RH850/U2Bx

Fully Connected Neural Network

Summary

This application describes the operation method of the neural network using the Floating-Point Unit (FPU) and the Extended Floating-Point Unit (FXU) that are supported by the RH850/U2Bx.

This application does not include the specification details information of FXU. Please refer to the APN “FXU Use for FP-SIMD Calculations” for the details. Also, please check the product specifications before using since the presence or absence of FXU and the position of the CPU equipped with FXU differ depending on the product. Refer to the appendix for the details.

Although the operation of the fully connected neural network example described in this application note has been confirmed, but please sure to confirm the operation before using it.

Operation Checked Device

RH850/U2Bx

Contents

1. Fully Connected Neural Network Overview

1.1 Fully Connected Neural Network Construction

Figure 1-1 shows the construction diagram of the fully connected neural network in this application. It is constructed by an input layer, two middle layers, and an output layer.

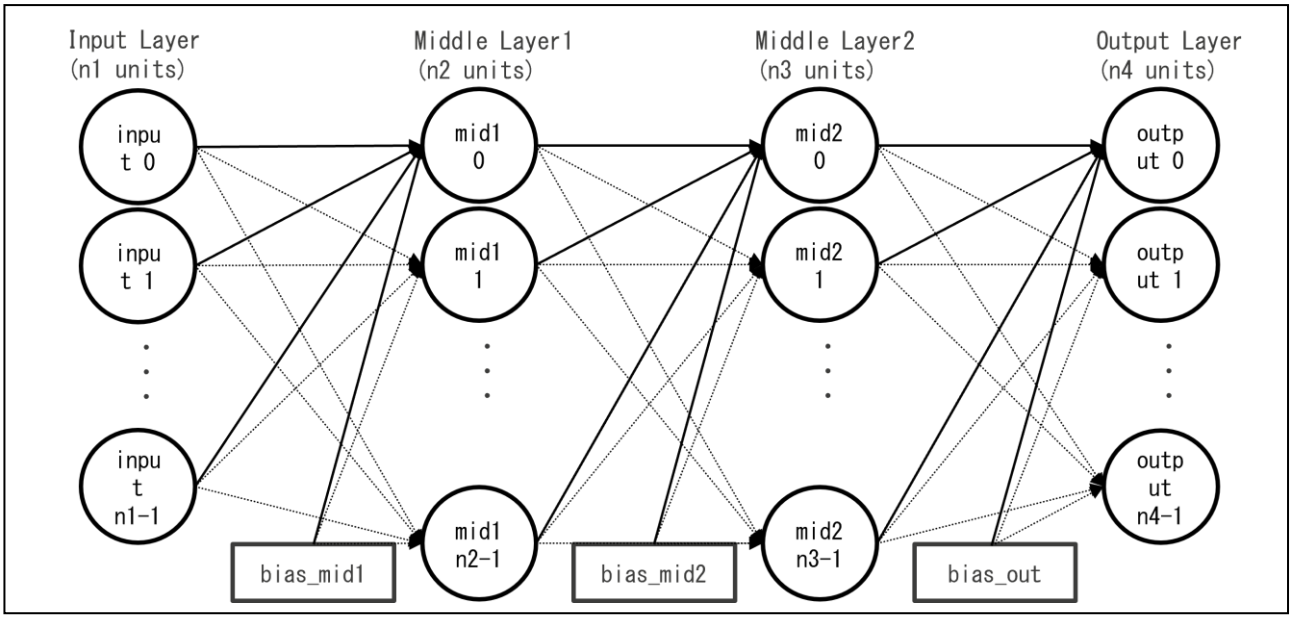


Figure 1-1 Fully Connected Neural Network Construction Diagram

1.2 Operation Method

Perform the fully connected processing and the activation function processing by each layer. Figure 1-2 shows the formula of the input layer to the middle layer.

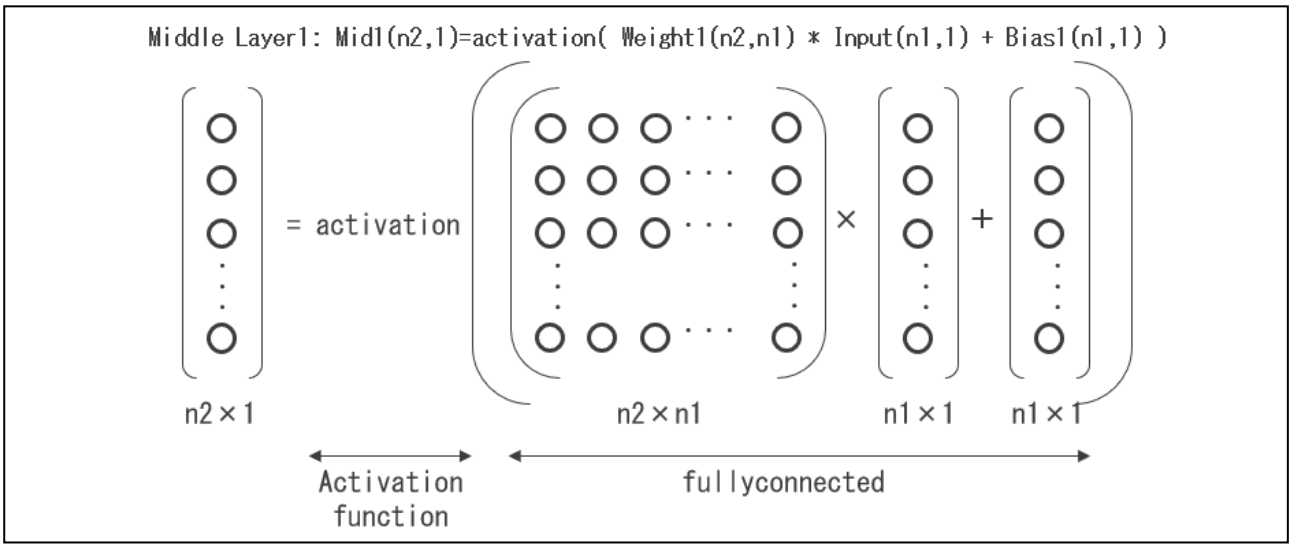


Figure 1-2 Each Layer Formula

1.3 Support Function

This sample software supports the following functions

Table 1-1 Support Function List

Function		FPU	FXU
Fully connected		fullyconnected	fullyconnected_fxu
Activation function	tanh	act_tanh	act_tanh_fxu
		act_tanh_usingexp	act_tanh_usingexp_fxu
	sigmoid	act_sigmoid	act_sigmoid_fxu
	ReLU	act_relu	act_relu_fxu

1.4 Use Hardware Function

The hardware functions of RH850/U2Bx using in this sample software are shown below.

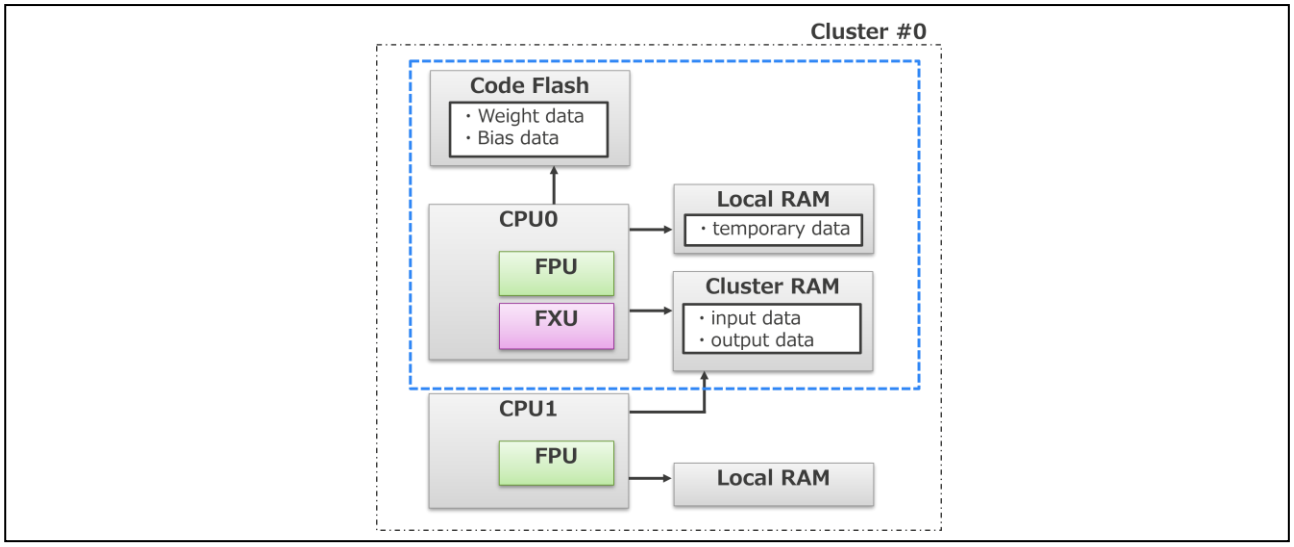
- Floating-Point Unit (FPU)
- Extended Floating-Point Unit (FXU)
- Various Memories (Code Flash、Cluster RAM、Local RAM)

This sample software performs the processing by inside of a cluster (Cluster #0) using CPU0. Refer to “2.4 Allocation of Constant and Variable” for the details of the constant and variable data allocation.

