

Các điểm sửa đổi so với lần báo cáo: Thêm các ảnh chụp chạy với các blockSize khác nhau (từ slide 36 đến hết)



Báo cáo đồ án cuối kỳ

1. Trần Nhật Huy – 1612272
2. Nguyễn Thành Đạt - 1612088

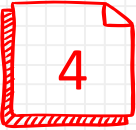


CONTENTS

- ❖ Quá trình tối ưu
- ❖ Tổng kết



Cài đặt tuần tự thuật toán Radix sort tuần tự



- ❑ File code: Version1.cu
- ❑ Ý tưởng: Đã học trên lớp

```
> nvcc Version1.cu 86 ./a.out 4 512 512
*****GPU info*****
Name: GeForce RTX 2070 with Max-Q Design
Compute capability: 7.5
Num SMs: 36
Max num threads per SM: 1024
Max num warps per SM: 32
GMEM: 8361279488 byte
SMEM per SM: 65536 byte
SMEM per block: 49152 byte
*****

Input size: 16777217

Num bits per digit: 4

Hist block size: 512, scan block size: 512

Radix sort by host
Time: 852.525 ms

Radix sort by device
Time: 561.608 ms
CORRECT :)
```

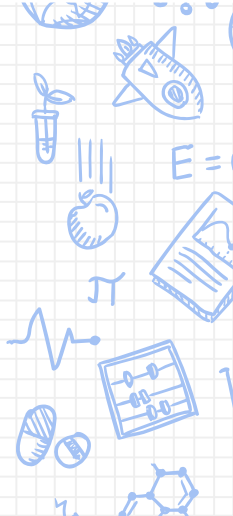




Cài đặt tuần tự thuật toán Radix sort tuần tự

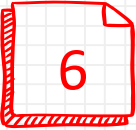


Version	histogramKernel	scanBlkKernel	addBlkKernel	transposeKernel	scatterKernel	Total
Baseline1						852.525
Thrust						44.17





Cài đặt song song 2 bước histogram và scan



- ❑ File code: Version1.cu
- ❑ Ý tưởng: Đã học trên lớp

```
> nvcc Version1.cu 86 ./a.out 4 512 512
*****GPU info*****
Name: GeForce RTX 2070 with Max-Q Design
Compute capability: 7.5
Num SMs: 36
Max num threads per SM: 1024
Max num warps per SM: 32
GMEM: 8361279488 byte
SMEM per SM: 65536 byte
SMEM per block: 49152 byte
*****

Input size: 16777217

Num bits per digit: 4

Hist block size: 512, scan block size: 512

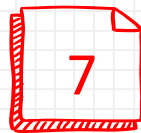
Radix sort by host
Time: 852.525 ms

Radix sort by device
Time: 561.608 ms
CORRECT :)
```





Cài đặt song song 2 bước histogram và scan

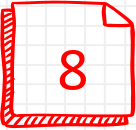


Version	histogramKernel	scanBlkKernel	addBlkKernel	transposeKernel	scatterKernel	Total
Baseline1						852.525
Baseline2						561.608
Thrust						44.17





Cài đặt song song Radix sort với $k=1$



- ❑ File code: Version2.cu
- ❑ Ý tưởng: Đã học trên lớp

```
> nvcc Version2.cu 86 ./a.out 4 512 512
*****GPU info*****
Name: GeForce RTX 2070 with Max-Q Design
Compute capability: 7.5
Num SMs: 36
Max num threads per SM: 1024
Max num warps per SM: 32
GMEM: 8361279488 byte
SMEM per SM: 65536 byte
SMEM per block: 49152 byte
*****

Input size: 16777217

Num bits per digit: 4

Hist block size: 512, scan block size: 512, scatter block size: 512

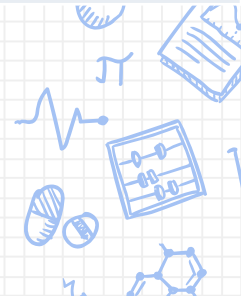
Radix sort by host
Time: 856.344 ms

Radix sort by device
Time: 1734.354 ms
CORRECT :)
```



Cài đặt song song Radix sort với $k=1$

Version	histogramKernel	scanBlkKernel	addBlkKernel	transposeKernel	scatterKernel	Total
Baseline1						852.525
Baseline2						561.608
Baseline3						1734.354
Thrust						44.17





