GPU - Programming
Homework - 5

Topic - Reduce
Code -

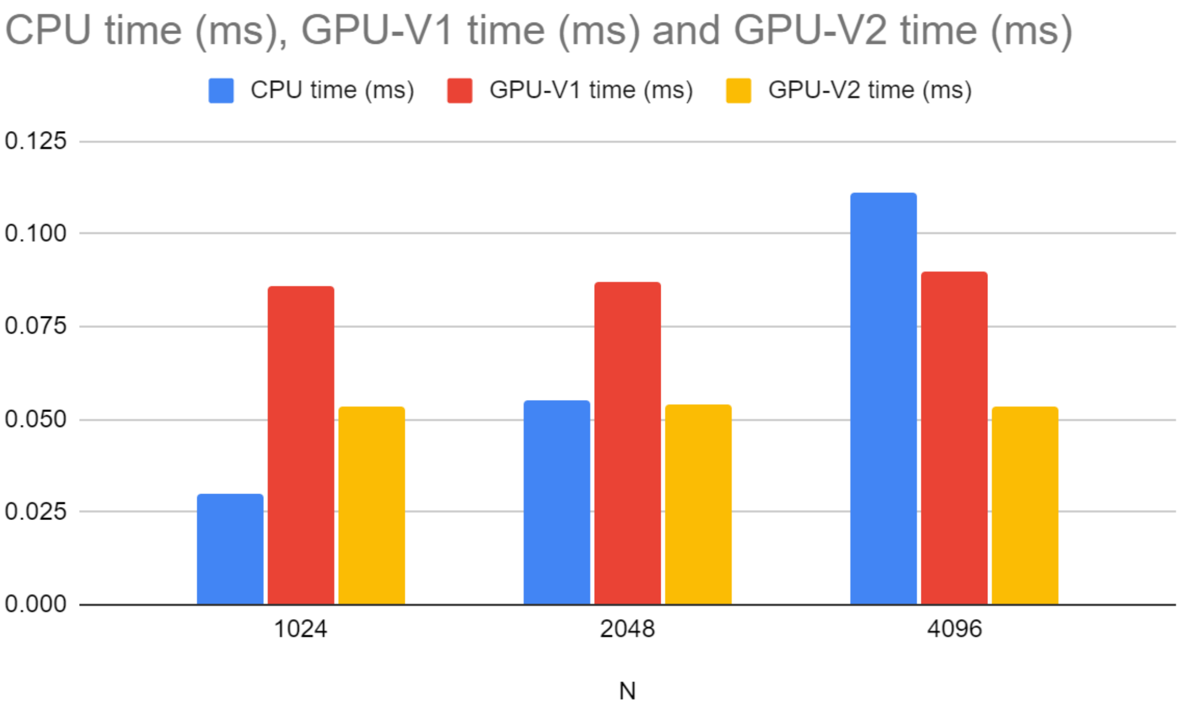
```
#include<stdio.h>

#define N 4096
#define BLOCK 1024

// i, n/2+i reduction
__global__ void reduceV1(int *elems){
	int id,i;
	id=threadIdx.x+blockIdx.x*blockDim.x;

	if(id>=N)
		return;

	for(i=N/2; i; i/=2) {
		if(id<i)
			elems[id] += elems[id+i];
		__syncthreads();
	}
	if(id==0)
		printf("GPU V1 Sum is %d\n",elems[0]);
}

// i,i+1 reduction
__global__ void reduceV2(int *elems)
{
	int tid = blockDim.x*blockIdx.x+threadIdx.x,threads =
blockDim.x*((N+BLOCK-1)/BLOCK),step=1,i1,i2;
	while(threads > 0)
	{
		if(tid < threads)
		{
			i1 = tid * step * 2;
			i2 = i1 + step;
			elems[i1] += elems[i2];
		}
		step = step<<1;
		threads = threads>>1;
```

```
        __syncthreads();
    }
    if(tid==0)
        printf("GPU V2 Sum is %d\n",elems[0]);
}

int main(){
    int host[N],i;
    long int sum=0;
    clock_t start,stop;

    printf("For N = %d\n",N);

    start = clock();
    for(i=0;i<N;i++){
        host[i]=rand()%20;
        sum+=host[i];
    }
    stop = clock();
    printf("CPU Sum is %d\n",sum);
    printf("CPU time taken is: %lf ms\n",((double)(stop-start)/CLOCKS_PER_SEC)*1e3);

    int *d_elems;
    float ms;
    cudaEvent_t s1,s2;
    cudaEventCreate(&s1);
    cudaEventCreate(&s2);
    cudaMalloc(&d_elems,N*sizeof(int));
    cudaMemcpy(d_elems,host,N*sizeof(int),cudaMemcpyHostToDevice);

    cudaEventRecord(s1);
    reduceV1<<<(N+BLOCK-1)/BLOCK,BLOCK>>>(d_elems);
    cudaEventRecord(s2);
    cudaEventSynchronize(s2);
    cudaEventElapsedTime(&ms,s1,s2);
    printf("GPU V1 time taken is: %lf ms\n",ms);
    //cudaDeviceSynchronize();

    cudaMemcpy(d_elems,host,N*sizeof(int),cudaMemcpyHostToDevice);
    cudaEventRecord(s1);
    reduceV2<<<(N+BLOCK-1)/BLOCK,BLOCK>>>(d_elems);
    cudaEventRecord(s2);
    cudaEventSynchronize(s2);
    cudaEventElapsedTime(&ms,s1,s2);
```

