Date:-

Assignment No - A1

Title:- CUDA programming

Problem Statement:-

a) Implement Parallel Reduction using min, max, sum and average operations.

b) Write a CUDA program that gives an N-element vector, find the maximum, minimum element, as well as arithmetic mean and standard deviation.

Objectives:-

- To understand parallel reduction operations
- To understand vector operations.

Outcomes:-

Understood the parallel reduction operations as well as vector operations.

Software requirements: Open NP (C++ library), g++, Google Colab, CUDA.

Hardware Requirements: 8 GB RAM.

Theory:-

a) CUDA (Compute United Device Architecture) is a parallel computing platform and application

programming interface model created by NVIDIA. It allows software developers and engineers to use CUDA enabled graphics processing unit for general purpose processing, an approach termed GPU.
Languages C, C++, Fortran can be used with CUDA. This accessibility makes it easier for the specialist in parallel programming to use GPU resources in contact to prior API's like direct 3D and openGL which required advanced skills in graphic programming CUDA also supports programming frameworks.

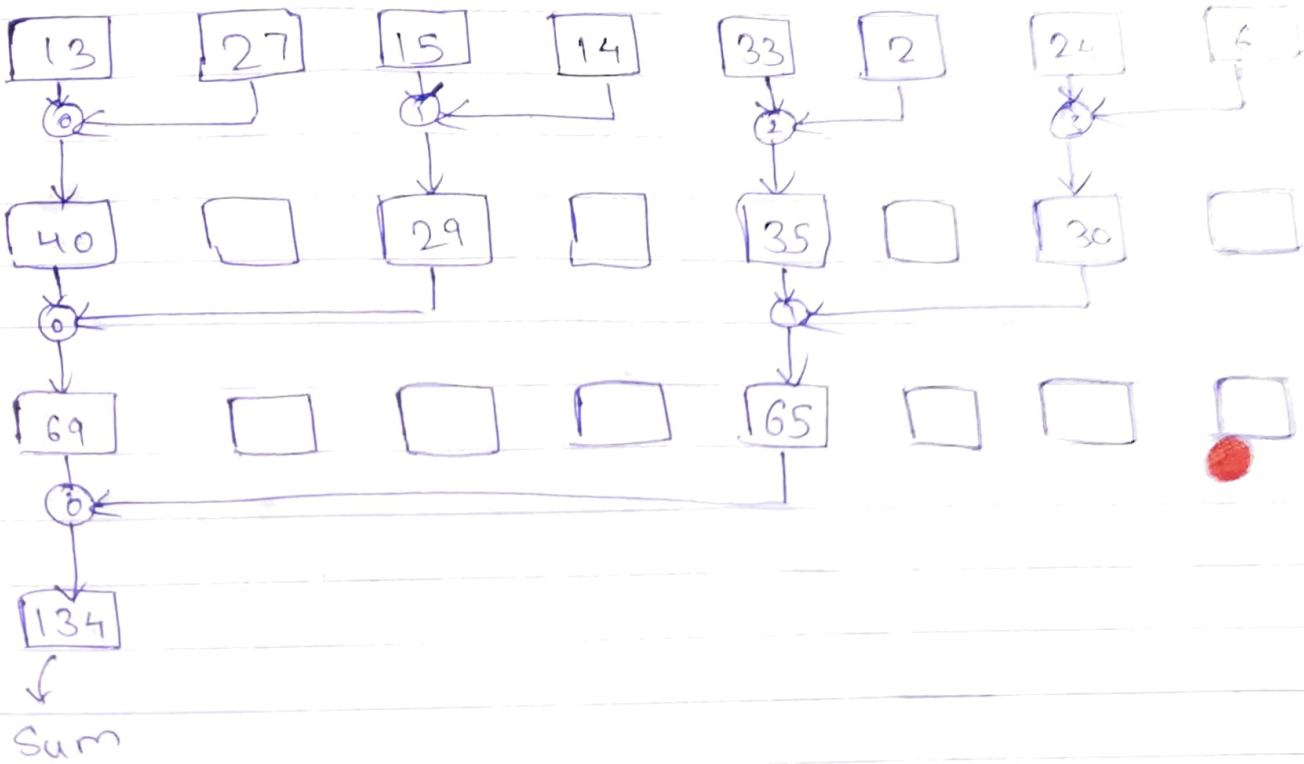
b) Min-Max operations:-
1) Max method: returns the larger element out of a, b compare function can be omitted By default, the compare functions is used to determine which object is larger in case they are non-numeric, otherwise the operator is used.
syntax:- max (object_type a, object_type b, compare ()).
2) Min Method: Returns smaller element of :xy out of a and b. Same rule applies for comparison as max function.
3) Arithmetic mean:- This value is found by taking the sum of all individual data

elements, and then dividing this sum by the total number of individual data elements. Mean is also referred to as 'average'.

4) Standard Deviation: represented by the greek letter σ (signal), standard deviation is a measure that is used to quantify the amount of variation or dispersion of a set of data values. This number is used to determine how measurements for a group of data are spread out from the average (mean) value, on either side.

5) Parallel Reduction:-

- Reduction operations are those which reduce a single collection of values to a single value
- Operations which are associate and commutative can be reduction operations.
- Some of them are: addition, manipulation, multiplication, bitwise AND, OR, XOR, logical AND, OR, XOR, finding minimum/maximum amongst a given set of numbers.
- Computation complexity likely to be $O(\log n)$
- Below is an example of sum of an array using parallel reduction.



Conclusion:-

I have studied parallel reduction using min, max, avg, sum; and CUDA program that given an N element array finds max, min, mean, standard deviation parallel and serially.

Both programs executed successfully and gave expected results.