Cài đặt tuần tự Radix sort tuần tự theo ý tưởng bài báo

10

- ☐ File code: Version3.cu
- ☐ Ý tưởng: Đã học trên lớp

```
> nvcc Version3.cu -x cuda -o ./a.out 4 512 512
*****GPU info*****
Name: GeForce RTX 2070 with Max-Q Design
Compute capability: 7.5
Num SMs: 36
Max num threads per SM: 1024
Max num warps per SM: 32
GMEM: 8361279488 byte
SMEM per SM: 65536 byte
SMEM per block: 49152 byte
*****

Input size: 16777217

Num bits per digit: 4

Hist block size: 512, scan block size: 512

Radix sort by library
Time: 44.152 ms

Radix sort by host
Time: 19602.135 ms
CORRECT :)

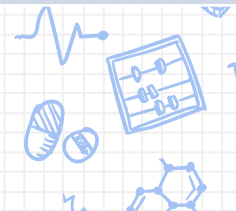
Radix sort by device
Time: 0.011 ms
INCORRECT :(
```



Cài đặt tuần tự Radix sort theo ý tưởng bài báo

11

Version	histogramKernel	scanBlkKernel	addBlkKernel	transposeKernel	scatterKernel	Total
Baseline1						852.525
Baseline2						561.608
Baseline3						1734.354
Baseline4						19602.135
Thrust						44.17





Tối ưu hóa lần 1

- ❑ File code: Parallel_v1.cu
- ❑ Ý tưởng: ta sẽ tiến hành song song hóa các bước ở baseline 4 để giảm tổng thời gian thực hiện. Code song song dựa theo bài báo

```
> nvcc Parallel_v1.cu 86 ./a.out 4 512 512
*****GPU info*****
Name: GeForce RTX 2070 with Max-Q Design
Compute capability: 7.5
Num SMs: 36
Max num threads per SM: 1024
Max num warps per SM: 32
GMEM: 8361279488 byte
SMEM per SM: 65536 byte
SMEM per block: 49152 byte
*****

Input size: 16777217

Num bits per digit: 4

Hist block size: 512, scan block size: 512

Radix sort by library
Time: 44.450 ms

Radix sort by host
Time: 0.010 ms
INCORRECT :(

Radix sort by device
Hist Time: 13.332
Scan Time: 0.782
Add Time: 0.179
Transpose Time: 0.529
Scatter Time: 117.443
Time: 171.905 ms
CORRECT :)
```



Tối ưu hóa lần 1

13

Version	histogramKernel	scanBlkKernel	addBlkKernel	transposeKernel	scatterKernel	Total
Baseline1						852.525
Baseline2						561.608
Baseline3						1734.354
Baseline4						19602.135
Parallel_v1	13.332	0.782	0.179	0.529	117.443	171.905
Thrust						44.17





Tối ưu hóa lần 2

14



- ❑ File code: Parallel_v2.cu
- ❑ Ý tưởng: Hiện tại, thuật toán chậm nhất ở phần Scatter. Lý do là vì nhóm đang sử dụng atomicAdd để tính rank. Nhóm sẽ viết lại thuật toán Scatter y như trong bài báo, sử dụng Radix Sort với $k = 1$ để sắp xếp sau đó tính rank của các phần tử trong Block

```
> nvcc Parallel_v2.cu 86 ./a.out 4 512 512
*****GPU info*****
Name: GeForce RTX 2070 with Max-Q Design
Compute capability: 7.5
Num SMs: 36
Max num threads per SM: 1024
Max num warps per SM: 32
GMEM: 8361279488 byte
SMEM per SM: 65536 byte
SMEM per block: 49152 byte
*****

Input size: 16777217

Num bits per digit: 4

Hist block size: 512, scan block size: 512

Radix sort by library
Time: 43.824 ms

Radix sort by host
Time: 0.010 ms
INCORRECT :(

Radix sort by device
Hist Time: 13.324
Scan Time: 0.780
Add Time: 0.179
Transpose Time: 0.531
Scatter Time: 73.052
Time: 128.528 ms
CORRECT :)
```





Cài đặt tuần tự thuật toán Radix sort tuần tự

Version	histogramKernel	scanBlkKernel	addBlkKernel	transposeKernel	scatterKernel	Total
Baseline1						852.525
Baseline2						561.608
Baseline3						1734.354
Baseline4						19602.135
Parallel_v1	13.332	0.782	0.179	0.529	117.443	171.905
Parallel_v2	13.324	0.780	0.179	0.531	73.052	128.528
Thrust						44.17



