



Processor NM6403

NeuroMatrix[®] Engine

Revision 1.1

This page was intentionally left blank.

The NeuroMatrix® architecture provides the unique flexibility of choice of the desired level of performance and precision for multiply/accumulation procedure. According to application requirements, the necessary length of sliced operands and products can be packed into a 64-bit data word. The number of multiplications/accumulations will depend on the length and number of operands. The highest performance - 14.400 MMACs (million multiplication and accumulations) is achievable with one-bit length operands and 50MHz clock rate. It is possible to increase the precision of calculation by using any operand length up to 32-bit. In this case, the performance will be 50 MMACs (million arithmetic multiplication and accumulations). That provides a trade-off between precision and speed. An example of usage of NeuroMatrix engine is shown in Fig. 1.

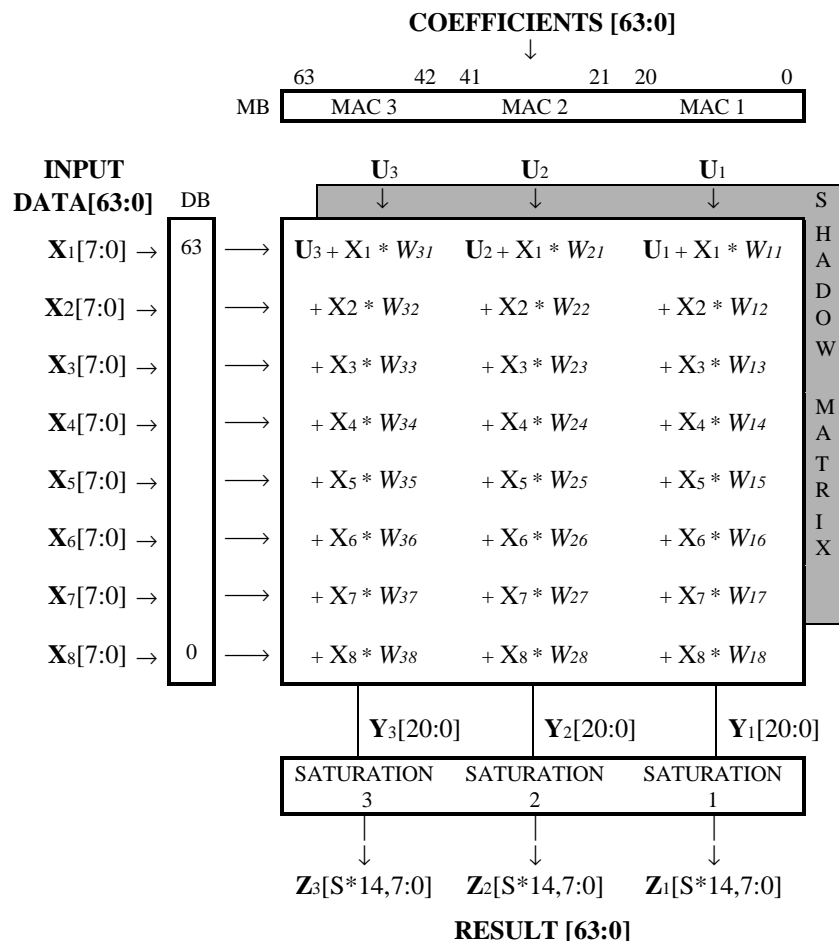


Fig. 1

The engine's core is an active matrix which looks like an array multiplier. The structure comprises cells that include a 1-bit memory (flip-flop) surrounded by several logical elements. The cells can be combined into several macrocells by using two 64-bit programmable registers: DB (data boundary) and MB (MAC boundary). These registers define the borders between rows and columns with macrocells. Each macrocell performs the multiplication on variable input words using preloaded coefficients (W_i) and accumulates the result from the macrocells in the column above it. The columns simultaneously calculate the results in one processor cycle. For 8-bit data (X_i) and coefficients, the engine performs 24 multiplication/accumulations with 21-bit results in one 20-nsec processor cycle. In this mode the performance is 1.200 MMACs (million arithmetic multiplications/accumulations per second). The number of multiplications/accumulations depends on the length and number of words packaged into a 64-bit block. The engine's configuration can change dynamically during the calculations. You can start the application with maximum precision and minimum performance and then dynamically increase the performance by reducing the data-word lengths.