Although performs the function by the single precision in this sample soft.

This sample software supports single precision (32-bit).

Figure 1-3 System Configuration



2. Software Explanation

2.1 Operation Flow

Figure 2-1 shows the operation flow in this sample software. The following operation flow is the example using tanh function for the activation function. This sample software supports the sigmoid function and the ReLU function, so replace them if necessary.

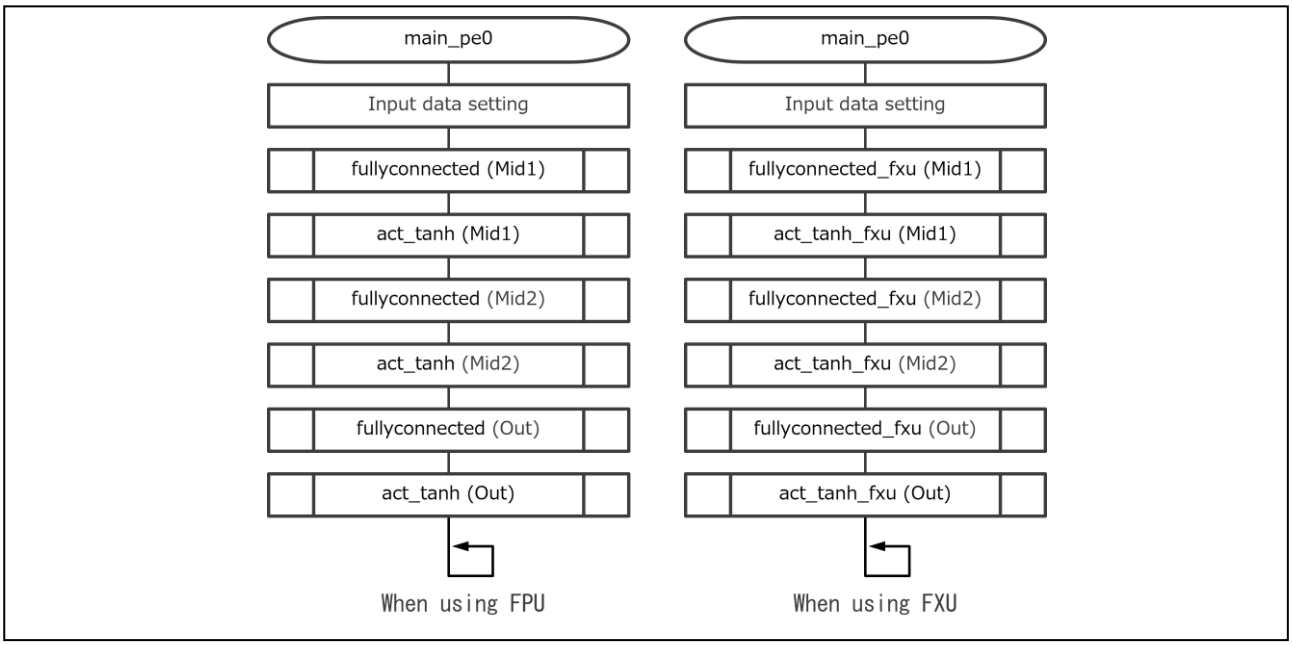


Figure 2-1 Operation Flow

2.2 Sample Software Configuration

Table 2-1 shows the file configuration of sample software.

Table 2-1 Sample Software File Configuration

File Name					Overview
FNN	FPU	U2Bx_Sample.gpj			Master project file
		U2B10_Sample.gpj			Master project file
		src	U2B10_Sample.gpj		Project file
			core0	fnn_fpu.c	Fully connected processing function, activation function
				weight_data_fpu(a/b/c).c	File of each pattern Refer to ”2.5 Change of The Number of Units
				weight_data_fpu(a/b/c).h	Header file of each pattern Refer to ”2.5 Change of The Number of Units”
				sub_timer_benchmark.c	File for processing load measurement
				sub_timer_benchmark.h	Header file for processing load measurement
				main_pe0.c	main function for CPU0
				intprg.c	Interrupt processing function No particular processing content
			core1	main_pe0.c	main function for CPU1
				intprg.c	Interrupt processing function No particular processing content
			core2	main_pe0.c	main function for CPU2
				intprg.c	Interrupt processing function No particular processing content
			core3	main_pe0.c	main function for CPU3
				intprg.c	Interrupt processing function No particular processing content
		startup			Start-up routine
	FXU	U2Bx_Sample.gpj			Master project file
		U2B10_Sample.gpj			Master project file
		src	U2B10_Sample.gpj		Project file
			core0	fnn_fxu.c	Fully connected processing function, activation function
				weight_data_fxu(a/b/c).c	File of each pattern

					Refer to ”2.5 Change of The Number of Units
				weight_data_fxu(a/b/c).h	Header file of each pattern Refer to ”2.5 Change of The Number of Units”
				sub_timer_benchmark.c	File for processing load measurement
				sub_timer_benchmark.h	Header file for processing load measurement
				main_pe0.c	main function for CPU0
				intprg.c	Interrupt processing function No particular processing content
			core1	main_pe0.c	main function for CPU1
				intprg.c	Interrupt processing function No particular processing content
			core2	main_pe0.c	main function for CPU2
				intprg.c	Interrupt processing function No particular processing content
			core3	main_pe0.c	main function for CPU3
				intprg.c	Interrupt processing function No particular processing content
		startup			Start-up routine

2.3 Function Specification

2.3.1 Function for FPU

Table 2-2 shows the functions list for FPU in this operation example

Table 2-2 Functions List for FPU

Function Name	Overview
main_pe0	Performs the call of each function.
fullyconnected	Performs the fully connected processing.
act_tanh	Performs the activation function processing (tanh).
act_tanh_usingexp	Performs the activation function processing (tanh). According to the following conversion formula, executes the tanh function processing to use the exponential function and four arithmetic operations. $\tanh x = \frac{e^x - e^{-x}}{e^x + e^{-x}}$
act_sigmoid	Performs the activation function processing (sigmoid).
act_relu	Performs the activation function processing (ReLU).

Table 2-3 to Table 2.7 show the functions operations for FPU using in this operation example.

Table 2-3 Specification of fullyconnected Function

fullyconnected			
Overview	Performs the fully connected processing and stores the result to the specified array.		
Declaration	`void fullyconnected(float input[], const float weight[], const float bias[], float output[], unsigned int size_in, unsigned int size_out);`		
Argument	[IN]	float input[]	: Specifies the input data of the fully connected processing.
	[IN]	float weight[]	: Specifies the weight matrix data of fully connected processing.
	[IN]	float bias[]	: Specifies the bias data of the fully connected processing.
	[OUT]	float output[]	: Stores the result of the fully connected processing.
	[IN]	unsigned int size_in	: Specifies the input data size.
	[IN]	unsigned int size_out	: Specifies the output data size.
Return value	-		
Remarks	- Allocate the weight matrix data specified in the argument in the transposed state. (Refer to “3.3 Constant Data Placement to Code Flash”)		

Table 2-4 Specification of act_tanh Function

act_tanh			
Overview	Performs the activation function processing (tanh) and stores the result to the specified array.		
Declaration	`void act_tanh(float input[], float output[], unsigned int size_in);`		
Argument	[IN]	float input[]	: Specifies the input data of the activation function processing (tanh).
	[OUT]	float output[]	: Stores the result of the activation function processing (tanh).
	[IN]	unsigned int size_in	: Specifies the input data size.
Return value	-		
Remarks			

Table 2-5 Specification of act_tanh_usingexp Function

act_tanh_usingexp			
Overview	Performs the activation function processing (tanh) and stores the result to the specified array. Executes the tanh function processing to use the exponential function and four arithmetic operations.		
Declaration	`void act_tanh_usingexp(float input[], float output[], unsigned int size_in);`		
Argument	[IN]	float input[]	: Specifies the input data of the activation function processing (tanh).
	[OUT]	float output[]	: Stores the result of the activation function processing (tanh).
	[IN]	unsigned int size_in	: Specifies the input data size.
Return value	-		
Remarks	- According to the following, calculates tanh using the formula with the exponential function. $\tanh x = \frac{e^{x} - e^{-x}}{e^{x} + e^{-x}}$ - Please note the input range since this function uses the expf function and the output is e^{n} for the input n.		