Tối ưu hóa lần 4

16

- ❑ File code: Parallel_v4.cu
- ❑ Ý tưởng: Ở bước này, nhóm sẽ vẫn tối ưu hóa Scatter kernel. Ở phiên bản Parallel_v1 thì trong Scatter Kernel nhóm sử dụng mảng smem để lưu dữ liệu và kích thước của mảng này không hề nhỏ khi nhóm lưu tới 6 loại mảng. Song song với đó là nhóm sử dụng nhiều câu lệnh `__syncthreads()`
- ❑ Thiết kế: Thay vì smem trong Scatter kernel phải lưu 6 mảng khác nhau thì giờ đây nhóm sẽ thiết kế lại:
 - Phần 1 vẫn sẽ lưu input kích thước `blockDim.x` phần tử
 - Phần 2 lưu một phần tử giả, gọi là dummy, có 1 phần tử
 - Phần 3 lưu chuỗi nhị phân lấy được, kích thước `blockDim.x` phần tử
 - Phần 4 lưu chỉ số bắt đầu của mảng sau khi đã sắp xếp, kích thước 2^n phần tử





Tối ưu hóa lần 4

17

```
> nvcc Parallel_v4.cu 86 ./a.out 4 512 512
*****GPU info*****
Name: GeForce RTX 2070 with Max-Q Design
Compute capability: 7.5
Num SMs: 36
Max num threads per SM: 1024
Max num warps per SM: 32
GMEM: 8361279488 byte
SMEM per SM: 65536 byte
SMEM per block: 49152 byte
*****

Input size: 16777217

Num bits per digit: 4

Hist block size: 512, scan block size: 512

Radix sort by library
Time: 46.939 ms

Radix sort by host
Time: 0.018 ms
INCORRECT :(

Radix sort by device
Hist Time: 13.648
Scan Time: 0.772
Add Time: 0.182
Transpose Time: 0.535
Scatter Time: 67.406
Time: 123.201 ms
CORRECT :)
```





Tối ưu hóa lần 4

18

Version	histogramKernel	scanBlkKernel	addBlkKernel	transposeKernel	scatterKernel	Total
Baseline1						852.525
Baseline2						561.608
Baseline3						1734.354
Baseline4						19602.135
Parallel_v1	13.332	0.782	0.179	0.529	117.443	171.905
Parallel_v2	13.324	0.780	0.179	0.531	73.052	128.528
Parallel_v4	13.327	0.762	0.175	0.529	59.814	113.641
Thrust						44.17



Tối ưu hóa lần 5

19

- ❑ File code: Parallel_v5.cu
- ❑ Ý tưởng: Ở lần này nhóm vẫn tiếp tục tối ưu hóa hàm Scatter kernel vì trong hàm này vẫn còn có nhiều câu lệnh `__syncthreads()` dùng để debug nhưng chưa dám xóa và chưa load mảng `scanHistogramArrayTranpose` vào smem.
- ❑ Thiết kế: Ở lần trước thì smem gồm 4 phần dữ liệu, bây giờ ta sẽ thêm một phần chứa `scanHistogramArrayTranpose` cho từng block, sẽ có 2^n phần tử. Khi chép dữ liệu cần lưu ý rằng số thread có thể sẽ nhỏ hơn số `nBins`





Tối ưu hóa lần 5

20

```
> nvcc Parallel_v5.cu 66 ./a.out 4 512 512
*****GPU info*****
Name: GeForce RTX 2070 with Max-Q Design
Compute capability: 7.5
Num SMs: 36
Max num threads per SM: 1024
Max num warps per SM: 32
GMEM: 8361279488 byte
SMEM per SM: 65536 byte
SMEM per block: 49152 byte
*****

Input size: 16777217

Num bits per digit: 4

Hist block size: 512, scan block size: 512

Radix sort by library
Time: 42.396 ms

Radix sort by host
Time: 0.012 ms
INCORRECT :(

Radix sort by device
Hist Time: 13.318
Scan Time: 0.769
Add Time: 0.181
Transpose Time: 0.529
Scatter Time: 60.569
Time: 114.792 ms
CORRECT :)
```