To load new coefficients to the engine, thirty two clock cycles is needed. To avoid the delay appearing when the coefficients refresh, the shadow matrix is used. The new coefficients are loaded to the shadow matrix in a background mode and then re-written to the active matrix in one clock cycle.

To avoid arithmetic overflow, the engine uses the saturation functions with user-programmable saturation boundaries. The saturation functions reduce the number of significant bits and pack MAC products to 64-bit data word.

The following equations could be used for calculation of the necessary peak performance:

$$MMAC = \left[\frac{64}{N_x} \right] * \left[\frac{64}{N_x + N_w + \left\lceil \log_2 \frac{64}{N_x} \right\rceil} \right] * F ,$$

where: MMAC - million multiplications/accumulations per second

64 - NeuroMatrix® engine data word length;

N_x - input data word length;

N_w - coefficients word length;

$F = 50$ MHz - clock rate.

In case of $N_X \neq 1$ and $N_W = 1$, the equation is:

$$MMAC = \left\lfloor \frac{64}{N_X} \right\rfloor * \left\lceil \frac{64}{N_X + \left\lceil \log_2 \frac{64}{N_X} \right\rceil} \right\rceil * F .$$

In case of $N_X = 1$ and $N_W \neq 1$, the equation is:

$$MMAC = 32 * \left\lfloor \frac{64}{N_Y + 5} \right\rfloor * F .$$

In case of $N_X = N_W = 1$, the equation is:

$$MMAC = 1024 * F$$

And finally, in case of $N_X = N_W = 32$, the equation is:

$$MMAC = F$$

Some results of the performance calculation are shown in Fig. 2.

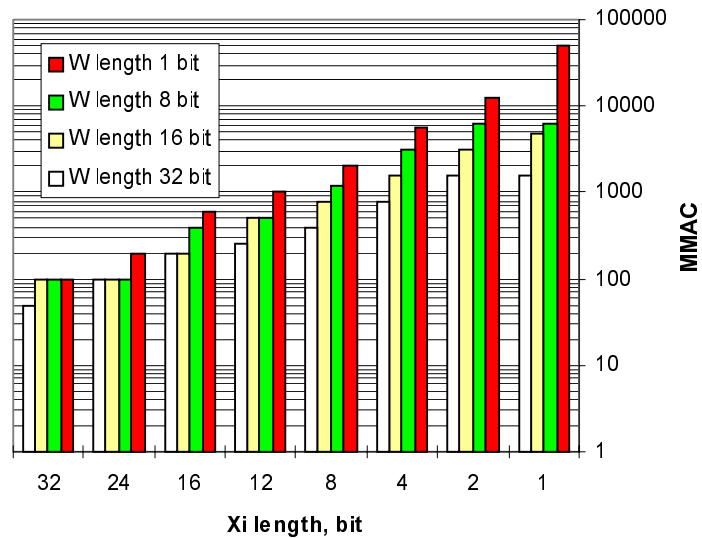


Fig. 2

The highest precision (32-bit coefficients and 32-bit input data) provides the lowest performance - 50 MMACs. The lowest precision (1-bit coefficients and 1-bit input data) provides the highest performance - 14.400 MMACs. When of using the "Boolean multuplications", the performance is 50.000+ MOPS.

Because of its flexibility, the engine can be used as a core for 3-rd generation DSP in the fields of signal processing, wireless communications, image processing, neural network accelerators and vector-matrix calculations.

This page was intentionally left blank.



**JOINT-STOCK COMPANY
RESEARCH CENTRE**

Research Centre Module
Box: 166, Moscow, 125190, Russia
Tel: +7 (095) 152-9335
Fax: +7 (095) 152-3168
E-Mail: info@module.ru
WWW: <http://www.module.ru>

©RC Module, 1999

All rights reserved.

Neither the whole nor any part of the information contained in, or the product described in this overview may be adapted or reproduced in any form except with the prior written permission of the copyright holder.

RC Module reserves the right to make changes without further notice to product herein to improve reliability, function or design. RC Module shall not be liable for any loss or damage arising from the use of any information in this overview or any error or omission in such information, or any incorrect use of the product.

Printed in Russia Data of release: 1999 September 15