Table 2-6 Specification of act_sigmoid

act_sigmoid			
Overview	Performs the activation function processing (sigmoid) and stores the result to the specified array.		
Declaration	`void act_sigmoid(float input[], float output[], unsigned int size_in);`		
Argument	[IN]	float input[]	: Specifies the input data of the activation function processing (sigmoid).
	[OUT]	float output[]	: Stores the result of the activation function processing (sigmoid).
	[IN]	unsigned int size_in	: Specifies the input data size.
Return value	-		
Remarks	- Please note the input range since this function uses the expf function and the output is e^n for the input n.		

Table 2-7 Specification of act_relu

act_relu			
Overview	Performs the activation function processing (ReLU) and stores the result to the specified array.		
Declaration	`void act_relu(float input[], float output[], unsigned int size_in);`		
Argument	[IN]	float input[]	: Specifies the input data of the activation function processing (ReLU).
	[OUT]	float output[]	: Stores the result of the activation function processing (ReLU).
	[IN]	unsigned int size_in	: Specifies the input data size.
Return value	-		
Remarks			

2.3.2 FXU Versions Function

Table 2-8 shows the functions list of the FXU versions using in this operation.

In this sample software, the built-in functions of FXU instructions standardly supported in GHS. Refer to “3.4.1.2 Details of FXU Built-in Function” for the built-in function details.

Table 2-8 Function List of FXU Ver.

Function Name	Overview
main_pe0	Performs the call of each function.
fullyconnected_fxu	Performs the fully connected processing.
act_tanh_fxu	Performs the activation function processing (tanh).
act_tanh_usingexp_fxu	Performs the activation function processing (tanh). According to the following conversion formula, executes the tanh function processing to use the exponential function and four arithmetic operations. $\tanh x = \frac{e^{x} - e^{-x}}{e^{x} + e^{-x}}$
act_sigmoid_fxu	Performs the activation function processing (sigmoid).
act_relu_fxu	Performs the activation function processing (ReLU).
tanhf_vector	Performs the tanhf function processing for each vector element.
expf_vector	Performs the expf function processing for each vector element.

Table 2-9 to Table 2-15 show the functions operations for FXU version using in this operation example.

Table 2-9 Specification of fullyconnected_fxu Function

fullyconnected_fxu			
Overview	Performs the fully connected processing using FXU, and stores the result to the specified array.		
Declaration	`void fullyconnected_fxu(float input[], const float weight[], const float bias[], float output[], unsigned int size_in, unsigned int size_out);`		
Argument	[IN]	float input[]	: Specifies the input data of the fully connected processing.
	[IN]	float weight[]	Specifies the weight matrix data of fully connected processing.
	[IN]	float bias[]	: Specifies the bias data of the fully connected processing.
	[OUT]	float output[]	: Stores the result of the fully connected processing.
	[IN]	uinsigned int size_in	: Specifies the input data size.
	[IN]	unsigned int size_out	: Specifies the output data size.
Return value	-		
Remarks	- Allocate the weight matrix data specified in the argument in the transposed state. (Refer to “3.3 Constant Data Placement to Code Flash”) - If the column size of the weight matrix data and the size of the bias data specified in the argument are not multiples of four, zero pad them until the size is a multiple of four. At the same time, make the argument size_out a multiple of four. (Refer to “3.4.2 Data Size”.) - Allocate the start address of the specified data of the argument: input[], weight[], bias[], and output[] to the 16Byte boundary. (Refer to “3.4.3 Alignment Specification”.)		

Table 2-10 Specification of act_tanh_fxu Function

act_tanh_fxu			
Overview	Performs the activation function processing (tanh) using FXU and stores the result to the specified array.		
Declaration	`void act_tanh_fxu(float input[], float output[], unsigned int size_in);`		
Argument	[IN]	float input[]	: Specifies the input data of the activation function processing (tanh).
	[OUT]	float output[]	: Stores the result of the activation function processing (tanh).
	[IN]	unsigned int size_in	: Specifies the input data size.
Return value	-		
Remarks	- Allocate the start address of the specified data of the argument: input[] and output[] to the 16Byte boundary. (Refer to “3.4.3 Alignment Specification”.)		

Table 2-11 Specification of act_tanh_usingexp_fxu Function

act_tanh_usingexp_fxu	
Overview	Performs the activation function processing (tanh) using FXU, and stores the result to the specified array. Executes the tanh function processing to use the exponential function and four arithmetic operations.
Deceleration	`void act_tanh_usingexp_fxu(float input[], float output[], unsigned int size_in);`
Argument	[IN] float input[] : Specifies the input data of the activation function processing (tanh). [OUT] float output[] : Stores the result of the activation function processing (tanh). [IN] unsigned int size_in : Specifies the input data size.
Return value	-
Remarks	- According to the following, calculates tanh using the formula with the exponential function. $\tanh x = \frac{e^{x} - e^{-x}}{e^{x} + e^{-x}}$ - Please note the input range since this function uses the expf function and the output is e^{n} for the input n. - Allocate the start address of the specified data of the argument: input[] and output[] to the 16Byte boundary. - (Refer to “3.4.3 Alignment Specification”.)

Table 2-12 Specification of act_sigmoid_fxu Function

act_sigmoid_fxu	
Overview	Performs the activation function processing (sigmoid) using FXU and stores the result to the specified array.
Declaration	`void act_sigmoid_fxu(float input[], float output[], unsigned int size_in);`
Argument	[IN] float input[] : Specifies the input data of the activation function processing (sigmoid). [OUT] float output[] : Stores the result of the activation function processing (sigmoid). [IN] unsigned int size_in : Specifies the input data size.
Return value	-
Remarks	- Please note the input range since this function uses the expf function and the output is e^{n} for the input n. - Allocate the start address of the specified data of the argument: input[] and output[] to the 16Byte boundary. (Refer to “3.4.3 Alignment Specification”.)

Table 2-13 Specification of act_relu_fxu Function

act_relu_fxu			
Overview	Performs the activation function processing (ReLU) using FXU and stores the result to the specified array.		
Declaration	`void act_relu_fxu(float input[], float output[], unsigned int size_in);`		
Argument	[IN]	float input[]	: Specifies the input data of the activation function processing (ReLU).
	[OUT]	float output[]	: Stores the result of the activation function processing (ReLU).
	[IN]	unsigned int size_in	: Specifies the input data size.
Return value	-		
Remarks	- Allocate the start address of the specified data of the argument: input[] and output[] to the 16Byte boundary. (Refer to “3.4.3 Alignment Specification”.)		

Table 2-14 Specification of tanhf_vector Function

tanhf_vector			
Overview	Performs floating-point tanh function calculation. Executes processing on each of the four elements of the vector specified in the argument and stores the result in the vector.		
Declaration	`__ev128_f32__ tanhf_vector(__ev128_f32__ x);`		
Argument	[IN]	__ev128_f32__ x	: Specifies the vector that performs the tanh function calculation.
Return value	Value of __ev128_f32__ type		: Return the vector that is stored the calculation result of tanh function.
Remarks			

Table 2-15 Specification of expf_vector Function

expf_vector			
Overview	Performs floating-point exponential calculation. Executes processing on each of the four elements of the vector specified in the argument and stores the result in the vector.		
Declaration	`__ev128_f32__ expf_vector(__ev128_f32__ x);`		
Argument	[IN]	__ev128_f32__ x	: Specifies the vector that performs the exponential calculation.
Return value	Value of __ev128_f32__ type		: Return the vector that is stored the calculation result of tanh function.
Remarks			

2.4 Allocation of Constant and Variable

In this sample software, performs the processing in a cluster#0 using CPU0. As shown in Figure 2-2 and Table 2-16, allocate the input/output data to Cluster RAM, and the constant (weight matrix data, bias data) to Code Flash.

Please note that there is a possibility the processing performance is degreased caused by the data access delaying if the data is not allocated properly to the resource corresponding to the CPU used.