Tối ưu hóa lần 5

21

Version	histogramKernel	scanBlkKernel	addBlkKernel	transposeKernel	scatterKernel	Total
Baseline1						852.525
Baseline2						561.608
Baseline3						1734.354
Baseline4						19602.135
Parallel_v1	13.332	0.782	0.179	0.529	117.443	171.905
Parallel_v2	13.324	0.780	0.179	0.531	73.052	128.528
Parallel_v4	13.648	0.772	0.182	0.535	67.406	123.201
Parallel_v5	13.318	0.769	0.181	0.529	60.569	114.792
Thrust						44.17



Tối ưu hóa lần 6

22

- ❑ File code: Parallel_v6.cu
- ❑ Ý tưởng: Ở lần này, nhóm chưa suy nghĩ ra cách nào để tối ưu hàm Scatter kernel nữa. Do đó, nhóm sẽ thực hiện tối ưu ở điểm khác. Tổng thời gian chạy các hàm kernel chiếm phần lớn thời gian thực thi nhưng thời gian còn nằm ở những phần không thuộc các hàm kernel và nhóm sẽ tối ưu các hàm cấp phát bộ nhớ xem có giảm tổng thời gian thực thi hay không
- ❑ Thiết kế: Ở phần cấp phát bộ nhớ, hiện tại thì nhóm đang lãng phí tài nguyên khi có thể tái sử dụng `d_histArr` thay vì tạo `d_scanHistArrTranspose`





Tối ưu hóa lần 6

23

```
> nvcc Parallel_v6.cu 86 ./a.out 4 512 512
*****GPU info*****
Name: GeForce RTX 2070 with Max-Q Design
Compute capability: 7.5
Num SMs: 36
Max num threads per SM: 1024
Max num warps per SM: 32
GMEM: 8361279488 byte
SMEM per SM: 65536 byte
SMEM per block: 49152 byte
*****

Input size: 16777217

Num bits per digit: 4

Hist block size: 512, scan block size: 512

Radix sort by library
Time: 43.375 ms

Radix sort by host
Time: 0.010 ms
INCORRECT :(

Radix sort by device
Hist Time: 13.324
Scan Time: 0.761
Add Time: 0.179
Transpose Time: 0.531
Scatter Time: 60.100
Time: 113.534 ms
CORRECT :)
```





Tối ưu hóa lần 6

24

Version	histogramKernel	scanBlkKernel	addBlkKernel	transposeKernel	scatterKernel	Total
Baseline1						852.525
Baseline2						561.608
Baseline3						1734.354
Baseline4						19602.135
Parallel_v1	13.332	0.782	0.179	0.529	117.443	171.905
Parallel_v2	13.324	0.780	0.179	0.531	73.052	128.528
Parallel_v4	13.327	0.762	0.175	0.529	59.814	113.641
Parallel_v5	13.318	0.769	0.181	0.529	60.569	114.792
Parallel_v6	13.324	0.761	0.179	0.531	60.1	113.534
Thrust						44.17





Tối ưu hóa lần 7

25

- ❑ File code: Parallel_v7.cu
- ❑ Ý tưởng: Với các phiên bản trước thì kernel tính hist sẽ sử dụng atomicAdd trên GMEM nên có tốc độ chậm
- ❑ Thiết kế: Nhóm sẽ tiến hành tối ưu kernel tính hist bằng cách sử dụng atomicAdd trên SMEM





Tối ưu hóa lần 7

26

```
> nvcc Parallel_v7.cu 86 ./a.out 4 512 512
*****GPU info*****
Name: GeForce RTX 2070 with Max-Q Design
Compute capability: 7.5
Num SMs: 36
Max num threads per SM: 1024
Max num warps per SM: 32
GMEM: 8361279488 byte
SMEM per SM: 65536 byte
SMEM per block: 49152 byte
*****

Input size: 16777217

Num bits per digit: 4

Hist block size: 512, scan block size: 512

Radix sort by library
Time: 45.707 ms

Radix sort by host
Time: 0.010 ms
INCORRECT :(

Radix sort by device
Hist Time: 4.457
Scan Time: 0.777
Add Time: 0.180
Transpose Time: 0.530
Scatter Time: 59.952
Time: 104.299 ms
CORRECT :)
```





