p_step_launch: CpuBindType=(null type)
srun: launching StepId=236630.0 on host GPU15, 1 tasks: 0
srun: route/default: init: route default plugin loaded
srun: launch/slurm: _task_start: Node GPU15, 1 tasks started
Generated 1000000 strings, total size 13504588 bytes.

--- Running CPU Implementation ---
[CPU] Time: 4289.558 ms | False Negatives: 0/1000000

--- Running GPU Implementation ---
[GPU] Time: 13.172 ms (325.66x speedup) | False Negatives: 0/1000000

srun: launch/slurm: _task_finish: Received task exit notification for 1 task of StepId=236630.0 (status=0x0000).
srun: launch/slurm: _task_finish: GPU15: task 0: Completed
[akmal@forest.usf.edu@gaivi-login1 proj3]$
```

/apps/GPU_course/runScript.sh proj3_akmal.cu 100000000 0.001 256

```
[akmal@forest.usf.edu@gaivi-login1 proj3]$ /apps/GPU_course/runScript.sh proj3_akmal.cu 100000000 0.001 256
---------------- nvcc Info: --------------------
nvcc: NVIDIA (R) Cuda compiler driver
Copyright (c) 2005-2021 NVIDIA Corporation
Built on Wed_Jun__2_19:15:15_PDT_2021
Cuda compilation tools, release 11.4, V11.4.48
Build cuda_11.4.r11.4/compiler.30033411_0
---------------- Compiling --------------------
ptxas info    : 0 bytes gmem
ptxas info    : Compiling entry function '_Z12check_kernelPKjPKcPKiPihym' for 'sm_52'
ptxas info    : Function properties for _Z12check_kernelPKjPKcPKiPihym
    0 bytes stack frame, 0 bytes spill stores, 0 bytes spill loads
ptxas info    : Used 44 registers, 376 bytes cmem[0], 8 bytes cmem[2]
ptxas info    : Compiling entry function '_Z10add_kernelPjPKcPKihym' for 'sm_52'
ptxas info    : Function properties for _Z10add_kernelPjPKcPKihym
    0 bytes stack frame, 0 bytes spill stores, 0 bytes spill loads
ptxas info    : Used 44 registers, 368 bytes cmem[0], 8 bytes cmem[2]
---------------- Executing --------------------
srun: defined options
srun: -------------------- --------------------
srun: exclusive           : exclusive
srun: gres                : gres:gpu:TitanX:8
srun: partition           : ClsParSystems
srun: reservation         : Spring2025Class
srun: time                : 01:00:00
srun: verbose             : 1
srun: -------------------- --------------------
srun: end of defined options
srun: jobid 236631: nodes(1):`GPU15', cpu counts: 16(x1)
srun: launch/slurm: launch_p_step_launch: CpuBindType=(null type)
srun: launching StepId=236631.0 on host GPU15, 1 tasks: 0
srun: route/default: init: route default plugin loaded
srun: launch/slurm: _task_start: Node GPU15, 1 tasks started
Generated 100000000 strings, total size 1350025518 bytes.

--- Running CPU Implementation ---
[CPU] Time: 569421.562 ms | False Negatives: 0/100000000

--- Running GPU Implementation ---
[GPU] Time: 1799.438 ms (316.44x speedup) | False Negatives: 0/100000000

srun: launch/slurm: _task_finish: Received task exit notification for 1 task of StepId=236631.0 (status=0x0000).
srun: launch/slurm: _task_finish: GPU15: task 0: Completed
```

I varied three parameters:

1. **Block size** (threads per block): 64, 128, 256, 512
2. **False-positive rate** (p): 1×10^{-2} and 1×10^{-5}
3. **Dataset size** (n): 1×10^{4}, 1×10^{6}, and 1×10^{8} strings

For each configuration, I launched enough blocks to cover all n threads (grid size = ceil(n/blockDim)) and measured the combined insertion + query time on both the GPU and the CPU reference implementation.

- **Optimal block size**: Across all tests, **256 threads per block** delivered the highest occupancy and best performance.
- **Speedups at $n = 1 \times 10^{6}$**:
  - $p = 0.01$ → from 307× to 326× faster than CPU
  - $p = 1 \times 10^{-5}$ → from 327× to 350× faster than CPU
- **Effect of dataset scaling** (with $p = 1 \times 10^{-3}$, block size = 256):
  - $n = 1 \times 10^{4}$ → 190× speedup

- $n = 1 \times 10^6 \rightarrow 325\times$ speedup
- $n = 1 \times 10^8 \rightarrow 316\times$ speedup

These results confirm that the GPU implementation consistently outperforms the serial CPU baseline by roughly **300×** across varying error rates and data sizes, demonstrating both scalability and robustness of the chosen optimizations.

Conclusion

GPU - accelerated Bloom filter achieves dramatic performance gains - on the order of **300x** - over the serial CPU implementation, while preserving the same false-positive characteristics. By moving all SipHash computations into registers, mapping one thread per string, and employing fine-grained `atomicOr()` updates alongside coalesced memory accesses and early-exit logic, I fully exploit the GPU's massive parallelism and memory bandwidth.