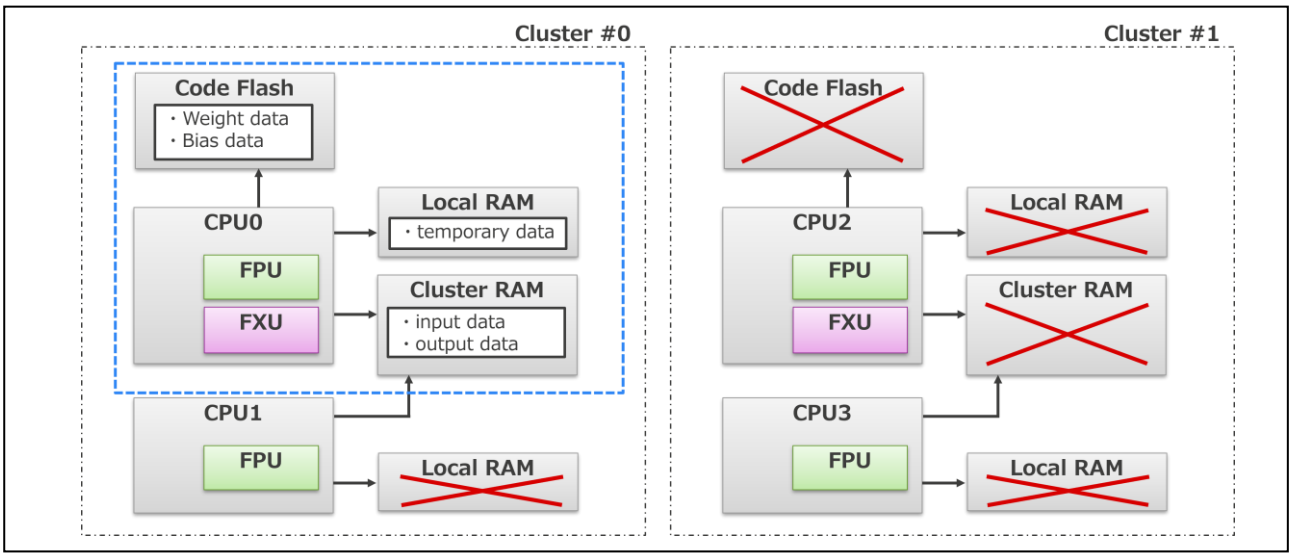


Figure 2-2 Allocation of Constant and Variable

Table 2-16 Type and Allocation of Constant and Variable

Type		Data Name	Constant/Variable	Allocation
Input/output data		input_data	Global variable	Cluster RAM in the same cluster as the CPU used
		output_data		
Weight data	Intermediate layer	weight_mid1	Constant	Code Flash in the same cluster as the CPU used
	Intermediate layer 2	weight_mid2		
	Output layer	weight_out		
Bias data	Intermediate layer 1	bias_mid1		
	Intermediate layer 2	bias_mid2		
	Output layer	bias_out		
Intermediate data	Intermediate layer 1	mid1_tmp	Local variable	Local RAM
		mid1_out		
	Intermediate layer 2	mid2_tmp		
		mid2_out		
	Output layer	out_tmp		

2.5 Change of The Number of Units

This sample software is selectable from the 3 patterns of data sets for both FPU/FXU.

Table 2-17 Pattern of The Number of Unit

Pattern		Number of Unit				File	
		Input layer	Middle layer1	Middle layer2	Output layer		
FPU	A	10	10	10	10	weight_data_fpua.h	weight_data_fpua.c
	B	10	20	20	10	weight_data_fpub.h	weight_data_fpub.c
	C	10	30	30	10	weight_data_fpuc.h	weight_data_fpuc.c
FXU*1	A	10	10	10	10	weight_data_fxua.h	weight_data_fxua.c
	B	10	20	20	10	weight_data_fxub.h	weight_data_fxub.c
	C	10	30	30	10	weight_data_fxuc.h	weight_data_fxuc.c

【Note1】The column size of the weight matrix data and the bias data size for each layer must be the multiple of four since the FXU processes four elements at a time. In that case, extend them by filling the data element with zeros and change the macro constant that indicate the data size. Refer to “3.4.2 Data Size” for the details.

When change the pattern, specify the header file and the constant data file corresponding the pattern.

(a) Change the definition of the macro name in main_pe0.c. This will change the header file to read.

Ex.) #define PATTERN_A when setting to the pattern A.

(b) Specifies the constant data file in U2B10_Sample_src.gpj.

Ex.) .¥weight_data_fpua.c when setting to the pattern A (FPU).

3. Precautions and Restrictions

3.1 FPU/FXU Initial Setting

The PSW register setting is required when using FPU/FXU. Also, the option byte setting is required when using FXU. Refer to each product user's manual for the detailed setting method.

Also, please check the product specifications since the presence or absence of FXU and the position of the CPU equipped with FXU differ depending on the product. Refer to the appendix for the details.

PSW Register

FPU：Enabled by setting the bit 16 (CU0) of the program status word (PSW) of the CPU to "1".

FXU：Enabled by setting the bit 17 (CU1) of the program status word (PSW) of the CPU to "1".

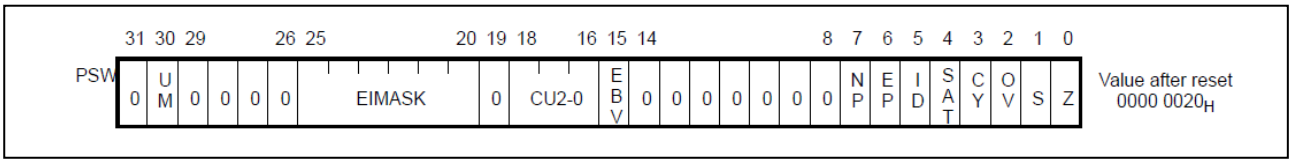


Figure 3-1 PSW Register

Option Byte

FXU mounting CPU0：Enabled by setting the bit 16 (PE0_FPSIMD_EN) of the OPBT3 to "1".

FXU mounting CPU2：Enabled by setting the bit 18 (PE2_FPSIMD_EN) of the OPBT3 to "1".

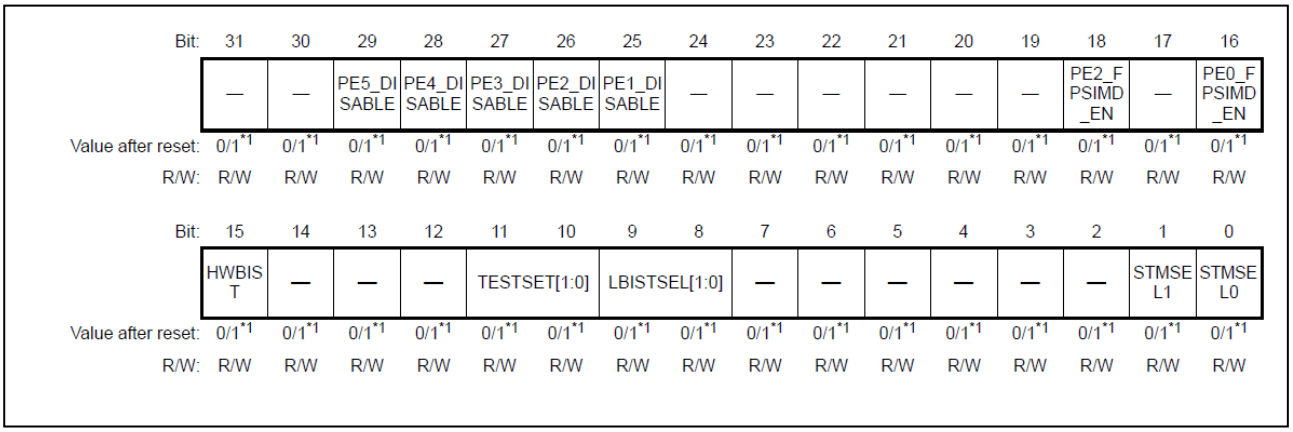


Figure 3-2 Option Byte Setting

3.2 Upper Limit and Low Limit of Single-Precision Floating-Point Type

The range of values that a single-precision floating-point type can represent is limited. Especially, when using an exponential function, it is necessary to note the input range because the output is e^n for the input n. In this sample software, the following function uses an exponential function.

act_tanh_usingexp, act_tanh_usingexp_fxu, act_sigmoid, act_sigmoid_fxu

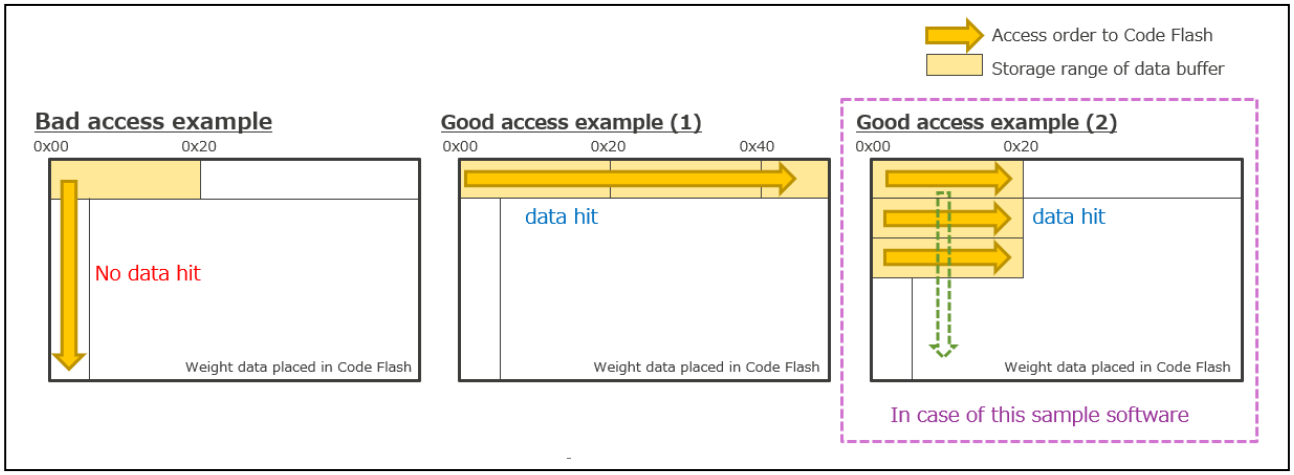
3.3 Constant Data Placement to Code Flash

This section describes how to read the constant data from Code Flash in this sample software and how to allocate the constant data in Code Flash.

