NeuroMatrix® NM6403
neuroprocessor performance evaluation report

(preliminary information)

Abstract

This paper describes the evaluation results of TMS320C40 and NM6403 comparison on vector-matrix operations listed below. The goal was to assess the performance gain of NM6403 in neural network solutions in comparison with TMS320C40 that we used in acceleration boards before.

The comparison has been made on the base of the clocks number calculation for each test. A clock rate has not been taken into account.

All tests have been written by assembly language.

Main assumptions:

TMS:

1. Most instructions executes within 1 clock cycle (except call and jump instructions, that are delayed and execute within 4 clock cycles).
2. Memory access executes within 1 clock cycle.
3. Double memory access instructions execute within 1 clock cycle (access to on-chip memory).
4. Parallel memory access instructions execute within 1 clock cycle.

NM6403:

1. Memory access executes within 1 clock cycle.
2. Jump and call instructions are delayed and execute within 3 clock cycles.
3. Vector instructions that access different buses execute in parallel (Parallel instructions bit is set into *enable* position).

Data types and sizes:

- Input vector size is 4096 bytes, input matrix size is 1024*1024 bytes.
- Each byte of input data has 6 significant bites, so data range is from 0 to 63 for unsigned data or from -32 to 31 for signed data.
- Overflow is not detected while multiply-accumulation operations. If we take overflow into account it will reduce TMS320C40 performance. NM6403 includes overflow protection mechanism that can be turned on by using of internal activation function.
- Output data are byte arrays of appropriate size.

The result table

Test Name	**TMS** Cycles	**NM6403** Cycles	**Performance ratio (NM6403/TMS)**
Vectors sum	$2*10^3$	10^3	2.0
Vectors subtraction	$4*10^3$	10^3	4.0
Scalar - vector multiplication	10^3	$6*10^2$	1.7
Vector - vector multiplication by components	$1.6*10^4$	$1.6*10^4$	1.0
Outer vector - vector multiplication	$4.2*10^6$	$2.6*10^6$	1.6
Scalar vector - vector multiplication	$1.6*10^4$	$1.7*10^4$	0.9
Scalar vector - matrix multiplication	$7.9*10^5$	$1.4*10^5$	5.6
Scalar matrix - matrix multiplication	$8.1*10^8$	$1.8*10^7$	45.0
Matrix columns maximum search	$2.9*10^6$	$6.9*10^5$	4.2
Matrix rows maximum search	$2.7*10^6$	$7.0*10^5$	3.9
Whole matrix maximum search	$2.9*10^6$	$7.0*10^5$	4.1
Matrix transpose operation	$9.4*10^6$	$2.3*10^6$	4.1
Semi-linear function	$2.5*10^4$	$5.1*10^2$	48.0
Table function	$1.2*10^4$	$1.6*10^4$	0.75
Function Θ (x)	$5.1*10^3$	$5.3*10^2$	9.7
Matrix compression (ratio 2/1)	$2.3*10^5$	$1.4*10^5$	1.6
Function A(x) = A(x) - A(x)/4	$4.1*10^3$	$1.3*10^3$	3.2
Convolution function (mask size: m = 3?3):	$4.6*10^6$	$8.3*10^5$	5.5