Tối ưu hóa lần 7

27

Version	histogramKernel	scanBlkKernel	addBlkKernel	transposeKernel	scatterKernel	Total
Baseline1						852.525
Baseline2						561.608
Baseline3						1734.354
Baseline4						19602.135
Parallel_v1	13.332	0.782	0.179	0.529	117.443	171.905
Parallel_v2	13.324	0.780	0.179	0.531	73.052	128.528
Parallel_v4	13.327	0.762	0.175	0.529	59.814	113.641
Parallel_v5	13.318	0.769	0.181	0.529	60.569	114.792
Parallel_v6	13.324	0.761	0.179	0.531	60.1	113.534
Parallel_v7	4.457	0.777	0.18	0.53	59.952	104.299
Thrust						44.17



Tối ưu hóa lần 8

28

- ❑ File code: Parallel_v8.cu
- ❑ Ý tưởng: Hàm Scatter Kernel hiện tại đang vừa sort với scatter trong một hàm. Do đó, hàm này đang sử dụng rất nhiều smem để lưu trữ mặc dù nhóm đã có tối ưu phần này.
- ❑ Thiết kế: Nhóm sẽ tách Scatter Kernel làm 2 hàm riêng biệt là sortBlock với scatterBlock. Phần code thì vẫn tương tự như Scatter Kernel nhưng nhóm muốn thử nghiệm xem nó có nhanh hơn hay không





Tối ưu hóa lần 8

29

```
> nvcc Parallel_v8.cu 86 ./a.out 4 512 512
*****GPU info*****
Name: GeForce RTX 2070 with Max-Q Design
Compute capability: 7.5
Num SMs: 36
Max num threads per SM: 1024
Max num warps per SM: 32
GMEM: 8361279488 byte
SMEM per SM: 65536 byte
SMEM per block: 49152 byte
*****

Input size: 16777217

Num bits per digit: 4

Hist block size: 512, scan block size: 512

Radix sort by library
Time: 43.975 ms

Radix sort by host
Time: 0.010 ms
INCORRECT :(

Radix sort by device
Hist Time: 5.011
Scan Time: 0.760
Add Time: 0.183
Transpose Time: 0.531
Scatter Time: 63.540
Time: 109.859 ms
CORRECT :)
```





Tối ưu hóa lần 8

30

Version	histogramKernel	scanBlkKernel	addBlkKernel	transposeKernel	scatterKernel	Total
Baseline1						852.525
Baseline2						561.608
Baseline3						1734.354
Baseline4						19602.135
Parallel_v1	13.332	0.782	0.179	0.529	117.443	171.905
Parallel_v2	13.324	0.780	0.179	0.531	73.052	128.528
Parallel_v4	13.327	0.762	0.175	0.529	59.814	113.641
Parallel_v5	13.318	0.769	0.181	0.529	60.569	114.792
Parallel_v6	13.324	0.761	0.179	0.531	60.1	113.534
Parallel_v7	4.457	0.777	0.18	0.53	59.952	104.299
Parallel_v8	5.011	0.76	0.183	0.531	63.540	109.859
Thrust						44.17

Tổng kết

31

- ❑ Version nhanh nhất: Parallel_v7.cu
- ❑ Block Sizes là: 256 512
- ❑ $K=4$

```
> nvcc Parallel_v7.cu 86 ./a.out 4 256 512
*****GPU info*****
Name: GeForce RTX 2070 with Max-Q Design
Compute capability: 7.5
Num SMs: 36
Max num threads per SM: 1024
Max num warps per SM: 32
GMEM: 8361279488 byte
SMEM per SM: 65536 byte
SMEM per block: 49152 byte
*****

Input size: 16777217

Num bits per digit: 4

Hist block size: 256, scan block size: 512

Radix sort by library
Time: 45.181 ms

Radix sort by host
Time: 0.010 ms
INCORRECT :(

Radix sort by device
Hist Time: 4.297
Scan Time: 1.346
Add Time: 0.347
Transpose Time: 2.759
Scatter Time: 51.924
Time: 99.488 ms
CORRECT :)
```





Tổng kết

32

```
> nvcc Parallel_v7.cu && ./a.out 1 512 512
*****GPU info*****
Name: GeForce RTX 2070 with Max-Q Design
Compute capability: 7.5
Num SMs: 36
Max num threads per SM: 1024
Max num warps per SM: 32
GMEM: 8361279488 byte
SMEM per SM: 65536 byte
SMEM per block: 49152 byte
*****

Input size: 16777217

Num bits per digit: 1

Hist block size: 512, scan block size: 512

Radix sort by library
Time: 43.469 ms

Radix sort by host
Time: 0.010 ms
INCORRECT :(

Radix sort by device
Hist Time: 27.568
Scan Time: 1.036
Add Time: 0.385
Transpose Time: 0.418
Scatter Time: 75.325
Time: 145.759 ms
CORRECT :)
```