3.3.1 Effective Use of Data Buffer

Describes the optimization method when the data reading of the Code Flash. When reading the data allocated in Code Flash, if the data is not placed continuously in the memory, the data hit get bad, and it takes long time to read the data, resulting in lower processing performance. Therefore, this sample software performs the data reading by the configuration as shown in Figure 3-3. Thereby, it is possible to read data while making effective use of the data buffer. The next section, “3.3.2 Transpose of Weight Matrix Data” describes the details.

Figure 3-3 Access Order Optimization to Code Flash



3.3.2 Transpose of Weight Matrix Data

As shown in Figure 3-4, allocate the weight matrix data with the rows and columns transposed.

The FXU instruction processes four elements at a time. Therefore, when calculating the product of the matrix and the vector in the general data processing direction (matrix column direction), it is necessary to sum the four elements of vector register in order to calculate one element of the output value. In this sample software, as shown in Figure 3-4, transposes the rows and columns of the matrix data, and the product sum of multiple output values is performed in the parallel. Thereby, it is no longer necessary to sum the four elements of the vector register, and faster processing speed can be expected.

In this sample software, the data processing is performed with the same configuration even in the case of FPU. Therefore, transpose and allocate the weight matrix data regardless of whether you use FPU or FXU,

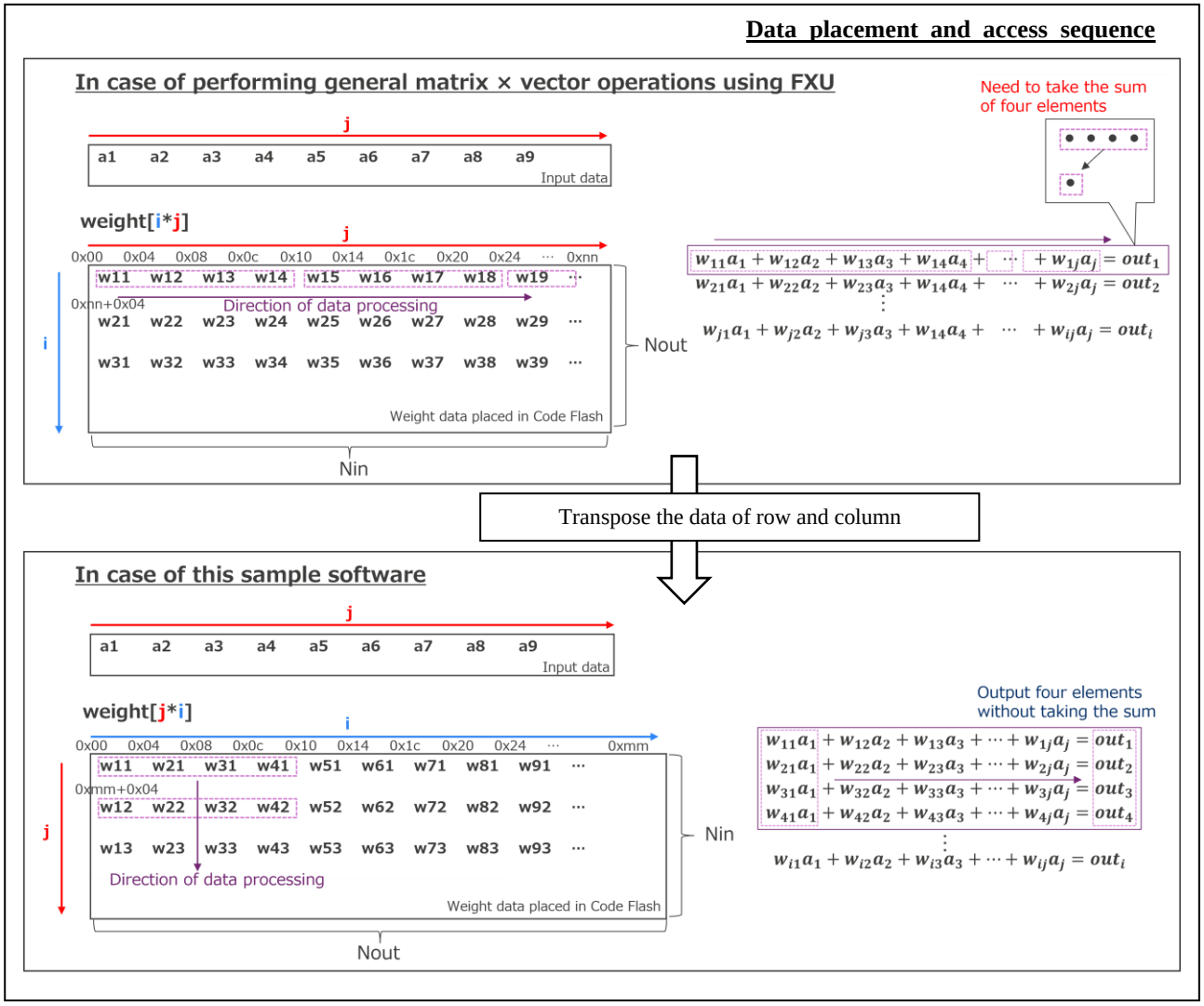


Figure 3-4 Placement Optimization of Weight Matrix Data

【Note】The above diagram is for illustration purposes. Actually, loop unrolling is performed to speed up the processing.

3.4 Notes on FXU use

3.4.1 FXU Built-in Functions

3.4.1.1 Setting when Compiling

To use the FXU built-in functions, it is necessary to enable the FXU support and include the header file v800_fxu.h or the header file v800_ghs.h. The latter automatically includes the former when the FXU support is enabled.

- FXU Support Enabling：-rh850_fxu
- Header File Including：#include <v800_fxu.h> or <v800_ghs.h>

3.4.1.2 Details of FXU Built-in Function

Table 3-1 shows the built-in function of FXU using in this sample software.

Table 3-1 FXU Built-in Function List

Bilt-in Function Name	Overview
__ev128_ldvqw	Loads the quad word to the vector register.
__ev128_ldvw_mask	Loads the word to the specified element of the vector register.
__ev128_stvqw	Stores the quad word of the vector register to the specified address.
__ev128_addfs_4	Performs single-precision floating-point addition for each element of the vector register.
__ev128_subfs_4	Performs single-precision floating-point subtraction for each element of the vector register.
__ev128_divfs_4	Performs single-precision floating-point division for each element of the vector register.
__ev128_fmafs_4	Performs single-precision floating-point fused multiply-addition for each element of the vector register.
__ev128_get_f32	Extracts the element of the specified vector register.

In the FXU built-in function, vector data type "__ev128_f32__" is used. It represents the vector with four 32-bit single-precision floating-point elements.

Table 3-2 to 3-9 show the specification of FXU built-in function using in this operation example.