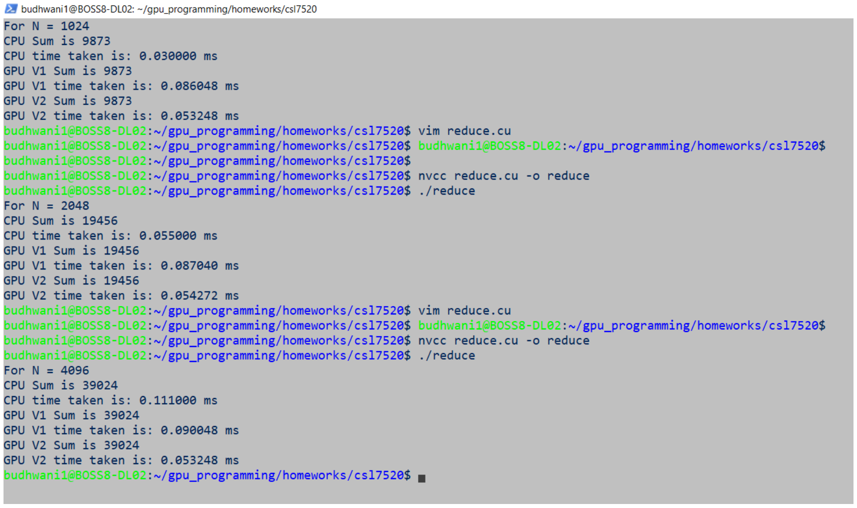
```
        printf("GPU V2 time taken is: %lf ms\n",ms);

        return 0;
}
```

Plots -

N	CPU time (ms)	GPU-V1 time (ms)	GPU-V2 time (ms)
1024	0.030000	0.086048	0.053248
2048	0.055000	0.087040	0.054272
4096	0.111000	0.090048	0.053248



Outputs -



```
For N = 1024
CPU Sum is 9873
CPU time taken is: 0.030000 ms
GPU V1 Sum is 9873
GPU V1 time taken is: 0.086048 ms
GPU V2 Sum is 9873
GPU V2 time taken is: 0.053248 ms
budhwani1@BOSS8-DL02:~/gpu_programming/homeworks/csl7520$ vim reduce.cu
budhwani1@BOSS8-DL02:~/gpu_programming/homeworks/csl7520$ budhwani1@BOSS8-DL02:~/gpu_programming/homeworks/csl7520$
budhwani1@BOSS8-DL02:~/gpu_programming/homeworks/csl7520$
budhwani1@BOSS8-DL02:~/gpu_programming/homeworks/csl7520$ nvcc reduce.cu -o reduce
budhwani1@BOSS8-DL02:~/gpu_programming/homeworks/csl7520$ ./reduce
For N = 2048
CPU Sum is 19456
CPU time taken is: 0.055000 ms
GPU V1 Sum is 19456
GPU V1 time taken is: 0.087040 ms
GPU V2 Sum is 19456
GPU V2 time taken is: 0.054272 ms
budhwani1@BOSS8-DL02:~/gpu_programming/homeworks/csl7520$ vim reduce.cu
budhwani1@BOSS8-DL02:~/gpu_programming/homeworks/csl7520$ budhwani1@BOSS8-DL02:~/gpu_programming/homeworks/csl7520$
budhwani1@BOSS8-DL02:~/gpu_programming/homeworks/csl7520$ nvcc reduce.cu -o reduce
budhwani1@BOSS8-DL02:~/gpu_programming/homeworks/csl7520$ ./reduce
For N = 4096
CPU Sum is 39024
CPU time taken is: 0.111000 ms
GPU V1 Sum is 39024
GPU V1 time taken is: 0.090048 ms
GPU V2 Sum is 39024
GPU V2 time taken is: 0.053248 ms
budhwani1@BOSS8-DL02:~/gpu_programming/homeworks/csl7520$
```