Tổng kết

33

```
> nvcc Parallel_v7.cu && ./a.out 2 512 512
*****GPU info*****
Name: GeForce RTX 2070 with Max-Q Design
Compute capability: 7.5
Num SMs: 36
Max num threads per SM: 1024
Max num warps per SM: 32
GMEM: 8361279488 byte
SMEM per SM: 65536 byte
SMEM per block: 49152 byte
*****

Input size: 16777217

Num bits per digit: 2

Hist block size: 512, scan block size: 512

Radix sort by library
Time: 47.711 ms

Radix sort by host
Time: 0.010 ms
INCORRECT :(

Radix sort by device
Hist Time: 10.909
Scan Time: 0.713
Add Time: 0.217
Transpose Time: 0.282
Scatter Time: 65.906
Time: 118.000 ms
CORRECT :)
```





Tổng kết

34

```
> nvcc Parallel_v7.cu 86 ./a.out 4 512 512
*****GPU info*****
Name: GeForce RTX 2070 with Max-Q Design
Compute capability: 7.5
Num SMs: 36
Max num threads per SM: 1024
Max num warps per SM: 32
GMEM: 8361279488 byte
SMEM per SM: 65536 byte
SMEM per block: 49152 byte
*****

Input size: 16777217

Num bits per digit: 4

Hist block size: 512, scan block size: 512

Radix sort by library
Time: 45.524 ms

Radix sort by host
Time: 0.010 ms
INCORRECT :(

Radix sort by device
Hist Time: 4.496
Scan Time: 0.765
Add Time: 0.179
Transpose Time: 0.531
Scatter Time: 57.933
Time: 103.144 ms
CORRECT :)
```





Tổng kết

35

```
> nvcc Parallel_v7.cu 86 ./a.out 8 512 512
*****GPU info*****
Name: GeForce RTX 2070 with Max-Q Design
Compute capability: 7.5
Num SMs: 36
Max num threads per SM: 1024
Max num warps per SM: 32
GMEM: 8361279488 byte
SMEM per SM: 65536 byte
SMEM per block: 49152 byte
*****

Input size: 16777217

Num bits per digit: 8

Hist block size: 512, scan block size: 512

Radix sort by library
Time: 42.850 ms

Radix sort by host
Time: 0.010 ms
INCORRECT :(

Radix sort by device
Hist Time: 4.892
Scan Time: 4.729
Add Time: 1.952
Transpose Time: 27.414
Scatter Time: 56.347
Time: 137.222 ms
CORRECT :)
```





Tổng kết

36

```
> nvcc Parallel_v7.cu 66 ./a.out 4 128 128
*****GPU info*****
Name: GeForce RTX 2070 with Max-Q Design
Compute capability: 7.5
Num SMs: 36
Max num threads per SM: 1024
Max num warps per SM: 32
GMEM: 8361279488 byte
SMEM per SM: 65536 byte
SMEM per block: 49152 byte
*****

Input size: 16777217

Num bits per digit: 4

Hist block size: 128, scan block size: 128

Radix sort by library
Time: 43.980 ms

Radix sort by host
Time: 0.010 ms
INCORRECT :(

Radix sort by device
Hist Time: 4.488
Scan Time: 2.350
Add Time: 1.093
Transpose Time: 5.576
Scatter Time: 52.510
Time: 104.522 ms
CORRECT :)
```





Tổng kết

37

```
> nvcc Parallel_v7.cu && ./a.out 4 128 256
*****GPU info*****
Name: GeForce RTX 2070 with Max-Q Design
Compute capability: 7.5
Num SMs: 36
Max num threads per SM: 1024
Max num warps per SM: 32
GMEM: 8361279488 byte
SMEM per SM: 65536 byte
SMEM per block: 49152 byte
*****

Input size: 16777217

Num bits per digit: 4

Hist block size: 128, scan block size: 256

Radix sort by library
Time: 44.944 ms

Radix sort by host
Time: 0.010 ms
INCORRECT :(

Radix sort by device
Hist Time: 4.493
Scan Time: 2.747
Add Time: 0.579
Transpose Time: 5.332
Scatter Time: 53.089
Time: 106.835 ms
CORRECT :)
```