Table 3-2 Specification of __ev128_ldvqw Function

__ev128_ldvqw			
Overview	Loads the quad word of the vector register. This instruction reads the quad word at the address specified in ptr and stores the value to the result vector register.		
Declaration	`__ev128_f32__ __ev128_ldvqw(void *ptr);`		
Argument	[IN]	void *ptr	: Specifies the start address of the quad word loads to the vector register.
Return value	Value of __ev128_f32__ type		: Returns the vector containing the quad words.
Remarks			

Table 3-3 Specification of __ev128_ldvw_mask Function

__ev128_ldvw_mask			

Overview	Loads/updates the word to the specified element of the vector register. This instruction reads the word at the address specified in ptr, and returns the vector whose elements are combined from the word and elements in vector register, according to the 4-bit immediate values in mask, as following:
	`val = *ptr`
	`res[w0] = ((mask & (1<<0)) == 1) ? val : x[w0]`
	`res[w1] = ((mask & (1<<1)) == 1) ? val : x[w1]`
	`res[w2] = ((mask & (1<<2)) == 1) ? val : x[w2]`
	`res[w3] = ((mask & (1<<3)) == 1) ? val : x[w3]`		

Declaration	`__ev128_f32__ __ev128_ldvw_mask(ghs_c_int__ mask, void *ptr, __ev128_f32__ x);`		
Argument	[IN]	__ghs_c_int__ mask	: Specifies the element of the vector register to update.
	[IN]	void *ptr	: Specifies the address of the word to load.
	[IN]	__ev128_f32__ x	: Specifies the vector register to load/update.
Return value	Value of __ev128_f32__ type		: Returns the vector containing the quad words.
Remarks			

Table 3-4 Specification of __ev128_stvqw Function

__ev128_stvqw			
Overview	Stores the quad word of the vector register to the specified address by ptr.		
Declaration	`void __ev128_stvqw(__ev128_f32__ x, void *ptr);`		
Argument	[IN]	__ev128_f32__ x	: Specifies the vector register to read.
	[IN]	void *ptr	: Specifies the address to store the read quad word.
Return value			
Remarks			

Table 3-5 Specification of __ev128_addfs_4 Function

__ev128_addfs_4			
Overview	Performs the single-precision floating-point addition. It is executed to the four elements of the vector register specified as the argument, and the result is stored to the vector register.		
Declaration	`__ev128_f32__ __ev128_addfs_4(__ev128_f32__ x, __ev128_f32__ y);`		
Argument	[IN]	__ev128_f32__ x	: Specifies the vector register that is added.
	[IN]	__ev128_f32__ y	: Specifies the vector register that is added.
Return value	Value of type	__ev128_f32__	: Returns the vector containing the addition result.
Remarks			

Table 3-6 Specification of __ev128_subfs_4 Function

__ev128_subfs_4			
Overview	Performs the single-precision floating-point subtraction. It is executed to the four elements of the vector register specified as the argument, and the result is stored to the vector register.		
Declaration	`__ev128_f32__ __ev128_subfs_4(__ev128_f32__ x, __ev128_f32__ y);`		
Argument	[IN]	__ev128_f32__ x	: Specifies the vector register that is subtrahend.
	[IN]	__ev128_f32__ y	: Specifies the vector register that is minuend.
Return value	Value of __ev128_f32__ type		: Returns the vector containing the subtraction result.
Remarks			

Table 3-7 Specification of __ev128_divfs_4 Function

__ev128_divfs_4			
Overview	Performs the single-precision floating-point division. It is executed to the four elements of the vector register specified as the argument, and the result is stored to the vector register.		
Declaration	`__ev128_f32__ __ev128_divfs_4(__ev128_f32__ x, __ev128_f32__ y);`		
Argument	[IN]	__ev128_f32__ x	: Specifies the vector register that is divisor.
	[IN]	__ev128_f32__ y	: Specifies the vector register that is dividend.
Return value	Value of __ev128_f32__ type		: Returns the vector containing the division result.
Remarks			

Table 3-8 Specification of __ev128_fmafs_4 Function

__ev128_fmafs_4			
Overview	Performs the single-precision floating-point fused multiply-addition. It is executed to the four elements of the vector register specified as the argument, and the result is stored to the vector register.		
Declaration	`__ev128_f32__ __ev128_fmafs_4(__ev128_f32__ x, __ev128_f32__ y, __ev128_f32__ z);`		
Argument	[IN]	__ev128_f32__ x	: Specifies the vector register that is multiplied.
	[IN]	__ev128_f32__ y	: Specifies the vector register that is multiplied.
	[IN]	__ev128_f32__ z	: Specifies the vector register that is added.
Return value	Value of __ev128_f32__ Type		: Returns the vector register containing the fused multiply-add result.
Remarks			

Table 3-9 Specification of __ev128_get_f32 Function

__ev128_get_f32			
Overview	Extracts the element specified by eid in vector register and returns it as a 32-bit single-precision floating-point data.		
Declaration	`float __ev128_get_f32(__ev128_f32__ x,int eid);`		
Argument	[IN]	__ev128_f32__ x	: Specifies the vector register that is extracted.
	[IN]	int eid	: Specifies the elements of the vector register that is extracted.
Return value	Value of float type		: Returns the vector containing the 32bit single precision floating-point data.
Remarks			

3.4.2 Data Size

The FXU instruction processes the four elements at a time. Therefore, If the weight matrix data column size and bias data size are not multiples of four, you need to expand to multiples of four by zero padding.

Figure 3-5 shows the example of zero padding in the operation that adds a bias to the product of a matrix and a vector for the size of pattern A (number of input elements: 10, number of output elements: 10).

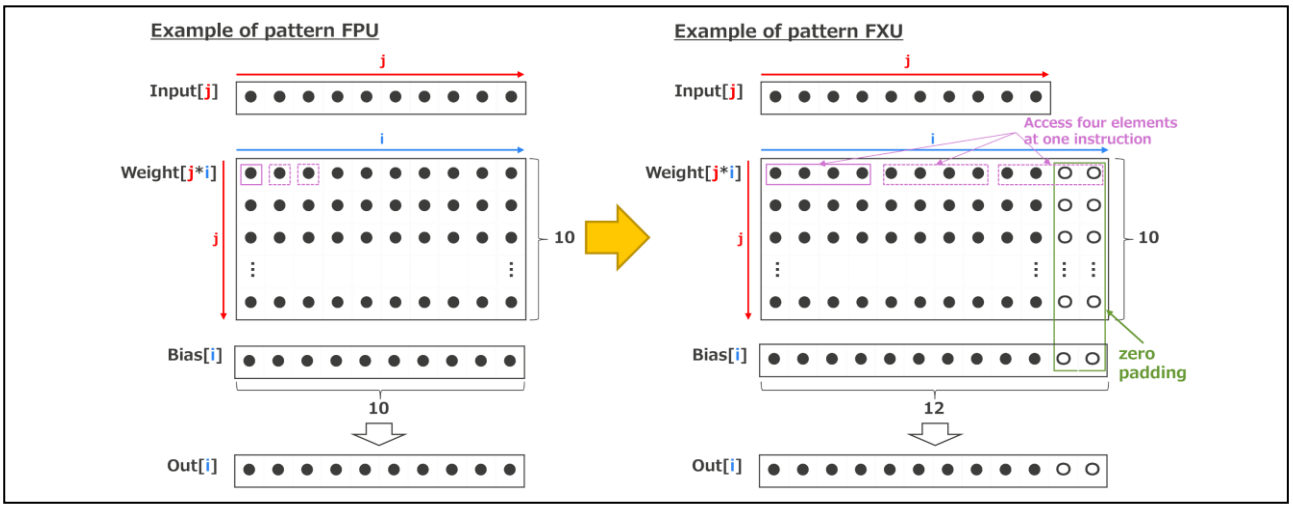


Figure 3-5 Zero padding of Weight Matrix Data and Bias Data

Also, the macro constants corresponding to the weight matrix data column size and the bias data size must be the multiples of four for the data size changes. At the same time, the size of the temporally variable that is stored the calculation result must be the multiples of four. In this sample software, all of these are defined by XX_OUTUNIT. The setting example of this sample software is shown below.

Table 3-10 Macro Constant Setting when Using FXU

Pattern		Input Data	Weight matrix data of Middle layer 1		Weight matrix data of Middle layer 2		Weight matrix data of output layer	
			Row size	Column size	Row size	Column size	Row size	Column size
		INPUT_UNIT	MID1_INUNIT	MID1_OUTUNIT	MID2_INUNIT	MID2_OUTUNIT	OUTPUT_INUNIT	OUTPUT_OUTUNIT
FXU	A	10	10	12	10	12	10	12
	B	10	10	20	20	20	20	12
	C	10	10	32	30	32	30	12

3.4.3 Alignment Specification

The data that becomes the source and the destination of the instruction for FXU must be aligned properly. If not, the misaligned error will occur. Table 3-11 shows the proper data alignment conditions.

Table 3-11 Data Align Condition

Execution Instruction		Data Align Condition		
FXU-Specific Instruction	Data Access Size	32b	64b	128b
LDV.W, STV.W	32b	OK	OK	OK
LDV.DW, STV.DW, LDVZ.H4, STVZ.H4	64b	NG	OK	OK
LDV.QW, STV.QW	128b	NG	NG	OK

The following is the example of allocating data on the 128bit (16bite) boundary by GHS compiler.

```
#pragma alignvar (16)
float data[8];
```

4. Performance Comparison of FPU and FXU

Measures the processing time of this sample software when FPU or FXU is used and compares them.

4.1 Measurement Condition

In this measurement example, the processing time is measure by the following conditions.

- OS timer is used for the measurement of processing time.

（1） Compiler Condition

- Using GHS Compiler v2021.1.5
- Option：-cpu=rh850g4mh -sda=all -large_sda -Ospeed -Onounroll -rh850_fxu -fastmath -prepare_dispose -no_callt

（2） Evaluation Environment

- Integrated Development Environment：GHS MULTI
- Emulator：E2 emulator
- Evaluation Board：RH850/U2B-468BGA PiggyBack board (Y-RH850-U2B-468PIN-PB-T1-V1)
- MCU：RH850/U2B10-FCC (R7F702Z21EDBG)

4.2 Measurement Result

Table 4-1 shows the processing time when using FPU and FXU. In this measurement, measurement of patterns (C-2, C-3) with an increased number of layers is also added. Figure 4-1 shows the graph plotted with the horizontal axis as the number of the Fused Multiply-add (FMA).

Table 4-1 Processing Time Measurement Result

Pattern		Unit Number				Number of Processing Executions			Processing Time [us]	
		Input layer	Middle layer	Output layer	Number of layer	FMA	tanh	exp	FPU	FXU
tanh	A	10	10	10	3	300	30	0	**11.0**	**12.5**
	B	10	20	10	3	800	50	0	**20.9**	**19.4**
	C-1	10	30	10	3	1500	70	0	**31.5**	**27.9**
	C-2	10	30	10	5	3300	130	0	**61.9**	**51.0**
	C-3	10	30	10	7	5100	190	0	**92.4**	**74.2**
tanh_usingexp	A	10	10	10	3	300	0	30	**7.5**	**8.0**
	B	10	20	10	3	800	0	50	**14.9**	**11.7**
	C-1	10	30	10	3	1500	0	70	**24.1**	**18.1**
	C-2	10	30	10	5	3300	0	130	**49.9**	**34.6**
	C-3	10	30	10	7	5100	0	190	**75.8**	**51.4**
sigmoid	A	10	10	10	3	300	0	30	**7.0**	**8.3**
	B	10	20	10	3	800	0	50	**14.3**	**12.1**
	C-1	10	30	10	3	1500	0	70	**23.1**	**18.4**
	C-2	10	30	10	5	3300	0	130	**47.5**	**35.0**
	C-3	10	30	10	7	5100	0	190	**71.9**	**51.7**
ReLU	A	10	10	10	3	300	0	0	**3.4**	**3.3**
	B	10	20	10	3	800	0	0	**8.1**	**4.6**
	C-1	10	30	10	3	1500	0	0	**14.4**	**7.5**
	C-2	10	30	10	5	3300	0	0	**31.2**	**14.4**
	C-3	10	30	10	7	5100	0	0	**48.0**	**21.4**

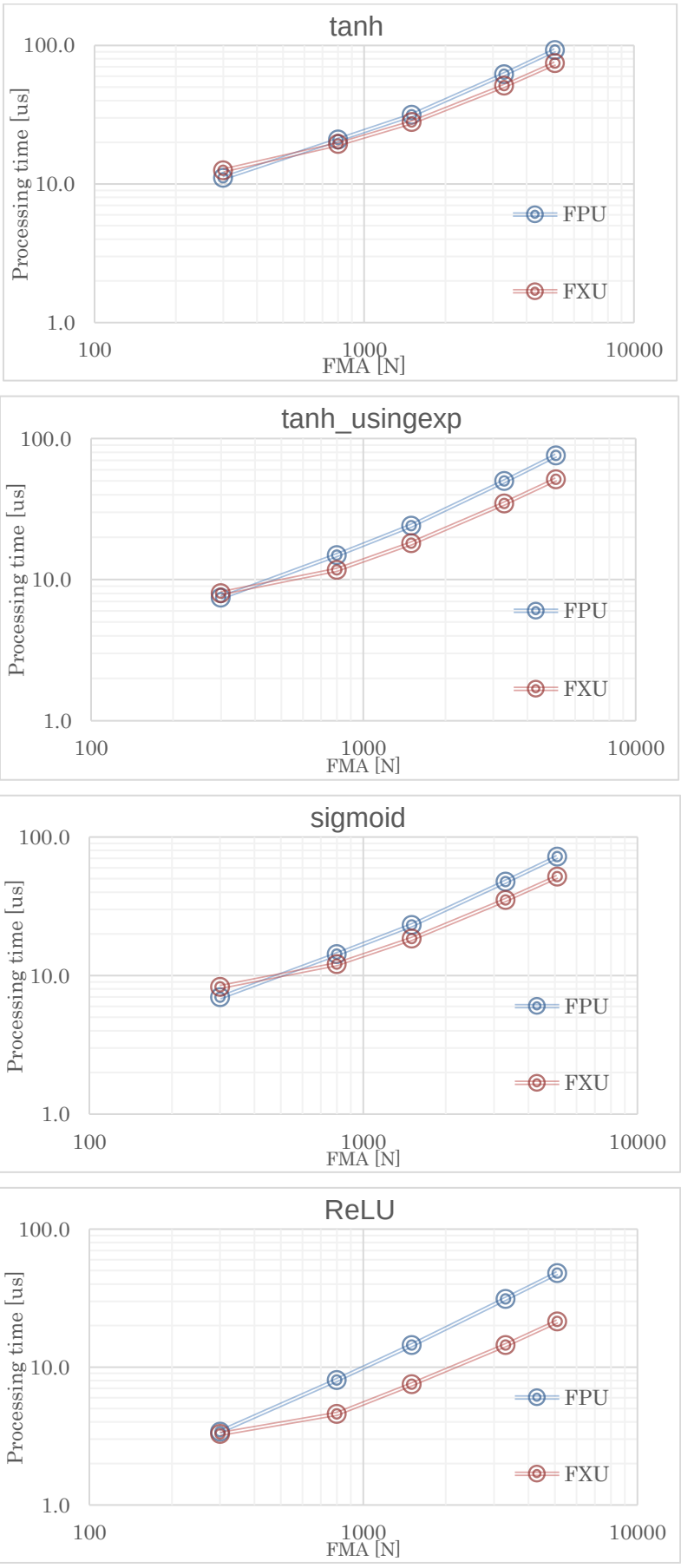


Figure 4-1 Processing Time Measurement Result Graph

Appendix

CPU Configuration of RH850/U2Bx Series

Table 5-1 shows the CPU configuration of RH850/U2Bx.

Please note that the placement of the FXU-equipped CPU is different for each product.

Table 5-1 CPU Configuration of RH850/U2Bx

Cluster	CPU (PEID)	U2B6	U2B10	
		3+2	4+2	3+3
0	0	DCLS **w/ FXU**	DCLS **w/ FXU**	DCLS **w/ FXU**
	1	DCLS	DCLS	DCLS
1	2	SNGL	SNGL **w/ FXU** *1	DCLS **w/ FXU** *1
	3	-	SNGL	-

【Note】DCLS：Dual Core Lockstep Core
SNGL：Single Core
Note 1. FXU is only in FCC device.

Revision History

Rev.	Date	Description	
		Page	Summary
1.00	July 16, 2024	-	New Release

General Precautions in the Handling of Microprocessing Unit and Microcontroller Unit Products

The following usage notes are applicable to all Microprocessing unit and Microcontroller unit products from Renesas. For detailed usage notes on the products covered by this document, refer to the relevant sections of the document as well as any technical updates that have been issued for the products.

1. Precaution against Electrostatic Discharge (ESD)

   A strong electrical field, when exposed to a CMOS device, can cause destruction of the gate oxide and ultimately degrade the device operation. Steps must be taken to stop the generation of static electricity as much as possible, and quickly dissipate it when it occurs. Environmental control must be adequate. When it is dry, a humidifier should be used. This is recommended to avoid using insulators that can easily build up static electricity. Semiconductor devices must be stored and transported in an anti-static container, static shielding bag or conductive material. All test and measurement tools including work benches and floors must be grounded. The operator must also be grounded using a wrist strap. Semiconductor devices must not be touched with bare hands. Similar precautions must be taken for printed circuit boards with mounted semiconductor devices.

2. Processing at power-on

   The state of the product is undefined at the time when power is supplied. The states of internal circuits in the LSI are indeterminate and the states of register settings and pins are undefined at the time when power is supplied. In a finished product where the reset signal is applied to the external reset pin, the states of pins are not guaranteed from the time when power is supplied until the reset process is completed. In a similar way, the states of pins in a product that is reset by an on-chip power-on reset function are not guaranteed from the time when power is supplied until the power reaches the level at which resetting is specified.

3. Input of signal during power-off state

   Do not input signals or an I/O pull-up power supply while the device is powered off. The current injection that results from input of such a signal or I/O pull-up power supply may cause malfunction and the abnormal current that passes in the device at this time may cause degradation of internal elements. Follow the guideline for input signal during power-off state as described in your product documentation.

4. Handling of unused pins

   Handle unused pins in accordance with the directions given under handling of unused pins in the manual. The input pins of CMOS products are generally in the high-impedance state. In operation with an unused pin in the open-circuit state, extra electromagnetic noise is induced in the vicinity of the LSI, an associated shoot-through current flows internally, and malfunctions occur due to the false recognition of the pin state as an input signal become possible.