Tổng kết

38

```
> nvcc Parallel_v7.cu 86 ./a.out 4 128 512
*****GPU info*****
Name: GeForce RTX 2070 with Max-Q Design
Compute capability: 7.5
Num SMs: 36
Max num threads per SM: 1024
Max num warps per SM: 32
GMEM: 8361279488 byte
SMEM per SM: 65536 byte
SMEM per block: 49152 byte
*****

Input size: 16777217

Num bits per digit: 4

Hist block size: 128, scan block size: 512

Radix sort by library
Time: 43.934 ms

Radix sort by host
Time: 0.010 ms
INCORRECT :(

Radix sort by device
Hist Time: 4.488
Scan Time: 2.504
Add Time: 0.612
Transpose Time: 5.332
Scatter Time: 52.952
Time: 104.607 ms
CORRECT :)
```





Tổng kết

39

```
> nvcc Parallel_v7.cu 86 ./a.out 4 256 128
*****GPU info*****
Name: GeForce RTX 2070 with Max-Q Design
Compute capability: 7.5
Num SMs: 36
Max num threads per SM: 1024
Max num warps per SM: 32
GMEM: 8361279488 byte
SMEM per SM: 65536 byte
SMEM per block: 49152 byte
*****

Input size: 16777217

Num bits per digit: 4

Hist block size: 256, scan block size: 128

Radix sort by library
Time: 44.227 ms

Radix sort by host
Time: 0.010 ms
INCORRECT :(

Radix sort by device
Hist Time: 4.270
Scan Time: 1.272
Add Time: 0.335
Transpose Time: 2.580
Scatter Time: 50.862
Time: 98.595 ms
CORRECT :)
```





Tổng kết

40

```
> nvcc Parallel_v7.cu 66 ./a.out 4 256 256
*****GPU info*****
Name: GeForce RTX 2070 with Max-Q Design
Compute capability: 7.5
Num SMs: 36
Max num threads per SM: 1024
Max num warps per SM: 32
GMEM: 8361279488 byte
SMEM per SM: 65536 byte
SMEM per block: 49152 byte
*****

Input size: 16777217

Num bits per digit: 4

Hist block size: 256, scan block size: 256

Radix sort by library
Time: 44.615 ms

Radix sort by host
Time: 0.010 ms
INCORRECT :(

Radix sort by device
Hist Time: 4.303
Scan Time: 1.212
Add Time: 0.348
Transpose Time: 2.855
Scatter Time: 55.095
Time: 106.838 ms
CORRECT :)
```





Tổng kết

41

```
> nvcc Parallel_v7.cu 86 ./a.out 4 512 128
*****GPU info*****
Name: GeForce RTX 2070 with Max-Q Design
Compute capability: 7.5
Num SMs: 36
Max num threads per SM: 1024
Max num warps per SM: 32
GMEM: 8361279488 byte
SMEM per SM: 65536 byte
SMEM per block: 49152 byte
*****

Input size: 16777217

Num bits per digit: 4

Hist block size: 512, scan block size: 128

Radix sort by library
Time: 43.205 ms

Radix sort by host
Time: 0.010 ms
INCORRECT :(

Radix sort by device
Hist Time: 4.459
Scan Time: 0.723
Add Time: 0.170
Transpose Time: 0.544
Scatter Time: 60.644
Time: 106.206 ms
CORRECT :)
```





Tổng kết

42

```
> nvcc Parallel_v7.cu 86 ./a.out 4 512 256
*****GPU info*****
Name: GeForce RTX 2070 with Max-Q Design
Compute capability: 7.5
Num SMs: 36
Max num threads per SM: 1024
Max num warps per SM: 32
GMEM: 8361279488 byte
SMEM per SM: 65536 byte
SMEM per block: 49152 byte
*****

Input size: 16777217

Num bits per digit: 4

Hist block size: 512, scan block size: 256

Radix sort by library
Time: 44.619 ms

Radix sort by host
Time: 0.014 ms
INCORRECT :(

Radix sort by device
Hist Time: 4.504
Scan Time: 0.713
Add Time: 0.179
Transpose Time: 0.535
Scatter Time: 59.453
Time: 104.468 ms
CORRECT :)
```