5. Clock signals

   After applying a reset, only release the reset line after the operating clock signal becomes stable. When switching the clock signal during program execution, wait until the target clock signal is stabilized. When the clock signal is generated with an external resonator or from an external oscillator during a reset, ensure that the reset line is only released after full stabilization of the clock signal. Additionally, when switching to a clock signal produced with an external resonator or by an external oscillator while program execution is in progress, wait until the target clock signal is stable.

6. Voltage application waveform at input pin

   Waveform distortion due to input noise or a reflected wave may cause malfunction. If the input of the CMOS device stays in the area between V_{IL} (Max.) and V_{IH} (Min.) due to noise, for example, the device may malfunction. Take care to prevent chattering noise from entering the device when the input level is fixed, and also in the transition period when the input level passes through the area between V_{IL} (Max.) and V_{IH} (Min.).

7. Prohibition of access to reserved addresses

   Access to reserved addresses is prohibited. The reserved addresses are provided for possible future expansion of functions. Do not access these addresses as the correct operation of the LSI is not guaranteed.

8. Differences between products

   Before changing from one product to another, for example to a product with a different part number, confirm that the change will not lead to problems. The characteristics of a microprocessing unit or microcontroller unit products in the same group but having a different part number might differ in terms of internal memory capacity, layout pattern, and other factors, which can affect the ranges of electrical characteristics, such as characteristic values, operating margins, immunity to noise, and amount of radiated noise. When changing to a product with a different part number, implement a system-evaluation test for the given product.

Notice

1. Descriptions of circuits, software and other related information in this document are provided only to illustrate the operation of semiconductor products and application examples. You are fully responsible for the incorporation or any other use of the circuits, software, and information in the design of your product or system. Renesas Electronics disclaims any and all liability for any losses and damages incurred by you or third parties arising from the use of these circuits, software, or information.
2. Renesas Electronics hereby expressly disclaims any warranties against and liability for infringement or any other claims involving patents, copyrights, or other intellectual property rights of third parties, by or arising from the use of Renesas Electronics products or technical information described in this document, including but not limited to, the product data, drawings, charts, programs, algorithms, and application examples.
3. No license, express, implied or otherwise, is granted hereby under any patents, copyrights or other intellectual property rights of Renesas Electronics or others.
4. You shall be responsible for determining what licenses are required from any third parties, and obtaining such licenses for the lawful import, export, manufacture, sales, utilization, distribution or other disposal of any products incorporating Renesas Electronics products, if required.
5. You shall not alter, modify, copy, or reverse engineer any Renesas Electronics product, whether in whole or in part. Renesas Electronics disclaims any and all liability for any losses or damages incurred by you or third parties arising from such alteration, modification, copying or reverse engineering.
6. Renesas Electronics products are classified according to the following two quality grades: “Standard” and “High Quality”. The intended applications for each Renesas Electronics product depends on the product’s quality grade, as indicated below.
   "Standard": Computers; office equipment; communications equipment; test and measurement equipment; audio and visual equipment; home electronic appliances; machine tools; personal electronic equipment; industrial robots; etc.
   "High Quality": Transportation equipment (automobiles, trains, ships, etc.); traffic control (traffic lights); large-scale communication equipment; key financial terminal systems; safety control equipment; etc.

   Unless expressly designated as a high reliability product or a product for harsh environments in a Renesas Electronics data sheet or other Renesas Electronics document, Renesas Electronics products are not intended or authorized for use in products or systems that may pose a direct threat to human life or bodily injury (artificial life support devices or systems; surgical implantations; etc.), or may cause serious property damage (space system; undersea repeaters; nuclear power control systems; aircraft control systems; key plant systems; military equipment; etc.). Renesas Electronics disclaims any and all liability for any damages or losses incurred by you or any third parties arising from the use of any Renesas Electronics product that is inconsistent with any Renesas Electronics data sheet, user’s manual or other Renesas Electronics document.
7. No semiconductor product is absolutely secure. Notwithstanding any security measures or features that may be implemented in Renesas Electronics hardware or software products, Renesas Electronics shall have absolutely no liability arising out of any vulnerability or security breach, including but not limited to any unauthorized access to or use of a Renesas Electronics product or a system that uses a Renesas Electronics product. RENESAS ELECTRONICS DOES NOT WARRANT OR GUARANTEE THAT RENESAS ELECTRONICS PRODUCTS, OR ANY SYSTEMS CREATED USING RENESAS ELECTRONICS PRODUCTS WILL BE INVULNERABLE OR FREE FROM CORRUPTION, ATTACK, VIRUSES, INTERFERENCE, HACKING, DATA LOSS OR THEFT, OR OTHER SECURITY INTRUSION (“Vulnerability Issues”). RENESAS ELECTRONICS DISCLAIMS ANY AND ALL RESPONSIBILITY OR LIABILITY ARISING FROM OR RELATED TO ANY VULNERABILITY ISSUES. FURTHERMORE, TO THE EXTENT PERMITTED BY APPLICABLE LAW, RENESAS ELECTRONICS DISCLAIMS ANY AND ALL WARRANTIES, EXPRESS OR IMPLIED, WITH RESPECT TO THIS DOCUMENT AND ANY RELATED OR ACCOMPANYING SOFTWARE OR HARDWARE, INCLUDING BUT NOT LIMITED TO THE IMPLIED WARRANTIES OF MERCHANTABILITY, OR FITNESS FOR A PARTICULAR PURPOSE.
8. When using Renesas Electronics products, refer to the latest product information (data sheets, user’s manuals, application notes, “General Notes for Handling and Using Semiconductor Devices” in the reliability handbook, etc.), and ensure that usage conditions are within the ranges specified by Renesas Electronics with respect to maximum ratings, operating power supply voltage range, heat dissipation characteristics, installation, etc. Renesas Electronics disclaims any and all liability for any malfunctions, failure or accident arising out of the use of Renesas Electronics products outside of such specified ranges.
9. Although Renesas Electronics endeavors to improve the quality and reliability of Renesas Electronics products, semiconductor products have specific characteristics, such as the occurrence of failure at a certain rate and malfunctions under certain use conditions. Unless designated as a high reliability product or a product for harsh environments in a Renesas Electronics data sheet or other Renesas Electronics document, Renesas Electronics products are not subject to radiation resistance design. You are responsible for implementing safety measures to guard against the possibility of bodily injury, injury or damage caused by fire, and/or danger to the public in the event of a failure or malfunction of Renesas Electronics products, such as safety design for hardware and software, including but not limited to redundancy, fire control and malfunction prevention, appropriate treatment for aging degradation or any other appropriate measures. Because the evaluation of microcomputer software alone is very difficult and impractical, you are responsible for evaluating the safety of the final products or systems manufactured by you.
10. Please contact a Renesas Electronics sales office for details as to environmental matters such as the environmental compatibility of each Renesas Electronics product. You are responsible for carefully and sufficiently investigating applicable laws and regulations that regulate the inclusion or use of controlled substances, including without limitation, the EU RoHS Directive, and using Renesas Electronics products in compliance with all these applicable laws and regulations. Renesas Electronics disclaims any and all liability for damages or losses occurring as a result of your noncompliance with applicable laws and regulations.
11. Renesas Electronics products and technologies shall not be used for or incorporated into any products or systems whose manufacture, use, or sale is prohibited under any applicable domestic or foreign laws or regulations. You shall comply with any applicable export control laws and regulations promulgated and administered by the governments of any countries asserting jurisdiction over the parties or transactions.
12. It is the responsibility of the buyer or distributor of Renesas Electronics products, or any other party who distributes, disposes of, or otherwise sells or transfers the product to a third party, to notify such third party in advance of the contents and conditions set forth in this document.
13. This document shall not be reprinted, reproduced or duplicated in any form, in whole or in part, without prior written consent of Renesas Electronics.
14. Please contact a Renesas Electronics sales office if you have any questions regarding the information contained in this document or Renesas Electronics products.

(Note1) “Renesas Electronics” as used in this document means Renesas Electronics Corporation and also includes its directly or indirectly controlled subsidiaries.
(Note2) “Renesas Electronics product(s)” means any product developed or manufactured by or for Renesas Electronics.

(Rev.5.0-1 October 2020)

Corporate Headquarters

TOYOSU FORESIA, 3-2-24 Toyosu,
Koto-ku, Tokyo 135-0061, Japan
www.renesas.com

Trademarks

Renesas and the Renesas logo are trademarks of Renesas Electronics Corporation. All trademarks and registered trademarks are the property of their respective owners.

Contact information

For further information on a product, technology, the most up-to-date version of a document, or your nearest sales office, please visit:
www.renesas.com/contact/.